

SITE EXPLORATION: A PROBABILISTIC APPROACH

by

GREGORY BERT BAECHER  
BS, University of California, Berkeley  
(1968)  
SM, Massachusetts Institute of Technology  
(1970)

Submitted in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy

Massachusetts Institute of Technology  
September, 1972

Signature of Author .. Department of Civil Engineering, August 14, 1972

Certified by .... Thesis Co-Supervisor

and ..... Thesis Co-Supervisor

Accepted by .....  
Chairman, Department Committee on Graduate Students of the  
Department of Civil Engineering



ABSTRACT

SITE EXPLORATION: A PROBABILISTIC APPROACH

by

GREGORY BERT DAECHER

Submitted to the Department of Civil Engineering  
on August 14, 1972 in partial fulfillment of the  
requirements for the degree of Doctor of Philosophy.

Probability theory is an aid which an engineer may employ in optimizing strategy decisions and in drawing inferences from observed data. The present study investigated the applicability of probability theory to the exploration of rock masses. Despite the widely voiced opinion that exploration is a "statistical" problem, no published attempt to comprehensively investigate the application of probability to exploration in civil engineering had been undertaken prior to the present work.

A five part, problematic model of the exploration process was developed which classifies specific exploration problems according to their analytical basis. The five problem classes are: reconnaissance, pattern recognition and reconstruction (mapping), search, joint surveys, and mechanical testing. Of these, the three treating geometric exploration were studied intensively (i.e., pattern recognition and reconstruction, search, and joint surveys). Because exploration depends heavily on judgement and treats "one-time" events, it is most appropriately handled (probabilistically) using subjectivist degree-of-belief theory; although, pragmatically, certain problems can be identified which are easily and satisfactorily handled using relative-frequentist methods.

The general conclusion of the study was that probability theory, beyond sampling statistics, is applicable to exploring rock masses, and may result in large savings of exploration effort. Specifically, well developed analytical techniques presently exist for optimizing effort allocation in problems of search and of joint surveys; their application to case studies is urged. Problems of pattern recognition and reconstruction, on the other hand, border on inductive reasoning, and the development of practical techniques to handle them will require considerable theoretical effort. Several recommendations for continued study are offered.

Thesis co-supervisor:	Ronald C. Hirschfeld
Title:	Associate Professor of Civil Engineering
Thesis co-supervisor:	C. Allin Cornell
Title:	Associate Professor of Civil Engineering

### ACKNOWLEDGEMENTS

I am deeply indebted to Professor Ronald C. Hirschfeld who has been my advisor on two theses, who has been an insightful critic, and who has constantly stood as a model of the ethical man.

I am also deeply indebted to Professor Herbert H. Einstein who offered daily input throughout the first half of the investigation, and whose ideas and opinions had a significant effect on the final result.

Professor C. Allin Cornell has served as my buttress for probability theory, and has led me out of innumerable theoretical difficulties, which often I was not even aware I was in. How he remained cheerful while reading and rereading the many less-than-refined uses of probability theory appearing in earlier drafts is almost beyond understanding, yet I am very grateful for it.

Professor Warren J. Baker has likewise read through several drafts of the work, and has offered many suggestions on the relationship of probability theory to soil engineering practice, as well as several intriguing special problems.

I would like particularly to express my gratitude to the National Science Foundation whose program of national graduate fellowships has allowed me to undertake an investigation of personal interest free from the pressures and confines of sponsored research, and in which the direction and topics of inquiry were purely of my own choosing.

## PREFACE

Site exploration is a subjective undertaking which relies heavily on the judgement of engineers associated with it. However, since large sums of money are spent on exploration, methods of making the process as efficient as possible are to be desired. This thesis represents an attempt to organize a comprehensive, rational approach to exploring rock masses, which can be used to optimize allocations of effort.

Uncertainties are rife in exploration, and one must deal with them if he is ever to analytically treat the process of strategy formulation. Although engineers have long recognized the "statistical nature" of exploration, they have given probabilistic approaches little attention.

The present study began as an attempt to coalesce existing work on the application of probability theory to exploration into an organized whole; however, as the investigation proceeded, an almost total lack of work on the topic became increasingly apparent, and the present study, more and more, began to require the development of a basic philosophy for the use of quantitative methods in exploration. Previous work, which had dealt primarily with sampling statistics and inference from laboratory testing, seemed to treat issues which were peripheral to the real problems of formulating strategies. As a result, we were required to structure the problems of exploration from "scratch," and to search for methods of treating them in

disciplines other than civil engineering and geology. The product of this effort we hope will serve as a foundation and catalyst for continued inquiry. The extent and difficulties of the work ahead are great, but the rewards, both academic and commercial, are great as well.

Unlike many problems of engineering geology, rationalization of site exploration strategies borders on inductive reasoning: it depends heavily on experience and "knowledge of geology." This makes the problem fascinating, but also requires that approaches to its solution be firmly founded in the philosophy-of-science.

Robert Frost's phrase, "good fences make good neighbors" applies nowhere more aptly than to academia. The present work, which straddles the "fences" of probability theory and engineering geology, will surely draw criticism from orthodox practitioners in both disciplines; yet the nature of the work requires input from both sources, and continued study must include specialists in each. We are in the midst of an interdisciplinary renaissance in which methods developed in one discipline are being brought to bear on problems in others, thus reaping great benefit. Hopefully, the present work will contribute to that trend.

This thesis is divided into three parts. In the first we discuss questions of the philosophical basis of probability and decision, and the inclusion of judgement in decision processes. In the second we present a model of the exploration process, and general conclusions from the study. In the third we discuss three components

of our model in detail and present analytical methods for treating them. As a first, casual review of this work one should probably read Section 2 (i.e., Chapters 3 and 4).

The present work represents an early effort into a subject whose breadth and depth are considerable. As such, a conscious decision was made to comprehensively study the whole topic rather than to intensively study one aspect of it. This policy decision, as any other, imposed limitations on the results. In places, mathematical proofs and techniques are not as sophisticated as they might have been were one investigating only a specific, isolated problem; and we have not been able to include major case studies using the techniques presented. However, these limitations are recognized as being inherent to any initial inquiry which attempts comprehensiveness.

Orthodox engineering geologists will undoubtedly criticize the use of decision models on the grounds that they require quantitative assessments of judgement which may be difficult to obtain in practice -- and, of course, they are correct. However, the major benefit of a decision-theoretic approach to exploration is not that it results in precise optimization of strategies, but rather that it presents the engineer with an entirely new perspective on the exploration process. It structures decisions and in so doing indicates facets of judgement which must be most critically reflected upon. Furthermore, it gives the engineer a better feeling for the uncertainties he must deal with no matter how he actually arrives at decisions.

## TABLE OF CONTENTS

	<u>Page</u>
TITLE PAGE.....	1
ABSTRACT.....	2
ACKNOWLEDGEMENTS.....	3
PREFACE.....	4
TABLE OF CONTENTS.....	7
<u>PART 1: SOME THEORETICAL AND PHILOSOPHICAL BACKGROUND.....</u>	<u>11</u>
CHAPTER 1: SOME PHILOSOPHICAL ISSUES OF PROBABILITY AND DECISION.....	11
Section 1.0: Introduction.....	11
Section 1.1: Probability Defined.....	11
1.1.1 Frequency Definitions.....	13
1.1.2 Degree-of-Belief Definitions.....	20
Section 1.2: Decision.....	31
1.2.1 Inference.....	32
1.2.2 Decision.....	35
1.2.3 Comment.....	45
CHAPTER 2: SUBJECTIVE PROBABILITY ASSESSMENT.....	46
Section 2.1: Theory.....	46
Section 2.2: Action Approach.....	48
2.2.1 Procedural Effects.....	50
2.2.2 Descriptive vs. Prescriptive Models.....	52
2.2.3 Ambiguity and Stability.....	58
Section 2.3: Methods of Assessment.....	60
2.3.1 Single, Non-repeatable Events.....	62
2.3.2 Random Variables.....	67
2.3.3 Multivariate Assessment: Dependent Variables.....	70

TABLE OF CONTENTS  
(continued)

	<u>Page</u>
Section 2.4: External Validity.....	77
Section 2.5: Some Comments in Summary.....	89
Section 2A. Appendix: Example Assessments.....	92
<u>PART II: PROBABILITY THEORY IN SITE EXPLORATION.....</u>	<u>98</u>
CHAPTER 3: SITE EXPLORATION: PROBABILITY.....	98
Section 3.1: Site Exploration.....	98
3.1.1 Nature of Uncertainty in Site Exploration..	102
3.1.2 Consequences of Uncertainty.....	105
3.1.3 Summary.....	106
Section 3.2: Probability Theory in Site Exploration...	106
3.2.1 Decision Models.....	107
3.2.2 Inference.....	111
3.2.3 Bayesian Methods.....	115
3.2.4 Size of Project to which Probability Theory is Applicable.....	117
3.2.5 Summary.....	118
Section 3.3: Goals of the Present Work.....	119
CHAPTER 4: CONCLUSIONS AND RECOMMENDATIONS.....	123
Section 4.1: Conclusions.....	123
4.1.1 General Conclusions.....	123
4.1.2 Specific Conclusions.....	125
Section 4.2: Recommendations for Additional Study.....	129
4.2.1 General Recommendations.....	130
4.2.2 Specific Recommendations.....	132
4.2.3 Topics of Related Interest.....	135



TABLE OF CONTENTS  
(continued)

	<u>Page</u>
<b><u>PART III: SPECIFIC PROBLEM CLASSES</u></b> .....	138
<b>CHAPTER 5: GEOLOGICAL MAPPING: PATTERN RECOGNITION AND RECONSTRUCTION</b> .....	138
Section 5.1: Introduction.....	138
Section 5.2: Discrete Attributes.....	139
5.2.1 Classification.....	140
5.2.2 Probabilistic Mapping and Rules for Inference.....	174
5.2.3 Uncertainty Functions and Entropy.....	216
5.2.4 Optimal Allocation.....	228
Section 5.3: Continuous Attributes.....	239
5.3.1 Inference.....	242
5.3.2 Allocating Sampling Effort.....	249
5.3.3 Summary.....	253
<b>CHAPTER 6: SEARCH THEORY</b> .....	255
Section 6.0: Introduction.....	255
6.0.1 Organization of Chapter.....	256
6.0.2 History of Search Theory.....	256
Section 6.1: Single-Stage Search.....	258
6.1.1 Random Search.....	259
6.1.2 Grid Search.....	262
6.1.3 Optimal Allocation.....	294
Section 6.2: Two-Stage Search.....	331
Section 6.3: Quantity of Total Effort.....	341

TABLE OF CONTENTS  
(continued)

	<u>Page</u>
Section 6.4: Sequential Search.....	349
6.4.1 General Solution.....	350
6.4.2 Optimal Stopping.....	360
6.4.3 Exponential Saturation.....	364
Section 6.5: Response Surfaces and "Optimal Seeking"..	367
CHAPTER 7: JOINT SURVEYS.....	372
Section 7.1: Principles of Sampling Joint Populations.	374
7.1.1 Statistical Sampling.....	374
7.1.2 Sampling Theory Applied to Joint Surveys...	381
7.1.3 A Sampling Plan for Joint Surveys.....	394
7.1.4 Intensity of Jointing.....	401
Section 7.2: Inferring the Nature of Joint Populations.....	404
7.2.1 Methods Not Using Distribution Models for Joint Populations.....	405
7.2.2 Mathematical Models for Orientation Dis- tributions.....	429
Section 7.3 Summary.....	459
Section 7A. Appendix: Classical Sampling Plans.....	461
7A.1 Simple Random Sampling.....	463
7A.2 Systematic Sampling.....	470
7A.3 Stratified Random Sampling.....	475
7A.4 Cluster Sampling.....	490
REFERENCES.....	502

## CHAPTER I

SOME PHILOSOPHICAL ISSUES OF PROBABILITY  
AND DECISIONSection 1.0 Introduction

Before applying probability theory to problems of site exploration we must understand the meaning of the term probability, and how it relates to decision making. The mathematical theory of probability is predicated upon three axioms in which probability is accepted as a primitive term, that is, in which the meaning of probability is never defined.\* The mathematical theory is internally consistent (all of the mathematical results can be derived from the three axioms) and universally accepted; but being a primitive term, probability is removed from "mathematics," and must be considered in a way which can only be called philosophical.

In this chapter we discuss existing schools of thought on the meaning of probability (section 1.1) and its use in decision making (section 1.2), so that the work of this thesis may be viewed in proper perspective.

Section 1.1 Probability Defined

There are two major schools of thought on the meaning of

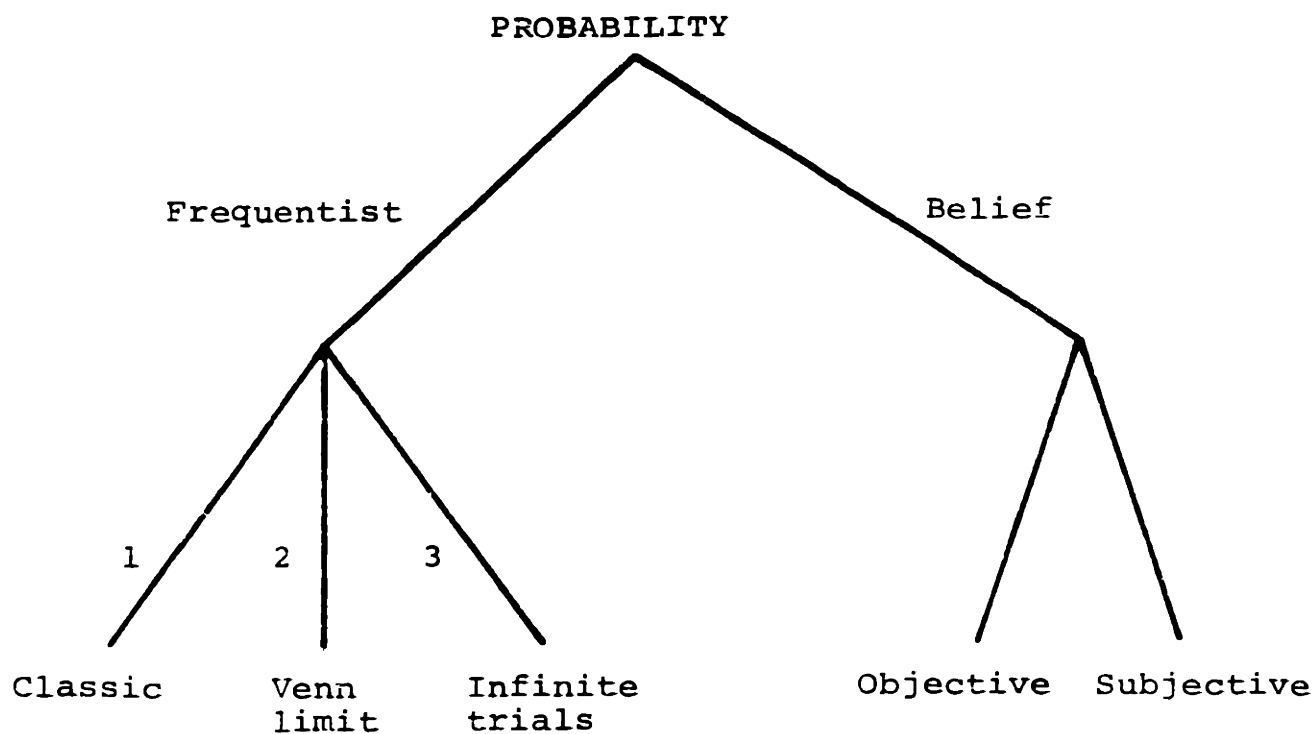
---

\*For reasons which we will not go into, Kolmogorov (1933) proposed predicating mathematical probability theory on a set of five axioms (which include the usual three).

probability. One school holds that probability should be defined as a relative frequency, while the other holds that it should be defined as a degree-of-belief. The two schools often arrive at the same conclusions in specific applications, but their precepts and methods sometimes differ considerably.\*

The structure of the two schools of thought and their respective sub-groups is shown schematically in Figure 1, which our discussion will follow.

Figure 1




---

\*To which Jeffreys (1961) offers the following anecdote attributed to the great British juror Lord Mansfield:

"Lord Mansfield gave the following advice to the newly

### Section 1.1.1 Frequency Definitions

The definitions based on frequency of occurrence are perhaps the definitions which most non-statisticians think of first when asked to define probability. They are based on the relative number of times that a certain result will occur in a series of trials. One definition (#1 below) determines this frequency, hence probability, before any trials have taken place, while the other two base probability on the empirical observation of several trials. Any probability stated before experimentation is known as "a priori" or "prior" probability, and any probability stated as a result of experimentation is known as "a posteriori" or "posterior" probability.

The three most frequent definitions of probability are (see Figure 1):

1. If there are  $n$  possible alternatives, for  $m$  of which  $p$  is true, then the probability of  $p$  is defined as  $m/n$ .
2. If an event occurs a large number of times, then the probability of  $p$  is the limit of the ratio of the number of trials on which  $p$  will be true to the whole number of trials, as the number of trials tends to infinity.

---

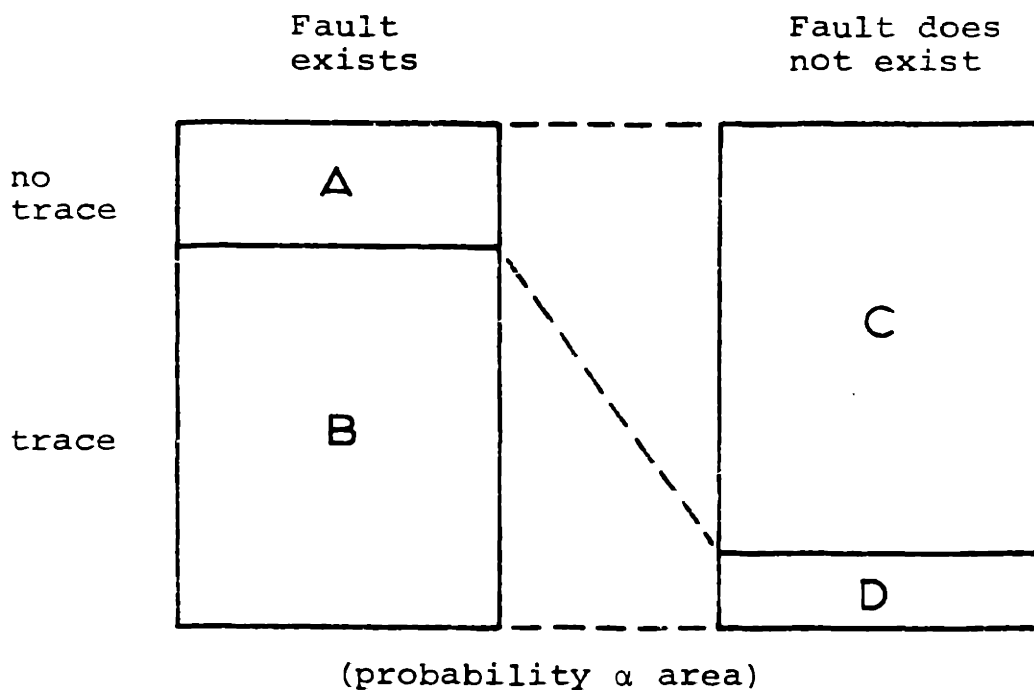
appointed governor of a West Indian island. 'There is no difficulty in deciding a case--only hear both sides patiently, then consider what you think justice requires, and decide accordingly; but never give reasons, for your judgement will probably be right, but your reasons will certainly be wrong.'

3. An infinite number of trials is assumed. The probability of  $p$  is defined as the ratio of the number of cases where  $p$  is true to the whole number.

The first definition is sometimes called the "classical" frequentist definition and has been associated with the work of Neyman, DeMoivre, and Cramer (Jeffreys, 1961); the second is known as the Venn limit, and has been associated with the work of von Mises; the third has been called the infinite trials or infinite ensemble definition, and has been associated with the work of Fisher (although it appeared earlier in the writings of Williard Gibbs to whom we also owe the term "ensemble") (Jeffreys, 1961).

Implicit in the definitions of the frequentists is the assumption that one cannot place probabilities on the state of nature. For example, a frequentist could not say, "the probability that a fault exists at a certain site is 30%," for the fault either exists or doesn't exist (i.e., the probability is either 0 or 1). In the frequentist school one can only place probabilities or confidences on the test or estimation procedure upon which inferences are made. For example, "if the fault existed, these aerial photographs would show evidence of it 90% of the time." This is not the same as saying, "there is a 10% chance that the fault exists if no trace of it is seen in the photographs".

Figure 2



The probability that an incorrect conclusion is drawn from a certain test or estimation procedure is called its significance level. For example, one significance level of the photographs mentioned above equals the probability that when the photographs show a trace and one concludes that there is a fault at the site, he is mistaken. In terms of Fig. 2 (in which probability is proportional to area) this would equal  $D/(C+D)$ , because  $D/(C+D)\%$  of the time when a fault did not exist one would have concluded that it did. We shall later see that the degree-of-belief school does allow the probability of the state of nature to be treated.

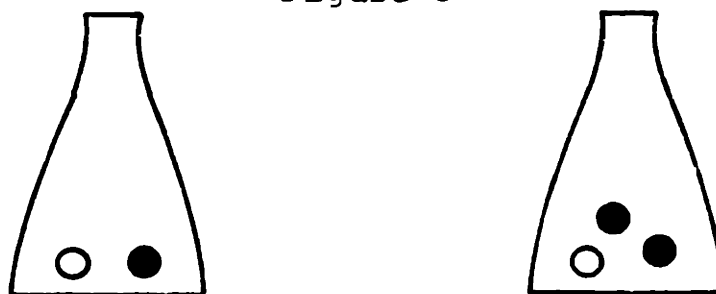
### Deficiencies in Frequentist Definitions

Degree-of-believers, obviously, find deficiencies in all three frequentist definitions, and in the following few pages we shall summarize some of their arguments.

The "Classical" frequentist definition, which determines probabilities a priori, leads in certain cases to assessments which do not obey the product rule (i.e., the probability of the simultaneous occurrence of independent events equals the product of the probabilities of their individual occurrences), and thus are considered unacceptable to degree-of-believers. This is true whenever the probabilities of individual outcomes are biased; that is, some occurring more frequently than others. Since this definition assigns probabilities a priori, bias is not taken into account. Consider a die biased in favor of the number 4, and against the number 3. A "classical" frequentist would say, "there are 6 possible outcomes (n), and there is one outcome favorable to the number 4 (m), therefore the probability of a 4 on any one throw is  $1/6$ . However, this is incorrect because the die is biased toward the 4, and, therefore, the probability of rolling a 4 is greater than  $1/6$ . As a less trivial example, consider a problem posed by Jeffreys (1961):



Figure 3



Two opaque bottles are given, one of which contains a white ball and a black ball, the other of which contains a white ball and two black balls. One bottle is selected at random, and a ball drawn from the bottle, also at random. What is the probability that the ball so drawn is white?

A "classical" frequentist would say that the probability of a white ball  $P(w)$  is  $2/5$ , because there are 5 possibilities and two white balls.

$$P(w) = m/n = 2/5$$

Whereas, many statisticians would hold that the probability of drawing the white ball in bottle 1 equals the probability of drawing bottle 1 times the probability of drawing a white ball once bottle 1 is chosen; similarly for bottle 2. So the total probability of drawing a white ball equals the sum of the probabilities of drawing one from bottles 1 and 2, or  $5/12$ .

$$P(w) = (1/2)(1/2) + (1/2)(1/3) = 5/12$$

Here, most statisticians would favor the latter probability because not all of the balls in the two bottles have an equally likely chance of being selected.

The infinite trials definition (that first put forward by Gibbs), degree-of-believers attack on two points. First, in practical cases one can never define a probability because he can never obtain an infinite number of trials. Second, even if there could be an infinite number of trials, the probability--which must be between 0 and 1 and, hence, finite--would be defined as the ratio of two infinite terms. Such a ratio is indeterminate.

This leaves only the Venn limit definition not countered by a strong argument. Nevertheless, Jeffreys (1961) has raised objection to this definition by noting that there is no a priori reason to believe that a limit at infinity exists for the ratio of successes to failures, nor is there any reason to believe that this limit would be unique if it did exist. So, a unique limit must be postulated in developing a probability theory based on the Venn limit. Furthermore, since the product rule does not follow from this definition, it, too, must be postulated (Jeffreys, 1961); however, from a philosophical view point, one would like to minimize the number of postulates he requires in deriving a theory. Jeffreys also holds that when one assumes a unique limit for the success ratio he necessarily constrains the way in which events occur, and affects the random nature of those events.

So, in summary, degree-of-believers present arguments against each of the three basic relative-frequentist definitions: The classical definition leads to probabilities which do not obey the product rule, the infinite trials definition leads to indeterminate

assessments, and the Venn limit requires one to postulate the existence of a unique limit for the ratio of successes to total trials and the product rule.

A frequentist definition which seems to have fewer critics than the others is a derivative of Fisher's (infinite trials) definition in which the hypothetical infinite population is replaced by a very large finite one. Such a definition avoids the problems of DeMoivre's definition in that it draws its probabilities empirically and not a priori, but there are still some problems: Since the size of the finite population cannot be specified, the actual observations have to be considered as a random sample from some larger (i.e., infinite) population. Probabilities obtained would be a function of the conditions of the experiment, and number of trials, and would have to be so defined.

The frequentist theories give rules for setting up hypotheses to be tested, and for rejecting them in certain circumstances. They do not place probabilities on the actual state of nature; and should a hypothesis be rejected, they do not say which hypothesis should replace it. One danger of such a set of rules is that a hypothesis which explains part, but not all, of the variation in a set of observations may be rejected while it is still of some use. Consider a hypothesis which explains 90% of the observed variation in some phenomenon. This hypothesis may be of considerable help in prediction even though to accept it as true we might not be willing to tolerate a 10% discrepancy.

The frequentist theories give answers in the form of the probability of different sets of observations occurring, given some hypothesis. The practical problem is just the reverse of this one. That is, the practical problem is what is the probability of some hypothesis being true, given the observations recorded.

Degree-of-believers maintain that frequentist theories have no advantage over degree-of-belief theories, rather they just attempt to give some meaning to "chance."

#### Section 1.1.2 Degree-of-Belief Definitions

The proponents of the degree-of-belief definition maintain that every statistician uses degree-of-belief whether or not he recognizes it, and in practice the relative frequency definition is almost never used (Jeffreys, 1961). In terms of real world problems the statistical measure needed when making decisions is not the relative frequency, but the degree-of-belief that some event will occur or that some hypothesis is true.

One of the earliest, perhaps the first (Raiffa, 1968), reference to the degree of belief definition is contained in J. Bernoulli's Ars Confectandi (1713). He suggested that probability is a "measure of the confidence" that an individual attaches to an uncertain event. Laplace expressed a similar view in A Philosophical Essay on Probabilities (1825) (Raiffa, 1968; Jeffreys, 1961).

Among modern statisticians, Keynes stated in his A Treatise on

Probability (1921) that "probability expresses the rational degree of belief that should hold logically between a set of propositions (as hypotheses) and another proposition (as conclusion)." Jeffreys (1961) adopted a similar definition, although he maintains that in practice the probability theory of Keynes is markedly different from his own (pp. v-vi). In 1933 Keynes modified his view of probability, as we will discuss later.

Degree-of-belief definitions implicitly assume that one can place probabilities on the state of nature: they define probability as the degree-of-belief that the state of nature concluded from the observations is the true one. Furthermore, the degree-of-belief definition leads directly to the so-called Bayesian or inverse probability rule of incorporating prior probabilities.

### Bayes' Theorem

The first serious attempt to rationalize the process of scientific inference as a means of gaining understanding into the operating of the real world can probably be traced to the Rev. Thomas Bayes (1702-1761) (Fisher, 1959), a Presbyterian Minister at Tunbridge Wells, and a Fellow of the Royal Society (Barnard, see References for data), who in his own time did not enjoy the standing he has been accorded in the past half century.

Bayes' most famous paper, "An essay towards solving a problem in the doctrine of chances," was not published until 1763, two years after his death. The paper was communicated to the Royal Society by

a Richard Price, a friend of Bayes, who added an introduction, and various examples and illustrations. The reason that Bayes did not publish the paper during his lifetime was probably because of the disputable nature of one of his prime postulates--one over which many words have passed since Bayes' time.

First we will consider the derivation (in modern notation) that Bayes offered, and then a second, simpler derivation common in modern textbooks.

Bayes' derivation is as follows: If  $p$  were the hypothetical probability of success in any given trial, the probability of observing  $a$  successes and  $b$  failures in  $a+b$  trials would be

$$\frac{(a+b)!}{a!b!} p^a (1-p)^b \quad 1$$

where  $(a+b)!/a!b!$  equals the number of combinations of  $(a+b)$  things taken  $a$  at a time. If we know that  $p$  itself has been chosen by some prior process such that the probability of  $p$  lying in any infinitesimal range  $dp$  between the limiting values of 0 and 1 is equal to the size of that range,  $dp$  (i.e., unif. dist 'd'); then the probability of the compound event of  $p$  lying in the assigned range, and of the observations  $a$  and  $b$  will be the product of the two independent probabilities,

$$\frac{(a+b)!}{a!b!} p^a (1-p)^b dp \quad 2$$

Such a compound event may only occur for those elements  $dp$  within the interval  $p = (0,1)$  (Figure 4), and the sum of all possible occurrences must by definition equal 1.0, or certainty. Therefore,

$$\frac{(a+b)!}{a!b!} \int_0^1 p^a(1-p)^b dp = 1 \quad 3$$

Multiplying the elemental probability by this normalizing factor gives

$$\frac{\frac{(a+b)!}{a!b!} p^a(1-p)^b dp}{\frac{(a+b)!}{a!b!} \int_0^1 p^a(1-p)^b dp}$$

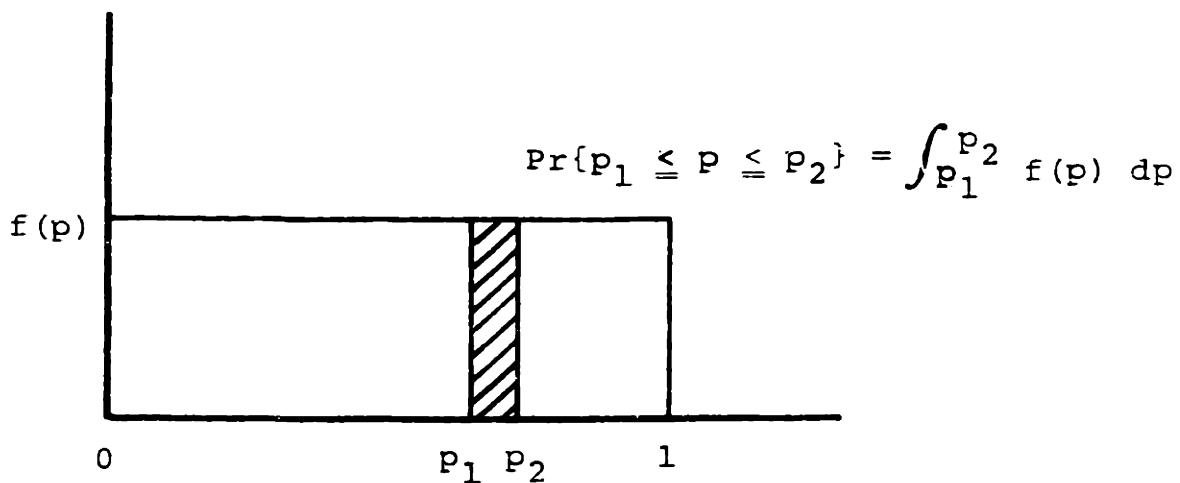


Figure 4

Since the denominator equals  $1/(a+b+1)$  (Fisher, 1970), the probability that  $p$  should lie between any assigned limits  $u$  and  $v$  (within the range 0 to 1) is

$$\frac{(a+b+1)!}{a!b!} \int_u^v p^a(1-p)^b dp \quad 4$$

The postulate which Bayes, along with many others, considered questionable was the assumption that the prior probability could be represented by  $dp$  (i.e., the uniform distribution). This will be discussed shortly (p. 25).

A somewhat simpler derivation, typical of those in modern textbooks, is the following: The probability of both events  $x_i$  and  $y$  occurring,  $f(x_i, y)$ , is (Figure 5)

$$f(x_i, y) = f(x_i) f(y|x_i) = f(y) f(x_i|y) \quad 5$$

so,

$$f(x_i|y) = \frac{f(x) f(y|x_i)}{f(y)} \quad 6$$

where  $f(x_i)$  is the probability of  $x_i$ ,  $f(y)$  is the probability of  $y$ ,

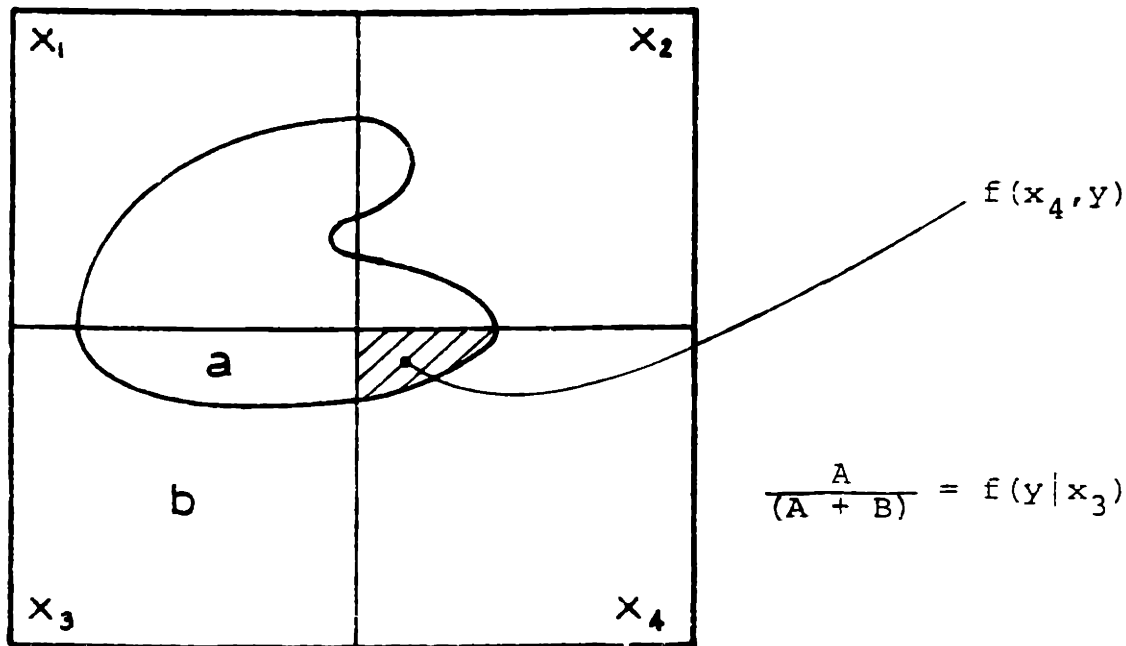


Figure 5

$f(x_i|y)$  is the probability of  $x_i$  given  $y$  and  $f(y|x_i)$  is the probability of  $y$  given  $x_i$ . The sum of the probability of both  $x_i$  and  $y$  occurring, taken over all possible values of  $x_i$  is just the probability of  $y$ , or,



$$f(y) = \sum_{i=1}^n f(x_i, y) = \sum_{i=1}^n f(x_i) f(y|x_i) \quad 7$$

placing the result of equation 7 into equation 6 gives,

$$f(x_i|y) = \frac{f(x_i) f(y|x_i)}{\sum_{i=1}^n f(x_i) f(y|x_i)} \quad 8$$

which is an incontrovertible statement of conditional probability, and is equivalent to Bayes' Rule.

The simple derivation of Bayes' theorem belies the extent of controversy which has raged over its application to practical problems.

Bayes' theorem is thought by most statisticians to be valid if cogent a priori knowledge is available of the distribution of  $x_i$ .

If a priori knowledge is not available, or if that knowledge is uncertain, then the spectrum of opinion on the proper role of Bayes' theorem goes from totally disregarding it (Fisher, 1961) to basing the prior distribution of  $x_i$  upon subjective "feelings" (Savage, 1954).

Fisher and his followers maintain that when cogent a priori information does not exist, the introduction of prior probabilities contains an arbitrary element, and must result in an arbitrary posterior probability assessment. Fisher's attitude toward the existence of valid prior distributions is summarized in the following excerpt from Statistical Methods and Scientific Inference (Fisher, 1959).

"A more important question, however, is whether in scientific research, and especially in the interpretation of experiments, there is cogent reason for inserting a corresponding expression representing probabilities a priori. This practical question cannot be answered peremptorily, or in general, for certainly cases can be found, or constructed, in which valid probabilities a priori exist, and can be deduced from the data. More frequently, however, and especially when the probabilities of contrasted scientific theories are in question, a candid examination of the data at the disposal of the scientist shows that nothing of the kind can be claimed."

The basic controversy over the definition of probability comes to light here because Fisher assumed a relative-frequency definition. Had he approached probability from the degree-of-belief definition, the existence of a prior probability would have followed immediately, even if the a priori data were less than conclusive.

An argument frequently voiced against using prior probabilities in cases with little good a priori data concerns the condition of total ignorance. In such a case the (non-subjectivist) degree-of-belief'ers would say to use the uniform distribution for prior probability (as Bayes did by assuming that the prior probability of  $p$  was equal to  $dp$ ). However, one might assert that in the example used by Bayes, we are a priori as equally ignorant of any arbitrarily chosen function of  $p$ , say  $p^2$ , as we are of  $p$  itself -- not that  $p^2$  necessarily has any physical meaning. So, in an equally rational manner we could use the uniform distribution for the prior probability of  $p^2$ , but if we do so we get different results. Using a uniform prior on  $p$  we get,

$$P(p) = dp \quad 9$$

$$P(p|A) \propto P(A|p) P(p) = P(A|p) dp \quad 10$$

Using a uniform prior on  $p^2$  we get,

$$P(p^2) = d(p^2) = p dp \quad 11$$

$$P(p|A) = P(p^2|A) \propto P(A|p^2) P(p^2) = P(A|p) p dp \quad 12$$

Clearly, the posterior probabilities expressed by 10 and 12 are not the same.

A (non-subjectivist) degree-of-belief'er would hold that the use of a prior distribution is the way real life scientific induction works, no matter how inconclusive is the a priori data. Furthermore, the definition of probability as a rational degree-of-belief leads directly to the use of uniform priors in the case of a priori ignorance -- "ignorance" means that each possible proposition is equally likely to the researcher.

What Bayes himself would have thought about these considerations is not certain. Bayes' definition of probability was (p. 376):

"The probability of any event is the ratio between the value at which an expectation depending on the happening of the event ought to be computed, and the value of the thing expected upon its happening."

In modern terminology, a definition based on expectation. One could infer that a definition based on expectation was the work of a degree-

of-belief'er, yet Fisher (1959, p. 14) claims otherwise:

"There is no room for doubt that Bayes would have regarded the 'expectation' to which he referred as capable of verification, to any required approximation, by repeated trials with sufficiently perfect apparatus. Subject to the latent stipulation of fair use, or of homogeneity in a series of tests, his definition is therefore equivalent to the limiting value of the relative frequency of success."

So the controversy continues.

### Degree-of-Belief'ers: Objectivists and Subjectivists

Within the degree-of-belief school there exists a group of statisticians who hold that there need be nothing logical about one's degree-of-belief, and that it need not follow necessarily from a priori information. Rather, they hold that one's degree-of-belief is a subjective entity, and is inseparable from his willingness to bet. These statisticians are known as "subjectivists," while those statisticians who hold probability to be a logical or necessary degree-of-belief are known as "objectivists."

A subjectivist would say that two people given the same a priori data may come up with different prior probabilities for the same event. This follows immediately from the fact that people's psychological make-ups differ from one to another. An objectivist would say that two people given the same a priori data, and following the same rules, would come up with the same prior probabilities. Information inadequately recorded, such as vague and half-forgotten empirical information, can be used in the subjectivist theory as a basis for

prior probabilities, but in the objectivist theory such information can only be used as suggestions of possible alternatives, and the prior probability used to express previous ignorance must still be used if no other information is available.

Example:

A dam is to be built in a V-shaped valley which may have formed either by differential weathering of a fault zone followed by erosion, or solely by stream action without differential weathering. Since seepage will be a problem if a fault exists in the bottom of the valley, the engineer would like to determine whether a fault does exist.

Two years ago the engineer worked on a job downstream of the proposed site at which he seems to remember a sizeable fault existing. What prior probability of the existence of a fault should he use?

A subjectivist would say that the prior probability of a fault existing is greater than 50% because the engineer sort-of-remembers a fault downstream and has a hunch that it exists at the site. The exact evaluation of the prior probability would be determined by comparison with artificial bets (e.g., "would you rather take a chance at winning \$10 conditional on rolling a 4, with an unbiased die, or on the fault existing?"). An objectivist would say that other than the vague information about a fault maybe existing downstream, the engineer is totally ignorant about the existence of a fault. Therefore, since existence and non-existence are equally likely, the correct prior probability is 50%.

The first uses of subjective probability with special reference to statistics (Ramsey, de Finette, and Koopman approached this topic from the standpoint of logicians) were made by Good (1950) and Savage (1954). Although the latter's book is often considered the primary reference on subjective probability, it is an early treatment of a theory which has been expanded greatly since 1954. In 1959 Schlaifer published the first comprehensive elementary textbook written entirely from the subjectivist Bayesian point of view. This book has itself become a minor classic for problems of application.

As one might guess, considerable arguments transpire over the use of subjective probabilities. Objectivists (and relative-frequentists also) argue that by its very nature statistical inference must be objective; allowing subjective beliefs to enter the analysis destroys this objectivity, and allows any arbitrary probability assessment to be the result of analysis. Subjectivists counter that there is no way to obtain the objectivists's unambiguous, universally acceptable value of degree-of-belief and that objectivists and relative frequentists actually allow more subjectivity to enter their analysis than do the subjectivists themselves. Experimental design is accomplished almost entirely by "judgement," and once the time for action comes objectivists and relative frequentists advise using "judgement" in acting upon the analytical results. Savage's comment is, "the (subjective) Bayesian approach is more objectivist than the frequentist in that it imposes a greater order on the subjective elements of the deciding process" (Savage, 1961).

One might comment, however, that the treatment of subjectivity by objectivists and relative frequentists causes the presence of subjectivity to be fully apparent at the moment of decision. The subjectivist's methods deal with subjectivity before the point of action, and may allow someone not fully cognizant of the limitations of the theory to place undue confidence in the analysis.

### Section 1.2 Decision

In the first section of this chapter we considered the definition of the primitive term probability, and saw that several schools of thought exist. In this section we will consider applying statistics to making decisions, and we will once again see that the several schools approach the problem differently.

Raiffa (1968) divides the practical use of statistics into three branches: data analysis, inference, and decision.

Data Analysis treats large masses of data which may be considered as composing the entire population in itself. The statistician in treating such data is not concerned with making inferences upon the data, but with bringing order to it. His job is to evaluate interrelationships, and to "explain" the data. An example of data analysis is the treatment of the national census.

Inference treats problems associated with increasing scientific knowledge. Inference attempts to "back-figure" the probabilistic mechanism responsible for certain observations from the observations, and the results may be used by persons unknown to, and for decisions

not anticipated by the scientist (Wallis and Roberts, 1956). An example of inference from data is determining whether rock strength is a function of effective stress or of total stress.

Decision treats problems of practical action such as deciding which of two uncertain remedial measures to take in the event of a leaky dam foundation. In decision problems the alternatives being considered can be listed and the consequence of taking each action evaluated as a function of the true state of nature.

Suppose that two remedial measures were available to the engineer, chemical grouting and emptying the reservoir and placing a clay blanket; and that the source of the trouble could be either solution cavities or a weathered fault zone.

If the chemical grouting were used as a remedy and the source of the leakiness were solution cavities, then there would be, say, a probability of 0.9 that the leak would be stopped. However, if the trouble was a weathered fault zone, there would be only, say, a probability of 0.3 that the leak would be stopped. Now on the other hand, if the reservoir were drained and a clay blanket placed, there would be, say, a probability of 0.95 that the leakage would stop no matter what the cause. The second action increases the probability of stopping the leak, but it is also considerably more expensive.

#### Section 1.2.1 Inference

The tree of Figure 6 illustrates the relationship between the



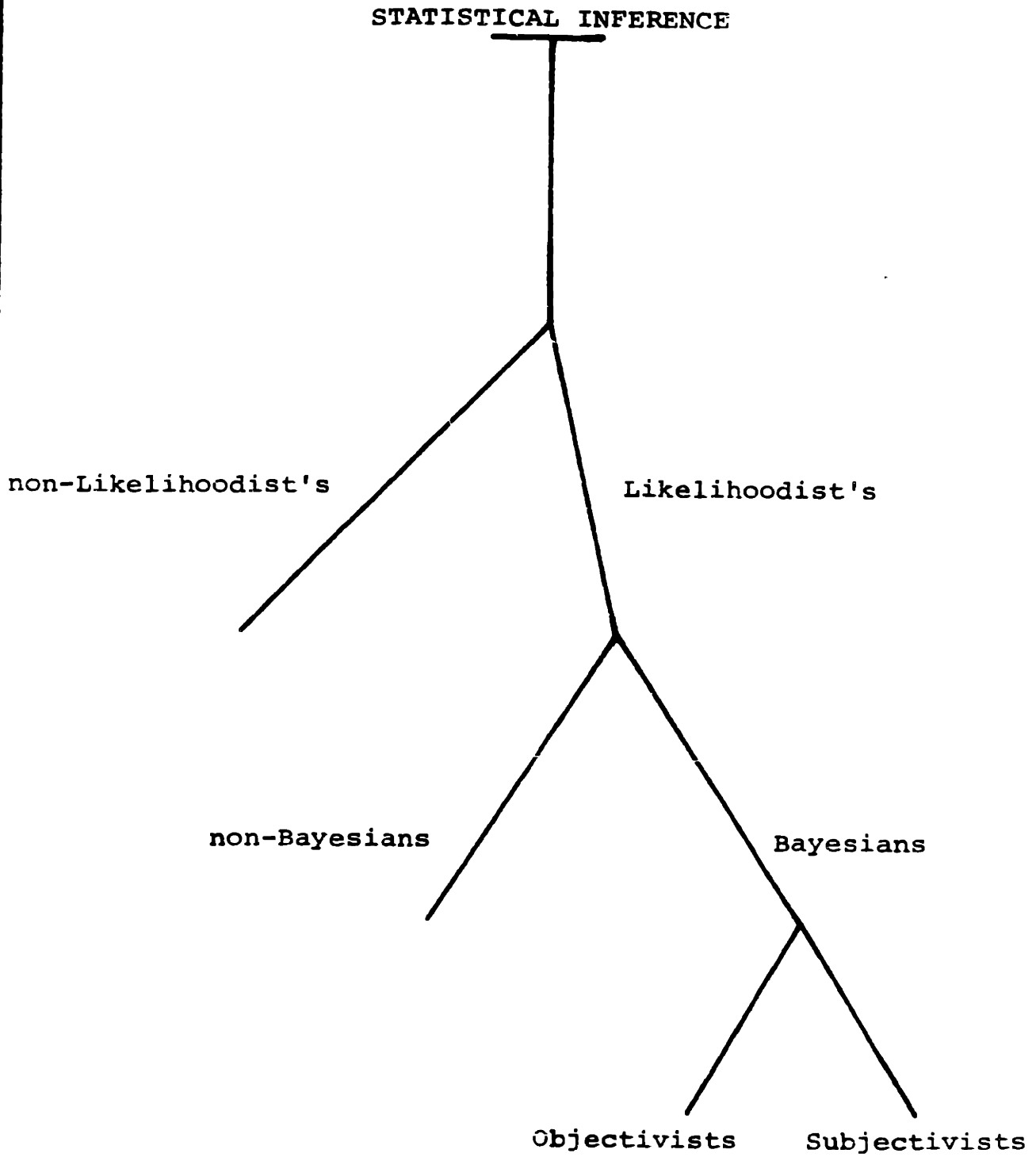


Figure 6 (after Raiffa, 1968)

major schools of statistical inference; we will note as we progress that the divisions are primarily based on the definition of probability adopted by each.

The initial division occurs between those statisticians who believe that any information about an experiment and its outcome other than the likelihood function (i.e. the probability of the data actually observed conditioned on various states of nature) is irrelevant for inference, and those who believe that the experimental frame of reference (within which the outcomes are observed) cannot be ignored. Continuing a series of experiments until one has collected enough evidence to support a particular hypothesis and then stopping is perfectly acceptable to a likelihoodist. A non-likelihoodist would argue that by stopping an experimental series when there was just sufficient evidence one could prove any hypothesis: The experimental frame is needed to supply the conditional probability assessments for other possible sample outcomes, which, in fact, have not occurred.

Consider an engineer who is trying to decide whether 80% of the foundation material for a new dam has a permeability less than  $10^{-5}$  cm/sec. If he runs several tests and selects that part of the data which is favorable to the hypothesis, he is guilty of data selection which no statistician would condone. If, however, he tests specimens until he has built up a favorable case and presents all of his data in the analysis, a likelihoodist would say that this was legitimate because the criterion for stopping is irrelevant to

the analysis. A non-likelihoodist would say that this was flagrantly illegitimate because the test series was carried out in a manner which was biased in favor of the hypothesis. He would hold that in order to gain meaningful results the number of specimens, and the necessary level of significance would have to be determined before the start of the experiments.

Among likelihoodists there are two groups: "Bayesians," and "non-Bayesians." "Non-Bayesians" are frequentists; they would maintain that the state of nature may not have a probability placed upon it -- either 80% of the material has a permeability less than  $10^{-5}$  cm/sec, or it doesn't.\* A non-Bayesian would set confidence limits, which would tell him how often his experiments would result in an estimate of the permeability within, say, 5% of the true value. "Bayesians" are degree-of-belief'ers. They would assess prior probabilities on the proportion of material having  $k \leq 10^{-5}$ , and use observations to compute posterior probabilities.

### Section 1.2.2. Decision

The range of thought on how decisions should be made is as wide as that on the definition of probability, and indeed the forces line up in similar schools. Raiffa has once again organized the basic

---

\*The title "non-Bayesian" is somewhat misleading because "non-Bayesians" do accept the validity of Bayes' Theorem as a statement of conditional probability (i.e., in the case of cogent prior information).

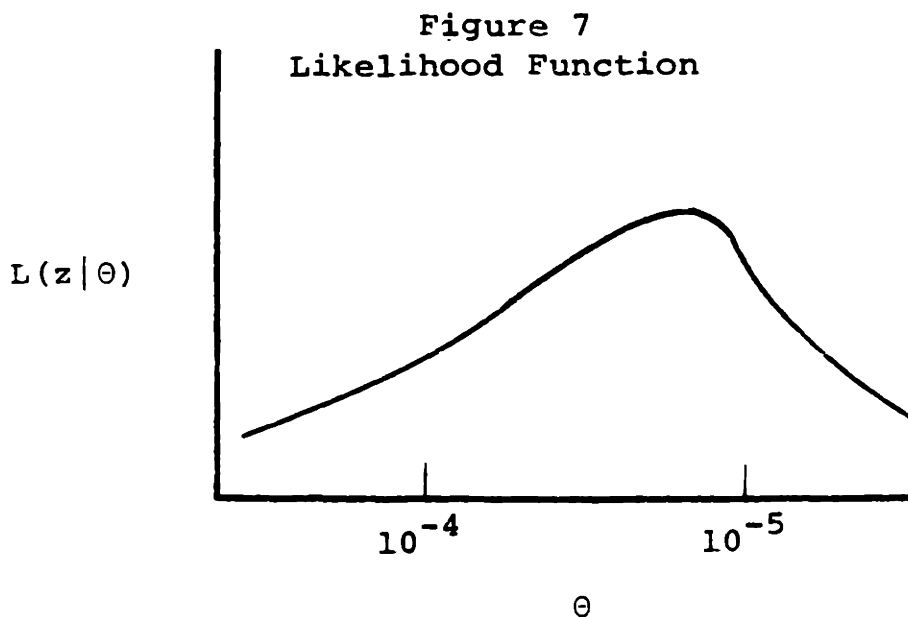
philosophies, and we shall use his classifications as a starting point. Raiffa divides the philosophies into four camps which he names by a prominent advocate(s) in each: Fisher, Neyman-Pearson, Wald and Savage.

The Fisherian school uses statistics to evaluate hypotheses on the state of nature while relegating decisions on action to judgment, this being outside the realm of statistics. Fisherians are primarily concerned with testing null hypotheses and the probability of rejecting the null hypothesis when it is really true. Fisherians suggest suspending judgement until there is enough hard experimental evidence that one cannot explain, beyond a reasonable level of doubt, except by concluding that the conjecture under consideration is correct. Decisions are made on the basis of the maximum of the likelihood function, which is the conditional probability of the observations given various states of nature. Therefore, probabilities are never placed on the state of nature.

An engineer, in designing a dam on a pervious foundation, must decide upon which of two foundation treatments to use for the control of leakage. His choice depends on the permeability of the foundation rock which he has attempted to measure by in-situ and laboratory testing. His analysis of the experimental data suggests that the permeability may be either about  $10^{-4}$  or about  $10^{-5}$  cm/sec. If he knew that the permeability was really  $10^{-4}$  cm/sec he would elect to use treatment #1 which is the more expensive and thorough method. If he knew that the permeability was really  $10^{-5}$  he would elect to use

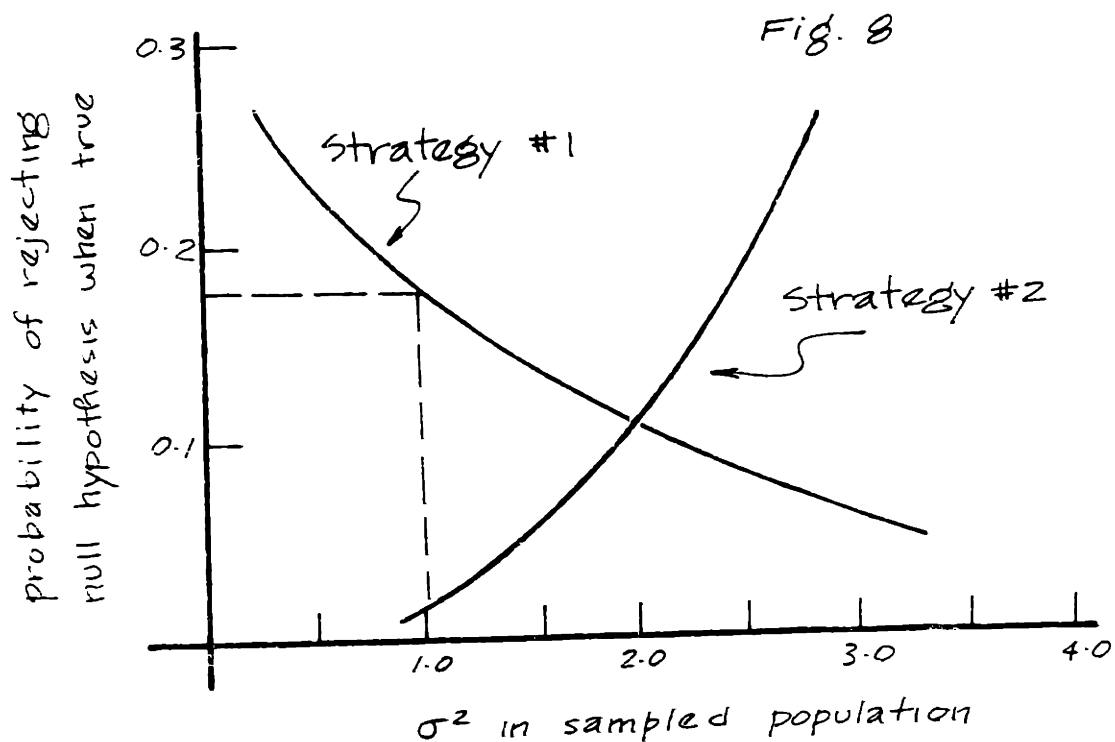
treatment #2 which is less expensive and thorough than treatment 1, but adequate for a permeability of  $10^{-5}$  cm/sec. Treatment 2 would not be adequate if the permeability was  $10^{-4}$  cm/sec, making expensive remedial action necessary. Which treatment should the engineer choose?

A Fisherian would proceed by evaluating the likelihood function of the experimental observations  $\underline{z}$  (Figure 7), which for any given state of nature  $\theta$  equals the conditional probability of observing  $\underline{z}$ . He would then conclude



that the best estimate of the real state of nature is that state which maximizes the likelihood function  $L(z|\theta)$ . The actual decision of which treatment to use, however, must be left to "judgement," since this is not part of statistics.

Neyman and Pearson hold (1933) that experimentation and accepting or rejecting the null hypothesis are action problems. When considering alternate strategies for accepting or rejecting hypotheses, each having the same probability of type I error (i.e., the same probability of rejecting the "null" hypothesis that two populations are the same when actually the hypothesis is true), one should favor those strategies with the lowest probability of a type II error (i.e., the probability of rejecting a difference when it actually does exist). The tool with which this comparison is made is called the "power function" of a strategy (or sometimes the "error characteristic"), and is a function which describes the conditional probability of a type II error (given the state of nature) resulting from the use of that strategy.



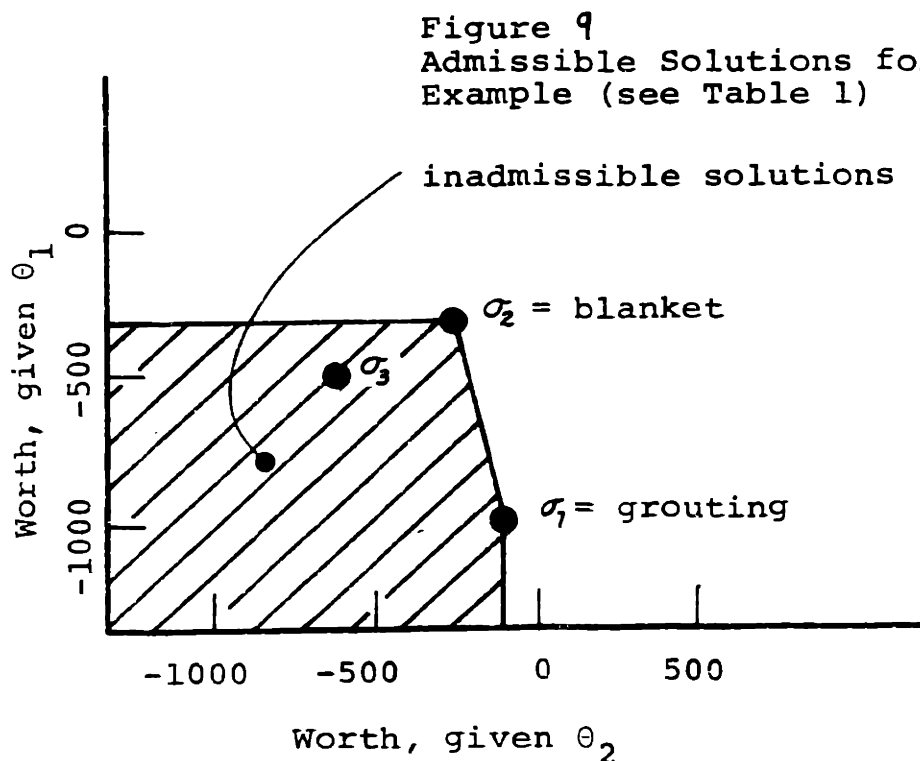
In Figure 8 are shown the power functions of two strategies for testing the hypothesis that there is no difference between the population sampled and a population having an average permeability of  $10^{-5}$ . Both strategies have the same probability of rejecting this hypothesis when it is true (type I error), and the power functions describe the probability of accepting the hypothesis when it is false (type II error). As indicated by the figure, we have assumed this latter probability to depend on the true variance of the population sampled. If the true variance was  $1.0 \times 10^{-5}$ , strategy #1 would have a probability of a type II error of 0.17, and strategy #2, 0.01. The choice of which strategy to use depends on the range within which the true variance is thought to lie. For  $0.5 \times 10^{-5} \leq \sigma^2 \leq 1.5 \times 10^{-5}$ , strategy #1 is preferred; for  $2.0 \times 10^{-5} \leq \sigma^2 \leq 2.5 \times 10^{-5}$ , strategy #2 is preferred.

The "best" strategy is the one which has the universally lowest power function. The difficulty with such a criterion is that one can usually obtain better error control over the entire parameter space only by increasing the cost of experimentation. So, we are faced with deciding between hypotheses whose "efficiency" depends on the region of the state space (i.e., the region which includes the actual state of nature). In deciding between these strategies one is advised to keep in mind the cost of experimentation, the relative seriousness of making different types of errors, and the decision maker's guess of the true state of nature. Since all probability assessments are conditional, one never assigns probabilities to the

state of nature, and one must make decisions, such as whether to grout or place a clay blanket for seepage control, purely through judgement.

Wald (1950) expanded the solution to the decision problem by adding the concepts of cost and worth of consequences into his formulation. As in the Neyman-Pearson and Fisherian schools, he did not place probabilities on the state of nature, but in arguing the worth of consequences he reduced the choices among which a decision must be made.

In deciding on a course of action, Wald would first evaluate the worth of the proposed strategies under each possible state of nature ( $\theta_i$ ). Figure 9 is a plot of the worth of strategy 1 ( $\sigma_1$  = use chemical grouting) and strategy 2 ( $\sigma_2$  = place clay blanket), given the true state of nature ( $\theta_1$  = fault zone;  $\theta_2$  = solution cavities).





Wald defines as "admissible solutions" those strategies for which no other strategy exists with a greater worth under all possible states of nature. For example, the hypothetical strategy  $\sigma_3$  would not be an admissible solution since  $\sigma_2$  has a greater worth no matter the true state of nature (both  $\sigma_1$  and  $\sigma_2$  are admissible solutions). The decision therefore is reduced to deciding among admissible solutions.

The criterion for deciding among admissible solutions is less clear; several have been proposed. Wald proposed the "minimax" rule which says to select the strategy having the least maximum loss. In Table 1 are shown losses (in arbitrary units of money) associated with the seepage example. The minimax rule specifies to choose a clay

Table 1

Strategy \ State of Nature:	Soln. Cavities	Fault
Grouting	-1000	-100
Clay Blanket	- 300	-300

blanket since its maximum loss (-300) is less (in absolute value) than the maximum loss associated with grouting (-1000).

Other criteria have been proposed, and in describing them we will let  $w_{ij}$  be the worth of electing strategy  $\sigma_i$  when the true state of nature is  $\theta_j$ .

**Maximax:** Maximizing the largest gain (or minimizing the minimum loss) associated with an action.

$$\max_i \max_j (w_{ij} | \theta_j)$$

**Hurwicz:** Maximizing a modified sum of the maximum gain and maximum loss of the strategy, using a constant called the "degree of optimism" ( $\alpha : 0 \leq \alpha \leq 1$ )

$$\max_i \{ \alpha [\max_j (w_{ij} | \theta_j)] + (1 - \alpha) [\min_j (w_{ij} | \theta_j)] \}$$

**Minimax Regret:** Minimizing the largest difference (largest over the possible states of nature) between the obtained and the most valuable consequence (i.e., the "regret")

$$\min_i \max_j (w_{ij} - \max_j w_{ij})$$

**Laplace:** Maximizing the arithmetic average of the consequences.

$$\max_i \left\{ 1/n \sum_{j=1}^n w_{ij} \right\}$$

Although the rationality of restricting consideration to admissible solutions cannot be questioned, the choice of secondary strategy may not be consistent with the decision maker's preferences and aversion to risk. If, in the present example, the probability of  $k$  equalling  $10^{-4}$  was very small (say less than 1%), the minimax criterion would still dictate choosing strategy  $\sigma_2$ . Actually, the engineer might be willing to risk this small probability of a large loss to save the 200 unit cost difference between treatments 1 and 2. In such a

situation the engineer's preferences and the dictates of the decision criterion would not be compatible!\*

The concept that probability and action are inseparable was introduced by Ramsey (1926) and leads to the use of subjective probability and "utility" in decision making. Utility might be called subjective worth, differing from direct monetary value by its inclusion of personal values and intangibles.

In his book, The Foundations of Statistics, Savage develops a theory of idealized behavior under uncertainty predicated upon the personalistic view of probability (i.e., that of Ramsey) and upon the theory of utility developed by von Neumann and Morgenstern (1947).

The value of a given monetary reward is not the same to all individuals: a wealthy person might view a \$1000 loss as minor while a poor person might view it as considerable. Furthermore, to most people the incremental value of a \$100 gain on top of a \$10,000 gain is minor compared to the incremental value of a \$100 gain on top of \$10; and certainly intangibles such as loss of professional reputation are part of the value of certain consequences (e.g., dam failure).

Beginning from a set of axioms describing what is meant by a consistent set of preferences (utilities), Savage shows that an ideally rational person should make decisions on the basis of expected utility. So, having already entered his subjective or

---

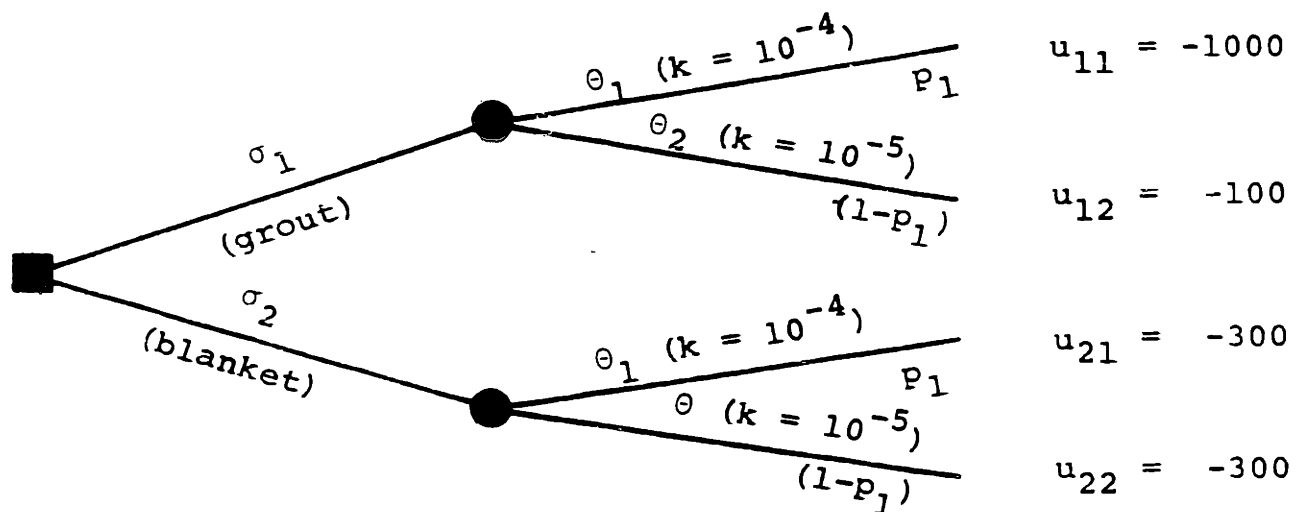
\*The difficulty of this particular example could be avoided by using Hurwicz's  $\alpha$ , but the fact remains that the secondary criterion may not reflect the engineer's true preferences.

"judgemental" feelings about the decision into the analytical formulation, the subjectivist would evaluate the expected utility of taking each of several actions and elect the action whose expected utility was the largest.

The problem of which foundation treatment to use would be approached by drawing a "decision tree" as shown in Figure 10, in which each branch represents a possible alternative of either decision or "chance" (i.e., a state of nature), a square representing a decision by the engineer, and a circle representing a "chance" occurrence by nature (one over which the engineer has no influence). The value at the tip of each branch at the extreme right of the diagram is the value of each outcome in utiles. Here, for the sake of simplicity, the values of monetary worth from Table 1 have been assumed to be values of utility.

Figure 10

Decision Tree for Seepage Problem



The expected utilities of taking actions  $\sigma_1$  and  $\sigma_2$  are:

$$E(u(\sigma_1)) = p(\theta_1)(-1000) + (1-p(\theta_1))(-100) = -100 - 900 p(\theta_1)$$

$$E(u(\sigma_2)) = p(\theta_1)(-300) + (1-p(\theta_1))(-300) = -300$$

For values of  $p(\theta_1) \leq 0.22$ ,  $E(u(\sigma_1)) \geq E(u(\sigma_2))$ ; and for values of  $p(\theta_1) \geq 0.22$ ,  $E(u(\sigma_1)) \leq E(u(\sigma_2))$ . So,

Choose  $\sigma_1$  if  $p(\theta_1) < 0.22$

Choose  $\sigma_2$  if  $p(\theta_1) > 0.22$

Indifferent if  $p(\theta_1) = 0.22$

The subjectivist approach to decision is frequently called "Bayesian Decision Theory," and has seen rapid development over the last 15 years.

### Section 1.2.3 Comment

The spectrum of decision models is wide and reflects philosophical differences in the nature of the primitive term probability. Despite the distinct philosophical differences between approaches to decision, in general we needn't be constrained to apply but one set of statistical tools in solving problems of site exploration (a comment sure to earn damnation from philosophers-of-science).

This survey has been brief, but should serve as a frame of reference within which the work of this thesis may be viewed. Books and articles elaborating these topics are to be found in the reference list.

## CHAPTER 2

SUBJECTIVE PROBABILITY ASSESSMENT

In Chapter 1 we reviewed the philosophical meaning of the primitive term probability, and demonstrated the impact of this meaning on decision. In this chapter we will discuss the task of assessing subjective probabilities.

Section 2.1 Theory and Philosophy of Subjective Probability Assessment

The Ramsey-deFinetti-Savage theory of subjective probability is based on the postulate that degree-of-belief and action are inseparable: subjective probabilities may be inferred from behavior, but they are not necessarily intuitive even to the decision maker. As a consequence, directly asking a subject what he judges a probability to be does not necessarily result in an evaluation compatible with his true "feelings" (as might be reflected in a decision situation); so to assess subjective probabilities one needs an "action" approach in which probabilities are inferred from the subject's behavior in controlled situations.

"In the first place, many doubt that the concept 'more probable to me than' is an intuitive one.... If the state of mind in question is not capable of manifesting itself in some sort of extraverbal behavior, it is extraneous to our main interest. If, on the other hand, it does manifest itself through some more material behavior, that should, at least

in principle, imply the possibility of testing whether a person holds one event to be more probable than another, by some behavior expressing, and giving meaning to, his judgement."

(Savage, 1954)

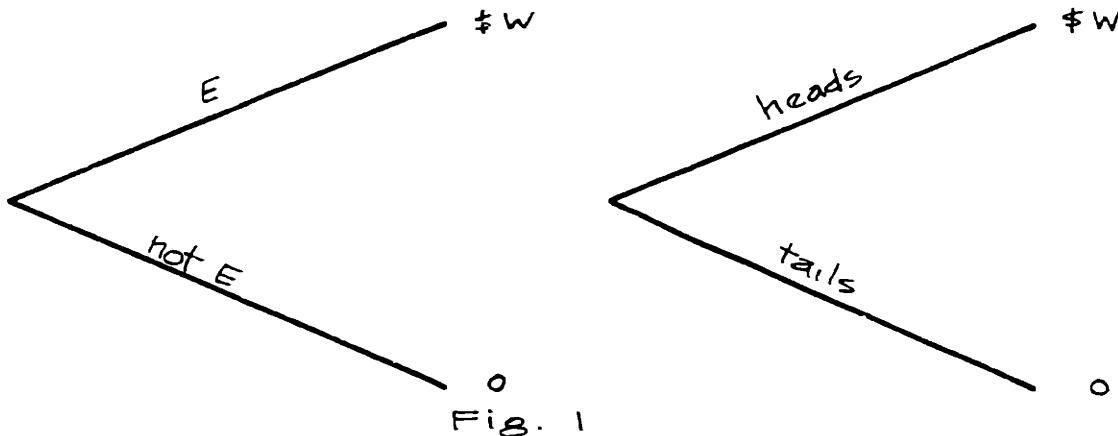
Succinctly, a subject's true feelings can be appraised only if he "has something at stake" in their assessment. Intuitive approaches (i.e., those asking the subject what he thinks a probability to be) make no attempt to foster careful weighting of judgemental feelings, and make no attempt to dissociate events from their consequences (e.g., by not isolating events from their consequences intuitive approaches lead an optimistic subject to overestimate the probability of an event having a favorable consequence, and vice versa). Furthermore, intuitive approaches make use of specific numerical scales (e.g., "the probability is 90%") which may be suggestive of things not intended (Fishburn, 1964).

An action approach to assessment presents the subject with decisions. It places before him gambles involving the probabilities to be assessed, in which he must choose one bet or another. Subjective probabilities are inferred from his decisions.

In practice one finds the intuitive approach used frequently -- despite the behavioral emphasis of the philosophical bases of subjectivism. This approach is easier and quicker to use than an action approach, but empirical evidence suggests that it does not lead to assessments which reflect the subject's true feelings (e.g., Phillips and Edwards, 1966; Luce and Suppes, 1965).

## Section 2.2 Action Approach to Assessment

In assessing probabilities through an action oriented approach, one asks a subject to express his preference between pairs of gambles: one dealing with real events and one dealing with simple "standard" events (such as flipping a coin or drawing a ball from an urn). One infers subjective probabilities by comparing the expected values of the gambles. If the subject prefers a gamble A to a gamble B, then one infers that the subject deems the expected value of A to be greater than the expected value of B. Since expected values are expressed in terms of the probabilities of success and failure, bounds can be determined on the desired subjective probability. As an example, consider the two gambles,



in which E is a real world event whose subjective probability we desire. The expected utility of gamble (a), assuming utility linearly related to monetary value, is<sup>\*</sup>

---

<sup>\*</sup>One must be cautious in constructing hypothetical gambles, for in



$$\begin{aligned} E(u_a) &= p(E)(w) + (1-p(E))(0) \\ &= p(E)(w) \end{aligned}$$

The expected utility of gamble (b) is,

$$\begin{aligned} E(u_b) &= p(\text{heads})(w) + (1-p(\text{heads}))(0) \\ &= (1/2)w \end{aligned}$$

If the subject prefers gamble (a) to (b), then  $E(u_a) \geq E(u_b)$  and  $p(E) \geq 1/2$ ; if he prefers (b) to (a), then  $E(u_b) \geq E(u_a)$  and  $p(E) \leq 1/2$ . By presenting a series of choices between gambles, these bounds may be drawn together and the subjective probability of E inferred.

In practice the use of real gambles is generally infeasible (it can be expensive, anger provoking, and in some places even illegal). Pragmatically, we are forced to use a compromise method, sacrificing some rigor to gain economy. One such method is that of "hypothetical gambles" -- a scheme of questioning which asks a subject what he would do in hypothetical situations involving risk (i.e., he is given a hypothetical choice between gambles). Whether this is a valid procedure or merely a restructuring of the intuitive approach, no one knows; experimental evidence is not available

---

addition to subjective probability the subject's choices also depend on his preference for rewards (i.e., his utility function). Utility functions are usually non-linear, so rewards must be limited in magnitude to a region in which preference is a linear function of worth; otherwise, assessed probabilities depend on empirically assessed utility functions.

comparing probabilities assessed by the two methods. The method of hypothetical gambles is widely used and is possibly the most satisfactory one at present, although assurances from experimental work would be comforting.

### Section 2.2.1 Procedural Effects

Probability assessment borders on experimental psychology, and as such is influenced by procedural details. For example, what effect does the precise wording of questions have on assessed probabilities? In current practice such details are often uncontrolled, and in fact, there are even few investigations of them appearing in the literature.\* The state-of-the-art of probability assessment is such that even questions as basic as, "do hypothetical gambles and real gambles result in the same assessments," are yet to be answered. In the future attention must be given to the importance of these details if only to obtain estimates of the variations they lead to in assessed probabilities. Here, we will mention three procedural details which, with our present state of understanding, seem to exert influences on assessments. Obviously, there are many others.

Subjective probabilities may change with time either because

---

\* Coomb and his associates have considered methodological problems in probability assessment as part of their investigations into psychological scaling in general (e.g., see Coomb and Beardsley, 1954); and more recently, Winkler has written a Ph.D. thesis on probability assessment which considers certain methodological questions (see Winkler, 1967a,b).

additional information comes to the subject's attention or because he reflects further on the problem. This is acceptable, because if they did not change with new information and reflection, the entire concept of subjective probability would not be appropriate for decision. Assessments should be made as near in time to implementation as is possible and convenient.

The "mental frame" of a subject can affect assessed probabilities, but in ways we are not sure of (Fishburn, 1964). For example, probabilities assessed immediately after a subject has returned from an extended business trip may differ from those assessed during a normal week. Therefore, one should try to achieve "standard" conditions for assessment; however, since we know so little of procedural effects, for now "standard" can only be defined intuitively.

The exact sequence of assessing the attributes of a probability distribution (e.g., mean, standard deviation, range) may also affect resulting subjective probabilities, since current learning theory tells us that subjects' responses to questions are directly influenced by their responses to previous ones (the closest in time having the greatest effect). Again, however, there simply has not been sufficient work performed in investigating these effects to draw substantial conclusions. In Section 2.4.2 we will discuss a specific example of a problem in assessing probabilities which may be influenced by the sequence of assessment.

### Section 2.2.2 Descriptive vs. Prescriptive Models

To infer probabilities from choices we require a descriptive model of behavior under uncertainty: a model which tells us how people behave. Bayesian decision theory is a normative model: it prescribes how people should behave. These are not the same.

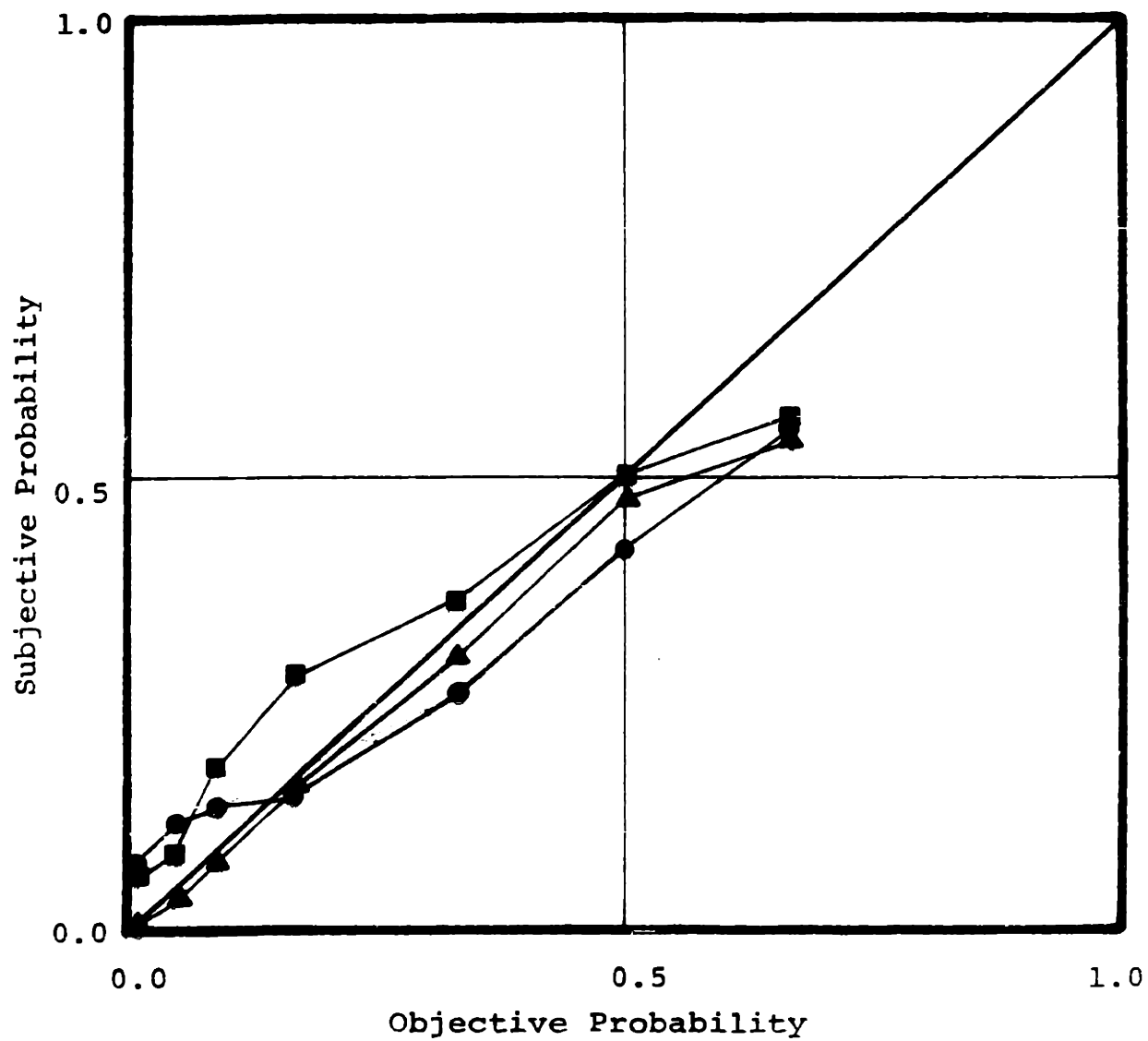
As Ellsberg (1961) and Fellner (1961) point out, cases can easily be constructed in which subjects do not wish to make the choices prescribed by Bayesian decision theory (i.e., those which maximize expected utility) -- even after the discrepancy between their behavior and that of an ostensibly "rational" person is pointed out. One can find ample evidence of this phenomenon in the literature of experimental psychology (e.g., Tversky, 1967; Luce and Suppes, 1965), which simply shows empirically that Bayesian decision theory is not a descriptive model. The consequence of the distinction between the way people actually behave and the way they "should" behave is that probabilities inferred using Bayesian decision theory may be incorrect. Nevertheless, since we lack a descriptive model, common practice is to assume that Bayesian decision theory is descriptive for simple gambles with low rewards.

One can begin to evaluate systematic errors in predictions of behavior made by Bayesian decision theory, and hence of systematic errors made in inferring subjective probabilities, by comparing the subjective probabilities inferred from subjects' willingness to accept gambles,  $s(p)$ , with the known objective probabilities of

success in those same gambles, p.

The results of three studies in which subjects were asked to bid for the privilege of accepting gambles involving known probabilities of success (e.g., "how much will you pay for a 75% chance at \$10?"), are shown in Figure 2. Although the results are different in detail, each set of subjects tended to overestimate low probabilities and underestimate high probabilities. In other words, subjects were willing to pay more than the expected value for gambles involving low probabilities of success, but less than the expected value for gambles involving high probabilities of success. Furthermore, these differences between objective and subjective probability increase as the number of people competitively bidding increases. Similar results have been found in other studies (e.g., Attneave, 1953; Dale, 1958, 1959; Sprowls, 1953). Aside from overestimation of low probabilities and underestimation of high probabilities, the indifference point at which  $s(p) = p$  has also been found to be consistent across studies. Most researchers have found indifference points to lie in the range 0.15 to 0.25 (although Mosteller and Noguee's data of Figure 2 indicate higher values).

The studies just discussed treated bidding experiments in which the objective probabilities associated with gambles were known to the subject at the start of the experiment. If we turn our attention to studies in which these probabilities were not known a priori, we still see consistent behavior among subjects and over separate studies, but the trends of behavior in this case are



- Preston and Baratta,
- ▲ Mosteller and Nogee,
- Mosteller and Nogee,

Figure 2

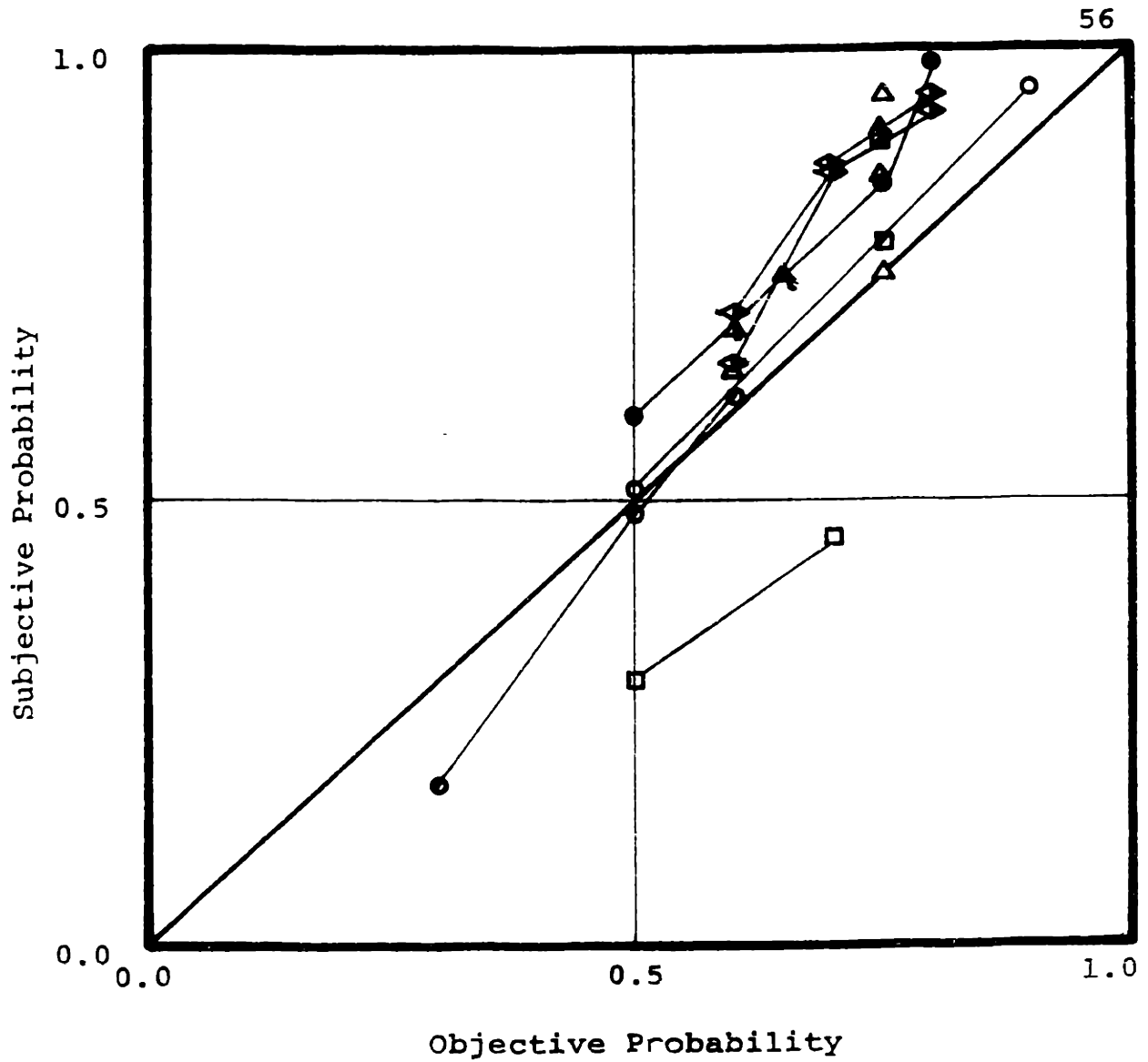
Subjective probabilities inferred from  
bidding experiments

reversed from those of Figure 2.

Studies in learning probabilities by repeated trials are usually conducted as follows: A subject is seated before two lamps and told that one of them will light in each trial; he will be asked to predict which one. If his prediction is correct he wins a prize; if his prediction is incorrect, he wins nothing. At several points in the experiment his subjective probabilities are assessed.

During early trials a subject's feelings about the relative likelihood of each lamp lighting change as he observes more and more trials, but eventually (usually after several hundred trials) his subjective probabilities become stable and are said to reach their asymptotic values,  $p_{\infty}$ . One might suggest that this is the same process through which people gain a "feel" for probabilities associated with common events -- like drawing a face card from a normal deck -- rather than by directly considering the numerical probability.

Results of several studies on learned probabilities are shown in Figure 3. Here asymptotic probability is plotted against objective probability. Note, that instead of under-estimating high probabilities, subjects consistently over-estimated them. An explanation should exist for the difference between the trends of Figure 2 and those of Figure 3, because the trends are consistent over many subjects and many experimenters, but no adequate explanation now exists. One suggestion has been that subjective probabilities



- Goodnow, 1955
- ◉ Edwards, 1956
- " " " "
- △ Siegel and Goldstein, 1959
- ▲ Suppes and Atkinson, 1960
- ▴ Siegel and Abelson, 1961
- Edwards, 1956
- ▣ Galanter and Smith, 1958
- Nicks, 1959
- ◊ Myers, et al, 1963

Figure 3  
Asymptotic Subjective Probabilities



depend on the sign of rewards (i.e., whether a subject can win/break-even or lose/break-even), even when account is taken of differences in the utility function for gains and losses; Irwin (1953) and Edwards (1955) have reported empirical evidence of such behavior. What this phenomenon would mean for the studies of Figure 2, which included both positive and negative rewards, is not clear.

Aside from differences between objective probabilities and their subjective interpretation, subjects have been reported to prefer certain probabilities and outcomes over others of equal expected value -- a phenomenon certainly not predicted by Bayesian decision theory. Edwards (1953, 1956) found that for gambles whose expected values were positive subjects preferred those involving the probability  $1/2$  and avoided those involving the probability  $1/3$ ; when the expected values were negative, he found precisely the opposite. Davidson, et al (1957) found in using simple devices like coins and dice, that subjects preferred specific outcomes over others (e.g., other things equal, subjects preferred winning-on-heads/losing-on-tails to winning-on-tails/losing-on-heads). Since these effects are not accounted for in the Bayesian model, they lead to errors in assessed probabilities.

As we have seen, the investigations of behavioral models for decisions under uncertainty are neither numerous, extensive, nor particularly conclusive. The relationship between objective probability and those inferred using Bayesian decision theory is complex and cannot be summarized in a few simple statements or formulae.

Given certain conditions,  $p$  and  $s(p)$  tend to coincide, and given others they diverge; but the relationship between them seems to be systematic and consistent over many subjects and experimenters. Subjective probabilities are influenced by the way questions are worded, the probabilities chosen for alternate gambles, and the rewards offered. Cohen and Hansel (1955) even present evidence that the relationship between objective probability and that inferred using the Bayesian model is dependent on age. However, despite Bayesian decision theory being only a normative model, in order to evaluate subjective probabilities we must assume it also to be a descriptive model. Available evidence shows that the validity of this assumption is open to question, despite its acceptance by many workers. The intention of this section has only been to point out that assuming subjects to behave as predicted by the Bayesian model at best leads to rough approximations of subjective probabilities.

### Section 2.2.3 Ambiguity and Stability

Were one asked to assess the probability of heads for any flip of a certain coin, with little hesitation he would answer "1/2." As individuals we have considerable experience with this uncertain quantity, and even though no coin is perfectly unbiased, the range of probabilities we expect is small: We feel little ambiguity. On the other hand, were one asked to assess the probability of the Dow Jones average increasing by more than a full point in a single day, after some reflection he might also answer "1/2," but the range of

probabilities we might expect is wide: We feel great ambiguity. In both cases the best estimate was the same; the assessments are different because the levels of ambiguity are different.

Ambiguity in a subjective probability manifests in the ease with which assessments are changed by additional information. If new information causes large changes one says that the probability assessment is unstable; if new information causes only small changes one says that it is stable.

Individuals are often reticent to quantify feelings when they feel great ambiguity; Ellsberg's and Fellner's work documents this phenomenon. This reticence is exhibited in the common objection from engineers that subjective probabilities are meaningless when experience is limited or the quality of a priori data is poor.\* Objections might subside, however, were assessment explicitly described as consisting of two parts, a "best estimate" and a measure of stability.

Stability is quantitatively expressed by the spread of assessed distributions. For predictions of the state of nature (e.g., average foundation permeability) this can be measured by the standard deviation of the assessment (e.g., "the average permeability is

---

\*Stability of assessments obviously depends on data quality and discriminability, but the relationship is not clear. Stability is a judgemental attribute, and judgement may "smooth" over deficiencies by adding intuitive feel to sparse information. This was a fundamental reason for using subjective methods to begin with; our "feel" for certain UQ's may excuse the quality of prior statistical information.

probably  $10^{-2} \pm 5 \times 10^{-3}$  cm/sec"); and for random variable UQ's by the standard deviation of the estimates of distribution parameters (e.g., "the average permeability of this type formation is probably  $10^{-2} \pm 5 \times 10^{-3}$  cm/sec, and the standard deviation is probably  $10^{-3} \pm 10^{-4}$  cm/sec"). Figure 4a shows two distributions for average permeability; both indicate  $10^{-2}$  as the most likely value, but assessment 2 is less stable than assessment 1. Were a sample of size  $n = 50$  taken which resulted in the likelihood function shown in Figure 4b, the posterior distributions based on prior 1 and on prior 2 would be as shown in Figure 4c. Note the larger change between prior and posterior distribution for #2 as compared with #1.

### Section 2.3 Methods of Assessing Probability Distributions

We assess subjective probabilities on two types of phenomena: single, non-repeating events, such as the shear strength of a specific laboratory specimen; and repeating events, which we model as random variables, such as the numbers of sinkholes in each square mile in a broad calcareous terrain.

For single, non-repeatable events, the event itself is the state of nature ("the strength of this specimen -- i.e., the state of nature -- is 10,000 psi"). For events which we model as random variables the parameters of that model become the state of nature ("the average number of sinkholes per square mile for this region is 3"). Note that given the state of nature for a single non-repeatable event, we perfectly know the value of the event; however,

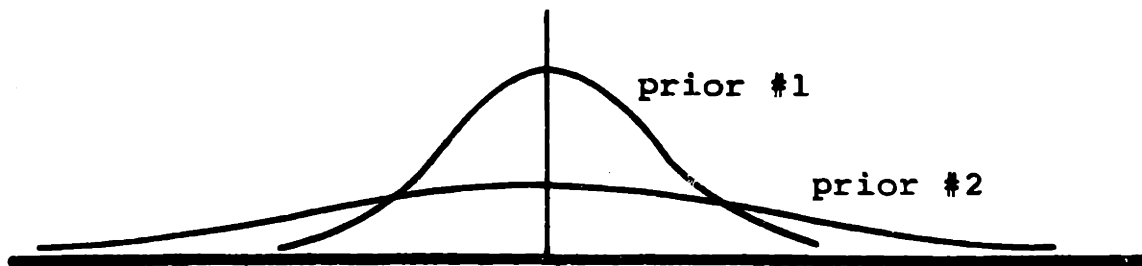


Figure 4a

given sample information  
likelihood function

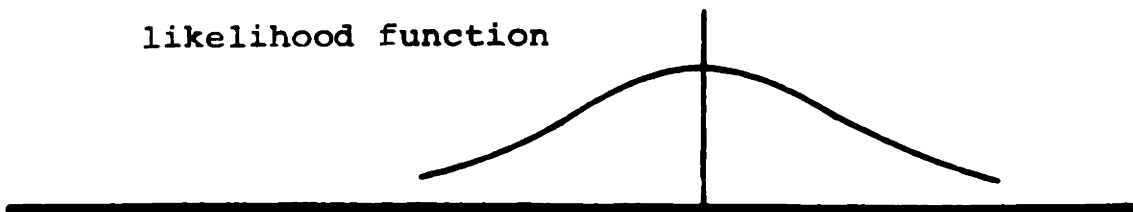


Figure 4b

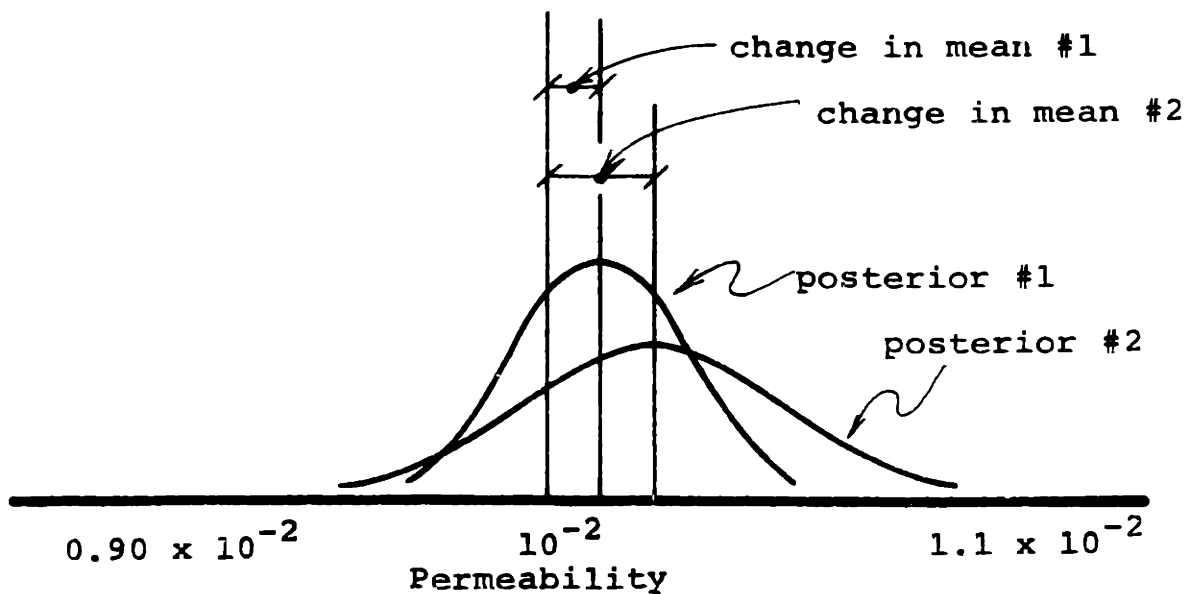


Figure 4c

given the state of nature for an event which we model as a random variable (i.e., the parameters of the model), we do not perfectly know the value of a single occurrence of the event (e.g., if we knew the average number of sinkholes per square mile to be 3.0, we would not perfectly know the number in the next square mile to be explored).

To assess subjective probabilities one offers the subject a choice between gambles, and from his preference determines a bound. By successively determining closer bounds, an estimate of the probability may be obtained. We considered this in Section 2.2, and discussed the validity of maximizing expected utility as a descriptive model with which to infer probabilities.

In this section we discuss what might be called "strategies of assessment": series of attributes of the subject's feelings for the UQ's which allow us to approximate probability distributions.

### Section 2.3.1. Single, Non-repeatable Events

There are primarily four strategies available for assessing subjective distributions: the first two are direct and the second two are indirect (meaning they assess information which is mathematically translatable into distributions).

1. Fractiles of the cumulative distribution can be assessed either directly or by successive subdivision of the range of the UQ into equally probable intervals.

2. Points on the probability density function can be determined by assessing relative densities and relative areas.
3. Judgement can be expressed in terms of an "equivalent prior sample" which, if observed, would have lead to the present "feelings."
4. The posterior distribution after a particular "hypothetical future sample," can be assessed and prior probabilities which would have lead to it, back-figured.

In this work we consider only direct methods since people are generally unable to accurately express their confidence in terms of hypothetical samples. As Phillips and Edwards (1966) found experimentally, "relative to Bayes' Theorem people are consistently conservative information processors, unable to extract from data anything like as much certainty as the data justify."

#### Fractile Method

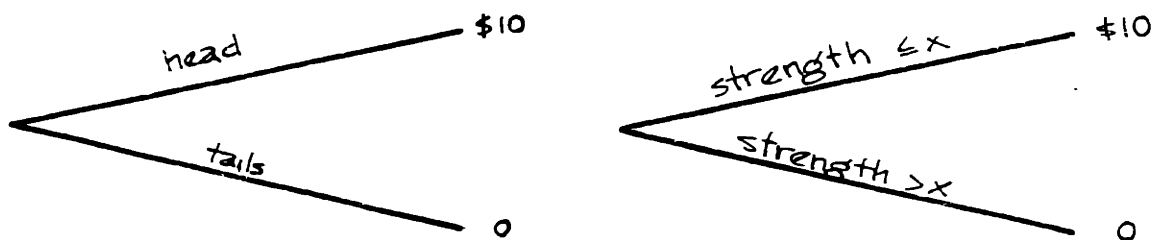
One usually assesses about five fractiles and approximates the entire subjective cumulative distribution on the state of nature by fitting a curve through the assessed points. Although in theory one can assess any fractile (e.g., 0.43), in practice a subject often becomes confused if too much refinement is sought, and his feelings are often not sufficiently precise to assess more than 5 or 6 fractiles (furthermore, greater precision often does not

change the optimal decision). Commonly used fractiles are 0.01, 0.25, 0.50, 0.75, and 0.99, because they are easily comprehended (e.g., 0.50 partition as the possible UQ values into two equally probable regions), and because a curve can be easily fit to them.

A line of questioning for assessing the subjective distribution of average shear strength in a formation might be as follows:

(1) "What value do you think the average shear strength is equally likely to be above as below? ... 20,000? ... 25,000? That is, for what value are you indifferent between the gambles:

Fig. 5

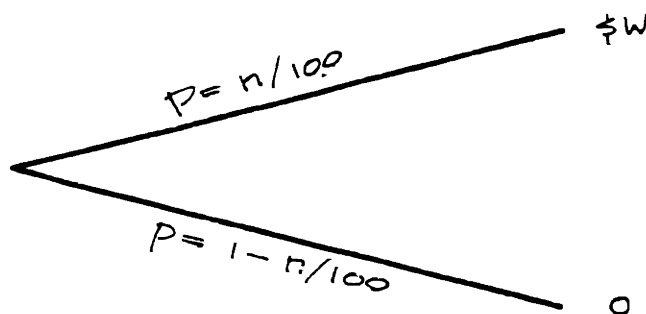


(2) "Now that we've established that 26,000 psi, say, is about the 0.50 fractile, we would like to assess the 0.75 fractile. Suppose you were told that (and only that) the average shear strength was greater than 26,000 psi. For what value of strength are you now indifferent between the gambles of Figure 5?" Continuing with similar questions we can assess several points on the cdf. Once an

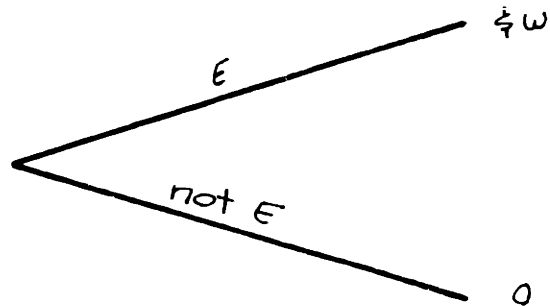


entire set of assessments has been made, additional questions (perhaps which re-assess one or more of the fractiles already assessed by using different gambles) should be asked as checks on the initial assessment.

An alternate way of constructing hypothetical gambles for fractile evaluation uses a cononical lottery (i.e., one in which each outcome has an equal probability of occurring) of 100 balls, each bearing a number from 1 to 100 (Schlaifer, 1961). One ball is drawn and placed in a closed box, and the subject is offered a chance to take part in the lottery by selecting one or more tickets, each also bearing a number from 1 to 100. If the number of any ticket he holds matches the number of the selected ball, he wins a prize, and if not, he receives nothing. The fractile of any value of the UQ is determined by the number of tickets the subject must hold to be indifferent between the lottery, in which  $n$  equals the number of



tickets held by the subject; and the gamble,



in which  $E$  is the occurrence of the event whose subjective probability is desired.

Obviously, the precise way in which questions are asked and gambles formulated affects the length of time required to assess distributions, and even the extent to which the assessment reflects the subject's true feelings (Section 2.2).

### Density Function Method

The only difference between the density function and fractile methods is the distribution assessed. The procedure of hypothetical gambles remains unchanged. One usually begins by assessing the most probable value of the UQ (i.e., the highest point on the pdf, or mode), followed by points which are "half-as-high", then by points which divide the area of the pdf into regions of certain relative

areas, and so forth. Curves fitted by eye are used to finally approximate the distribution (see Appendix for an example).

### Section 2.3.2 Random Variables

#### Univariate Random Variables

The assessment of subjective probabilities for univariate random variables (r.v.'s) is no more complex than for single events, but it does usually require more work. Although models for simple sequences of (independent, identically distributed) r.v.'s may have only one parameter (as in the Poisson distribution), they more commonly have two or more, and so in assessing subjective probabilities of the state of nature (i.e., the values of the parameters) we usually must assess distribution over more than one variable (e.g., the subjective pdf of mean and of variance).

#### Equivalent Sample Size

An alternate way of assessing subjective probabilities for r.v.'s is to assess the most probable values of the process parameters and an equivalent sample size which can be used as a measure of stability.

An equivalent sample size is the number of observations which our feelings might have been based on to give the present level of ambiguity. If our ambiguity were low, this sample size would be large; if our ambiguity were high, this sample size would be small.

Since we can relate sample size to the variance in estimates of parameters using sampling theory, we can obtain subjective probability distributions on the uncertain parameters.

Equivalent sample size is convenient, because it is concise (one number), and it fits easily into conjugate distribution techniques of Bayesian analysis. Unfortunately, since people are "conservative" information processors (i.e., since, relative to Bayes' Theorem, they do not extract as much confidence from sample data as they should -- Phillips and Edwards, 1966), subjects may unknowingly overestimate stability by stating equivalent sample sizes which are too large. Direct assessment of the standard deviation in estimates leaves less room for misinterpretation, and if equivalent sample sizes are desired for Bayesian analysis, they can be back-figured.

### Restrictive Families of Distributions

One finds that fitting restrictive families of distributions to the parameters whose probability distributions are being assessed requires less work than attempting to assess multivariate distributions by the fractile or density methods, and results in distributions which are analytically tractable -- of course, fitting families of distributions also restricts the shape of the assessment. When the number of variables exceeds two, this method becomes essentially a necessity, as we shall discuss in the next section.

The choice of distribution shape is not at all clear. One often fits families of distributions which are flexible, yet which have few

parameters. For example, the Beta distribution is frequently chosen. However, another common choice is the conjugate distribution to the likelihood function of the process whose parameters are being assessed. A conjugate distribution is one which has the property that: If the prior distribution of the parameters belongs to this family of distributions, then for any set of sample observations, the posterior distribution also belongs to this family of distributions. Obviously, in selecting a family of distributions one must consider trade-offs between distributions which seem acceptable models to the subject, and ones which will be easy to use in analysis.

Objectivist degree-of-belief'ers have long pondered the question of proper prior pdf's (and one might even suggest that -- philosophical issues aside -- this difficulty is a prime reason for there being few objectivists within the Bayesian ranks). There seems no way to answer this question unless one is willing to accept a measure of information content in probability distributions. If he is, he can assess those constraints about which there is low ambiguity (e.g., mean, standard deviation, range), and select the distribution which, while satisfying the constraints, is minimally prejudiced (i.e., contains the least information of all possible distributions).

One measure often suggested, and there are others, is Shannon's entropy function,  $H$ ,

$$H = \int_{-\infty}^{\infty} f(x) \log f(x) dx$$

1

in which  $f(x)$  is a probability distribution on  $x$ . For the case in

which  $x$  is a r.v. with known random process model  $f(x|\underline{\theta})$  (where  $\underline{\theta}$  is the set of parameters of the model), the minimally prejudiced parameter distribution,  $g(\underline{\theta})$ , is that which maximizes the entropy of the so-called Bayesian distribution of  $x$ ,

$$f(x) = \int_{\underline{\theta}} f(x|\underline{\theta}) g(\underline{\theta}) d\underline{\theta} \quad 2$$

This minimally prejudiced function,  $g(\underline{\theta})$ , is not necessarily the same as that which maximizes its own entropy,

$$H(g(\underline{\theta})) = \int_{\underline{\theta}} g(\underline{\theta}) \log g(\underline{\theta}) d\underline{\theta} \quad 3$$

Constraints for which the distributions of maximum entropy are common functions are listed in Table 1, and Zellner (1971) discusses minimally prejudiced distributions,  $g(\underline{\theta})$ , for normally distributed r.v.'s.

In view of the unknown effects of many factors which exert influence on assessed probabilities, such as procedure, increases in refinement due to minimum information considerations do not seem to justify their use. If one accepts subjective probability, yet admits the use of restrictive families of distribution, he might as well be entirely pragmatic and choose a convenient one.

### Section 2.3.3 Multivariate Assessments: Dependent Variables

The effort required to assess joint subjective probability distributions of multiple variables can be reduced if the variables

Table 1

Maximum Entropy Distribution

Constraints on prior information leading to common distributions (source: Tribus, 1969, 1962). For given constraints, distributions which maximize entropy can be found using Lagrangian multipliers.

Constraints	Maximum Entropy Distribution Family
1. $a \leq x \leq b$	uniform
2. $-\infty \leq x \leq +\infty$ mean, variance known	normal
3. $x > 0$ mean known	exponential
4. $x > 0$ mean and average of $\ln x$ known	gamma
5. $0 \leq x \leq 1$ average $\ln x$ , average $(1 - \ln x)$ known	beta

can be assumed independent. Denoting the variables whose subjective joint probability distribution is being assessed by  $x_i$ ,  $i = 1, \dots, n$ , the joint distribution can be found by,

$$f(x_1, \dots, x_n) = f(x_1, x_2, \dots, x_n) f(x_2, x_3, \dots, x_n) \dots f(x_n) \quad 4$$

However, since  $f(x_i | x_j) = f(x_i)$  for independent variables, equation 4 can be reduced to,

$$f(x_1, \dots, x_n) = f(x_1) f(x_2) \dots f(x_n) \quad 5$$

if we assume independence. Therefore, assuming the variables to be independent lessens the work required for assessment, because marginal distributions (i.e., the distribution of a parameter which expresses its behavior irrespective of other parameters) can be assessed one at a time without considering any but the one variable.

When independence cannot be assumed, assessments must be made by simultaneously considering more than one variable. There are primarily three techniques for assessing subjective distributions over dependent variables:

1. Relative probabilities of combinations of variables are assessed, and the entire distribution normalized. These assessments may be made using the fractile method, for example, but require several separate assessments.



2. The marginal distribution of one variable is assessed (e.g.,  $x_n$  in equation 5), then successive conditional distributions of the remaining variables. As with method 1, this requires several assessments, since conditional distributions must be assessed at more than one value of given variables.
3. A restrictive family of distributions is fit to the variables by assessing parameters of the model (this includes parameters accounting for dependence, such as covariances in the multivariate normal distribution).

The work required in assessment increases rapidly with increases in the number of variables. When this number is greater than 2, even fitting restricted families of distributions requires considerable effort. To fit a bivariate normal distribution requires 5 assessments; a trivariate, 9; and a quadrivariate, 14. Fortunately, since most site exploration problems involve spacial or planar distributions, one seldom must assess distributions of greater than three dimensions.

Although few investigations have been directed toward the assessment of multivariate distributions, certain comments can be made. In two dimensions both method 1 and method 2 can be satisfactorily used, but since to use method 1 the site must usually be discretized into large elements, the level of approximation it leads to may not be acceptable. In greater than two dimensions the only feasible method is to fit restrictive families of distributions, but

the assessment of these parameters is "tricky" (Pratt, et al, 1965) and redundancy should be designed into their assessment. Robinson (1971) discusses problems of assessing multivariate distributions as do Pratt, et al, and Keeney (1969) who treats multivariate utility functions.

In treating multivariate UQ's we would prefer to treat them as independent whenever possible, and in some cases we have means available for transforming dependent variables into independent ones. One such method is to find additional variables upon which the conditional distributions of the UQ's are independent. For example, the number of solution cavities in each of several contiguous strata of a thick limestone formation are highly dependent variables: the number in one stratum is highly correlated to the number in adjacent strata, since they depend on fracturing, the quantity of flowing water, and other features which are similar in adjacent strata. However, were we to conditionalize these variables on the regional factors which cause them to be dependent, we could assess the conditional distributions in isolation. To determine the marginal joint distribution (i.e., the joint distribution irrespective of the values of the various regional factors), which is the distribution we seek to obtain, we integrate over all possible values of the regional factors. If we let the number of cavities in the  $i$  th stratum be  $n_i$  and the set of regional factors on which the conditional distributions of the  $x_i$  are independent be  $\Theta$ ,

then,

$$f(x_1, \dots, x_n | \underline{\theta}) = f(x_1 | \underline{\theta}) f(x_2 | \underline{\theta}) \dots f(x_n | \underline{\theta}) \quad 6$$

and

$$f(x_1, \dots, x_n) = \int_{\underline{\theta}} f(x_1, \dots, x_n | \underline{\theta}) g(\underline{\theta}) d\underline{\theta} \quad 7$$

where  $g(\underline{\theta})$  is the subjective prior distribution on the regional factors.

Another method of making UQ's independent is by performing linear transformations. Figure 6 shows a hypothetical site with well defined regional dip. If the coordinate axes  $x$  and  $y$  are not taken parallel and perpendicular to this preferred orientation, the coordinates at which any stratum is likely to outcrop are dependent variables. For example, the probability of sandstone along the line  $x = x_0$  depends on  $y$  (i.e.,  $\text{Pr}(\text{sandstone @ } x_0, y_0) \neq \text{Pr}(\text{sandstone @ } x_0, y_1)$ ). Were we to rotate the coordinate system, making  $x$  parallel to the regional strike and  $y$  parallel to the regional dip,  $x$  and  $y$  would be independent.

In some cases independence, or lack of it, is obvious; in other cases it is not. In cases for which independence is not apparent, one is not clear on how to proceed. Attempts have been made to formulate direct measures of dependence which would allow us to verify that variables were sufficiently independent for us to treat them as such in assessment (Robinson, 1971), but a satisfactory measure is not now available. Since we are ultimately concerned with

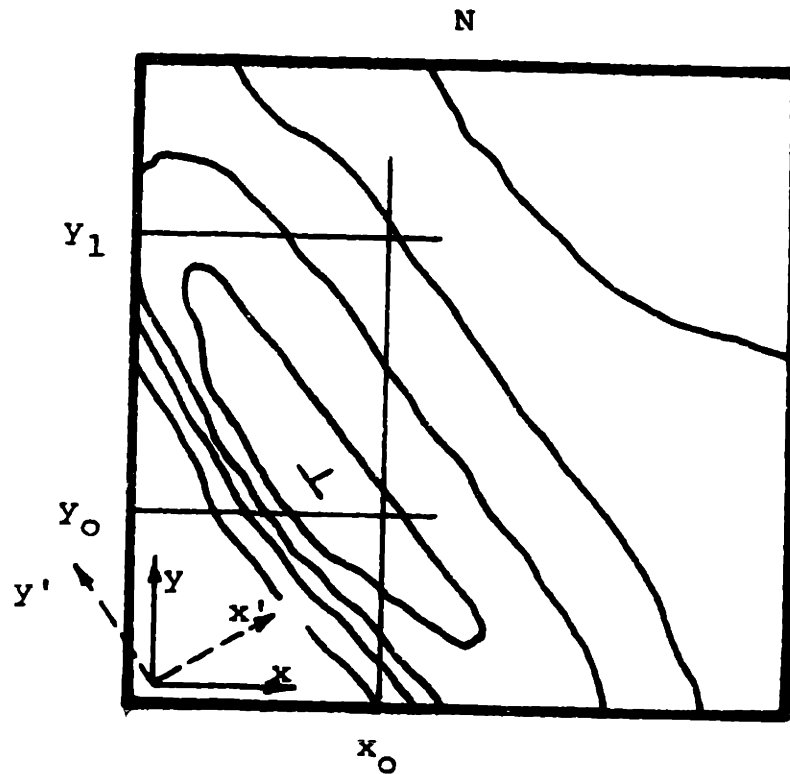


Figure 6

Hypothetical topographic data showing hogback ridge dipping towards the NE. With coordinate axes parallel to the N-S and E-W directions the unknown coordinates of outcrops of any one bed are dependent variables. Coordinates of outcrops may be made independent by rotation to  $x', y'$ .

how priors affect optimal decisions, Kaufman (see, Robinson, 1971) has suggested that two optimal decisions be evaluated: one using priors assuming independence, and one using priors assuming great dependence. If the two optimal decisions are approximately the same, further analysis is unnecessary; if the two are drastically different, one must go back and assess a more detailed dependent distribution. There are drawbacks to this simple procedure, which Robinson discusses, but until a satisfactory measure of dependence or similar test is developed we must rely on heuristic rules.

#### Section 2.4 External Validity

Subjective probability assessments are merely statements of what a subject knows or does not know about an UQ and therefore their "goodness" is related only to how well they mirror his feelings. If the true value of an UQ turns out to be one which the assessment indicates is improbable, all we can do is dismiss it as a rare event, or conclude that the subject's feelings were not very good.

Nevertheless, we hope that "in the long run" (i.e., over the assessment of many UQ's a subject has feelings which are reflected in reality. In other words, we hope that errors occur with frequency proportional to their subjective likelihoods. When this condition is met we say that assessed distributions are externally valid.\*

---

\*One has little trouble finding assessments which are not externally valid. For example, estimates of the time required to accomplish tasks are usually biased low and are almost always overconfident.

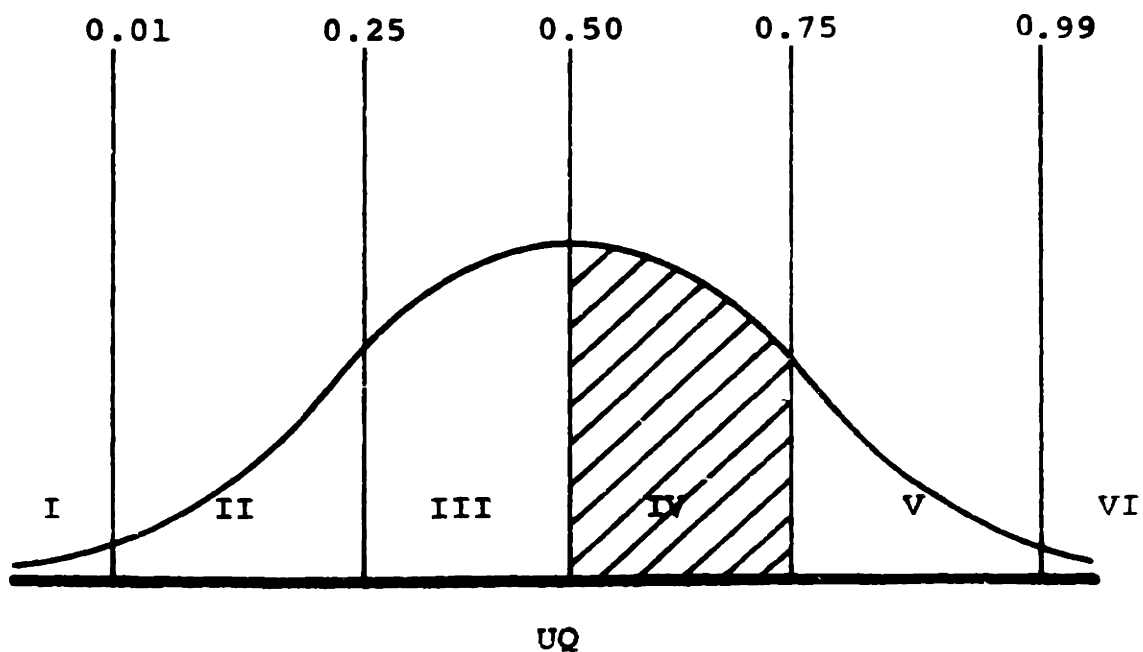
No matter how a subject makes decisions -- by closing his eyes and leaning back in his chair, or by probabilistic analysis -- his ultimate choice is based on his feeling for the likelihood of various contingencies. He can empirically evaluate the external validity of his feelings by using subjective probability as a vehicle for comparing them with reality. There is no discipline where need for such introspection is more necessary than in soil and rock mechanics, since technical decisions in these fields are made on remarkably scanty information.

Although few studies have been made of external validity, the ones which have are both interesting and disturbing. They seem to show similar, consistent, and unconservative behavior for very different types of subjects assessing very different types of UQ's.

#### Alpert and Raiffa (1969)

Alpert and Raiffa (1969) studied external validity by comparing subjects' assessments on several UQ's with the true values. They used graduate students as subjects, and fixed, but unknown "almanac type" statistics for UQ's (e.g., annual revenue of the Panama Canal, results of recent Gallup polls).

If the assessments in the study were externally valid, the number of true values of the UQ's falling into certain ranges of their assessed probability distributions would be proportional to the likelihood of those ranges. For example, about 1/2 of the true values would fall between the 0.25 and 0.75 fractiles of their



Subjective distribution of UQ

<u>Region</u>	<u>Nominal frequency</u>
I	0.01
II	0.24
III	0.25
IV	0.25
V	0.24
VI	0.01

Figure 8

Table 2  
 (after Alpert and Raiffa, 1969)

UQ	I	II	III	IV	V	VI	TOTAL
1	33	16	20	40	11	10	100
2	15	12	35	19	10	9	100
3	11	8	28	29	13	11	100
4	51	41	6	1	1	0	100
5	1	1	13	28	29	28	100
6	24	14	12	13	10	27	100
7	1	3	11	9	15	61	100
8	9	2	13	10	8	58	100
9	25	15	18	9	7	26	100
10	18	8	8	12	16	38	100
total	158	120	164	170	120	268	1000

number of real values falling in given intervals



respective assessed distributions (Figure 8). Table 2 shows the actual results for 100 subjects.

In this sample 33.4% of the true values fell into the interquartile range (0.25 to 0.75), compared with a nominal value of 50%. This indicates that the interquartile ranges were on the average too narrow. However, note that over 40% fell outside of the 0.01 or 0.99 fractiles. This is an astonishing statistic, since, were the assessments externally valid, only 2% should have fallen in this range. The apparent conclusion is that subjects grossly overestimate their confidence: the range in which subjects are surprised to see true values fall is far too large.

The authors then explained their findings to the subjects, emphasizing the overconfidence exhibited in the first round, and asked the subjects to assess their feelings about a new set of UQ's. Comparison of both rounds is shown in Table 3. Although the proportion falling in the interquartiles approached its nominal value in the second round (i.e., 43% compared with 50%), the proportion outside 0.01 or 0.99 was still 24% -- even after feedback from the first round. Feedback helped, but subjects were still grossly overconfident, and so external validity was still seriously deficient.

As further example, Alpert and Raiffa then asked different subjects to assess the "maximum" and the "minimum" values specific UQ's might have, and also values which were "astonishingly high" and "astonishingly low." When the true values of the UQ's were compared, 47% were outside of the assessed maximum or minimum, and

Table 3  
(after Alpert and Raiffa, 1969)

Comparison of first round  
external validity with  
second round external validity

	Round 1	Round 2
percent of true values inside interquartile range	33.1	43.3
percent of true values falling outside 0.01 or 0.99 fractiles	40.6	23.5

38% were outside the astonishingly high or low values (in a separate experiment the respective values were both 35%).\*

#### Folayan, Hoeg, and Benjamin (1970)

To the author's knowledge only one attempt to assess subjective probabilities in engineering geology has been published, but its results, when analyzed for external validity, are strikingly similar to the results of Alpert and Raiffa. Folayan, Hoeg, and Benjamin (1970) assessed the judgemental probabilities of five experienced engineers about the compression ratio of San Francisco Bay mud at a particular site, before any samples had been taken. The authors asked each engineer to assess a mean value, the variance about the mean, and an "equivalent number of samples" as a measure of stability (Table 4, Figure 9). After the assessments were made, 42 specimens were sampled and tested. The results of those tests are shown by the dashed distributions in Figure 9. Although the authors did not discuss external validity, upon inspection we see large discrepancies between priors and the sample.

All of the priors are biased low compared to the sample by as much as 50% (which in the case of settlement is unconservative),

---

\*The authors conclude that subjects tend to interpret "maximum" and "0.99 fractile" as about the same, and "astonishingly high" and "0.999 fractile" as about the same. Their subjects were mostly business and economics students; perhaps engineers would interpret "maximum" and "minimum" in a stricter sense, and make wider assessments for these limits.

Table 4  
 (after Folleyan, et al, 1970)

criteria	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5
mean	0.30	0.275	0.275	0.30	0.30
variance	$0.91 \times 10^{-3}$	$0.23 \times 10^{13}$	$0.23 \times 10^{-3}$	$0.91 \times 10^{-3}$	$0.81 \times 10^{-2}$
equivalent sample size	82	42	12	4	2

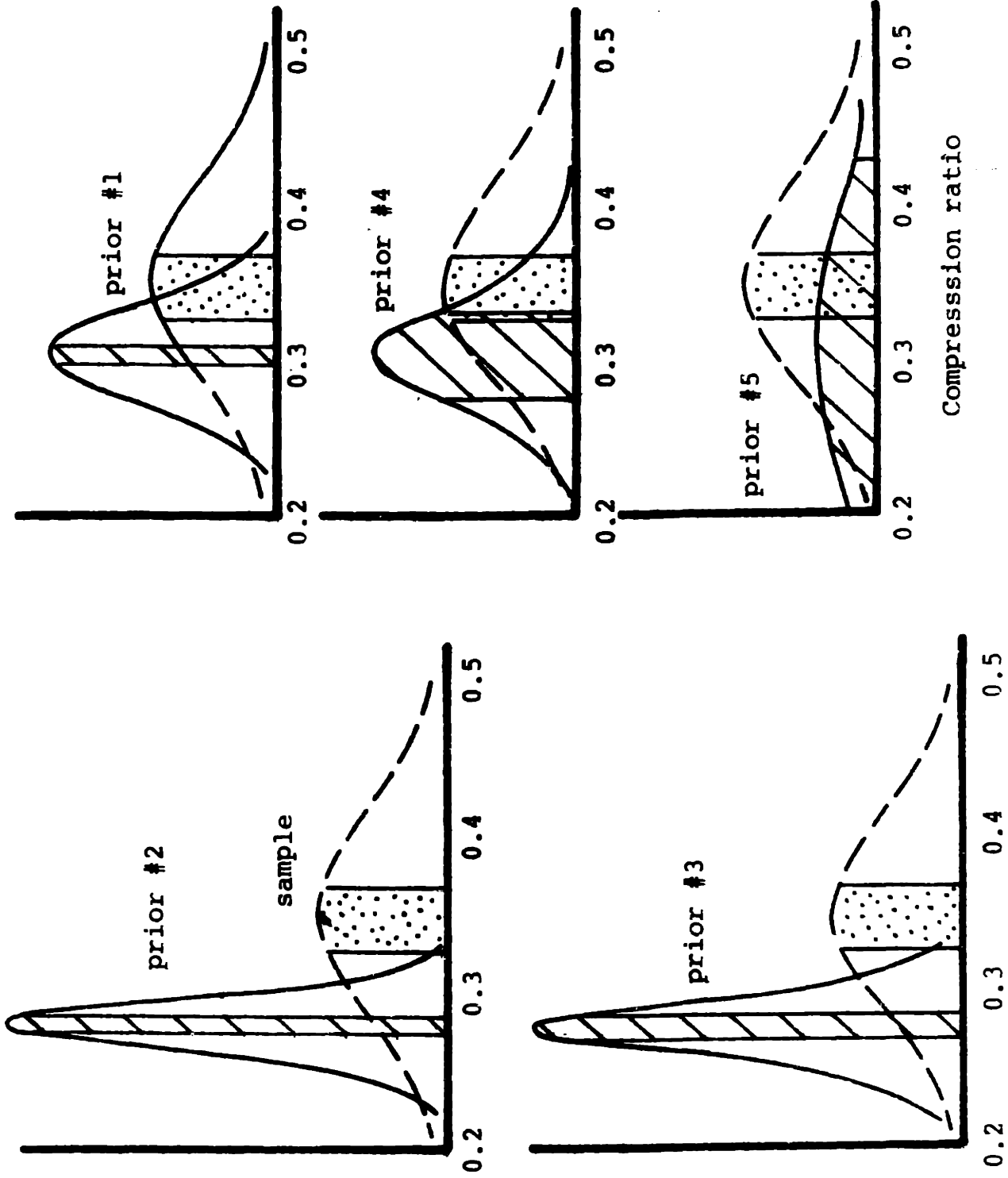


Figure 9: (data from Folayan, et al, 1970)

and the estimated variances (with one exception) are low by as much as an order of magnitude. Since a quick survey of the literature (e.g., see, Mitchell, 1963) suggests that this site was anomalously compressible, for the present we may dismiss bias errors. A more interesting observation -- particularly in light of Alpert and Raiffa -- is that the actual sample data is very unlikely, when conditioned on the prior distributions. Were we to assume the priors normally distributed, the distribution of the estimate of the mean (i.e., the subjective relative likelihoods of what the subject feels the average compression ratio is likely to be) would also be normally distributed; with variance  $\sigma^2/n$ , where  $\sigma^2$  is the estimated variance of the compression ratio and "n" is the equivalent sample size.\* In Figure 9 we have shown the 0.025 and 0.975 fractiles of these subjective distributions, and the interval between them is cross-hatched. The true average compression ratio should lie outside this interval only 5% of the time. We have plotted the similar interval for sample data and indicated it with stipples. Only for one of the five assessments do these intervals overlap (and that one was made by a student who simply rated this ambiguity as very high, "n" = 2). Again subjects (particularly, in this case, experienced engineers) have greatly overestimated their confidence.

---

\*This result for the mean is "robust" with respect to the assumption of normality of the individual samples; which simply means that it is insensitive to departures from the assumption (see, Raiffa and Schlaifer, 1961).

One might convincingly argue that using "equivalent sample size" as a measure of stability may mislead a subject if he is not familiar with statistics and the consequences of his assessment. We suggested this in Section 3, and argued for measuring stability directly. Let's say, to be extreme, that the subjects overassessed their equivalent sample size by a factor of 2. Recalculating the same fractiles for the distribution of the mean using  $n/2$ , we obtain the intervals indicated by dashed lines in Figure 10, and once again three of the five intervals do not overlap the interval from the sample data. One concludes from this study that the subjects were all more confident than they had reason to be -- precisely the conclusion one draws from Alpert and Raiffa.\*

On the basis of a few studies we do not propose that overconfidence in the stability of predictions is a universal trait, but were it common the consequences for engineering geology are

---

\*We do not know what effects the precise procedure of assessment had in causing over-confidence, since the authors did not discuss their procedure and so little is known of procedural effects anyway. Intuitive interrogation as opposed to action-approaches may induce over-confidence by allowing the subject to be flippant about the assessment. Furthermore, one knows from experimental psychology that once a subject states an opinion he is adamant in sticking by it, even long after it ceases to be tenable. This is sometimes called cognitive dissonance. Asking him to make a best estimate first may induce him to "stand by" that assessment by assessing narrow bounds on his feelings. Were we to request he assess bounds first, then a best estimate, his over-confidence might not be so extreme.

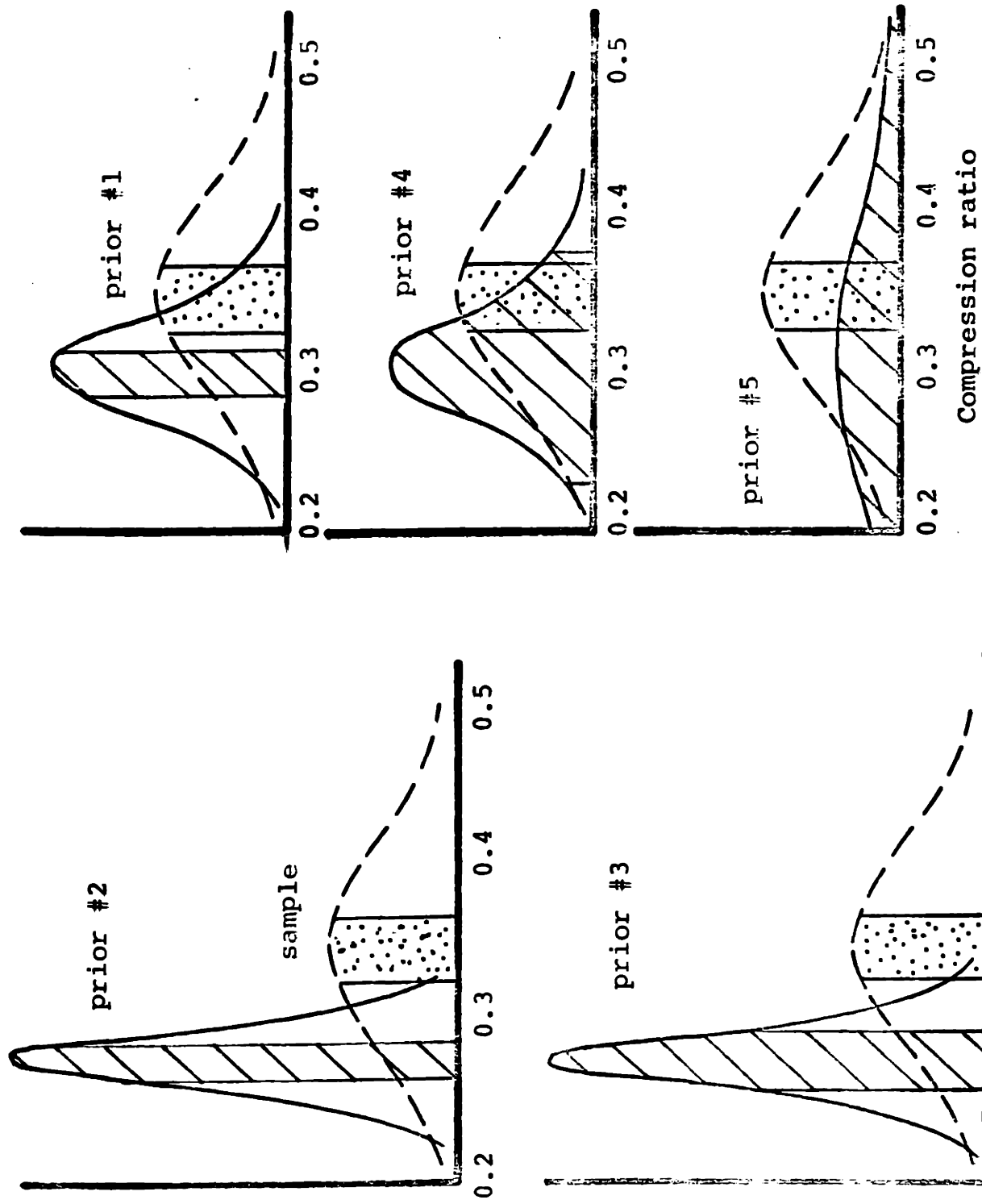


Figure 10 (data from Folleyan, et al, 1970)



apparent -- whether or not quantitative decisions are made. One is intrigued by the promise of further inquiry into the way technical decisions are made in engineering geology, using subjective probability and utility as research tools. The results of this inquiry would be equally informative for the qualitative decisions now made as for the quantitative decisions which may be made in the future. Of immediate interest is this question of over-confidence. An allied question is how much weight do engineers attach to low probability events -- are they given consideration appropriate to their expected consequences, or disproportionately greater as would be implied by Preston and Baratta's data? The latter would counteract tendencies toward "narrow" priors.

## Section 2.5 Some Comments in Summary

### Operational Difficulties

The practical assessment of subjective probabilities is hindered by two obstructions, winning the DM over to the rationality of quantifying his judgement, and the time required to assess distributions.

There is little one can say of the first obstruction that does not apply to all technical change. The best way and perhaps the only way) to win adherents is to demonstrate that the new techniques save time and money for someone else who is using them. At first appearance, the Folayan, et al., article seems to be having this effect. As we will see in the entire thesis, optimal strategies and

allocations of effort are not always intuitive, so quantifying judgemental probabilities can at times lead to significant savings.

Even though assessment requires time, the entire decision process may be shortened by using quantitative methods. Grayson (1960) suggests that for one dimensional r.v.'s and single, non-repeatable events, quantitative assessments may even be quicker than attempting to describe "feelings" through not sufficiently precise adjectives, especially when more than one person is involved in the decisions. As he further points out, quantification of judgement also simplifies the transfer of information between specialist and decision maker.

When the UQ becomes multivariate the work required for assessment rapidly increases, even if restrictive families of distributions are fit. At present there is no way around this problem. If the subject can narrow things down by simplification, distributions might be easily assessed, but this will often cause the distribution to be less than an acceptable model of the real situation. Fortunately, site exploration problems seldom treat more than three dimensions, but nevertheless, the work required for assessment may still present difficulty.

One might argue that subjective assessments are imprecise; too approximate to be of any use. The obvious answer is, "so are judgemental feelings, yet this is what decisions must ultimately be based on." Numbers do not necessarily imply objectivity. They are merely a language which allows "feelings" to be expressed more

precisely and interfaced with other facets of decision making. In situations where the optimal decision is insensitive to the exact form of priors, imprecision is admissible; in situations where the optimal decision is sensitive, quantification tells us which parts of our "feelings" we must be most careful in assessing.

### Behavioral (Psychological) Difficulties

A descriptive model of behavior under uncertainty (as opposed to Bayesian decision theory which is a normative model) is required to infer subjective probabilities from subjects' preferences between gambles. We do not now have such a model. Instead, we assume that maximizing expected utility describes subjects' behavior in simple, low-reward gambles. We have seen evidence in Section 2.2 that this assumption is questionable.

Hand-in-hand with the development of a better descriptive model must be standardization of the games and exact choices offered subjects during assessment. We know that procedural details affect other psychophysical experiments (Luce and Suppes, 1965) and there is no reason to think that they do not affect probability assessment. We have almost no empirical information about the affects of procedural details, since most workers have concentrated their efforts on the mathematics of subjectivist decision theory rather than on its application.

## APPENDIX TO CHAPTER 2

In this appendix we present two examples of subjective probability assessments: Both are of an engineering geologist with approximately 15 years experience.

### Assessment #1: Unconfined Shear Strength

Uncertain Quantity: The first UQ was a univariate random variable: the distribution of unconfined shear strength of a gypsum modeling material used for studies on the behavior of jointed rock. In his capacity as research director, the subject had familiarity with this specific material.

Procedure: The assessment was performed using both fractile and density methods. Because the first assessment using the density method followed closely in time the assessment using the fractile method, a second density assessment was made on a subsequent day.

Fractile method--The flip of a coin and Schlaifer's standard lottery were used as hypothetical gambles to assess the 0.50, 0.25, 0.75, 0.01, and 0.99 fractiles, in the order stated. Assessment of the interquartile range was made for redundancy.

Density method--The same hypothetical gambles were used to evaluate, in order, the mode, points "half-as-high," and points beyond each of which lay 1% of the area

under the curve.

Results: Plots of the assessments are shown in Figures A1 - A4.

Discrepancies between them were not great, and the first density assessment was actually more discrepant from the fractile assessment than was the second.\* In practice one would use some distribution averaged from these.

Contrary to the findings of Alpers and Raiffa, and Folayan, et al the subject assessed distributions which were more disperse than the actual distribution. The reason is probably that he was aware of the earlier findings on probability assessment indicating a trend toward over-confidence, and made a conscious effort to overcome such tendencies.

---

\*One is referred to Winkler (1967a) for discussion of discrepancies between the results of fractile and density methods in controlled studies.

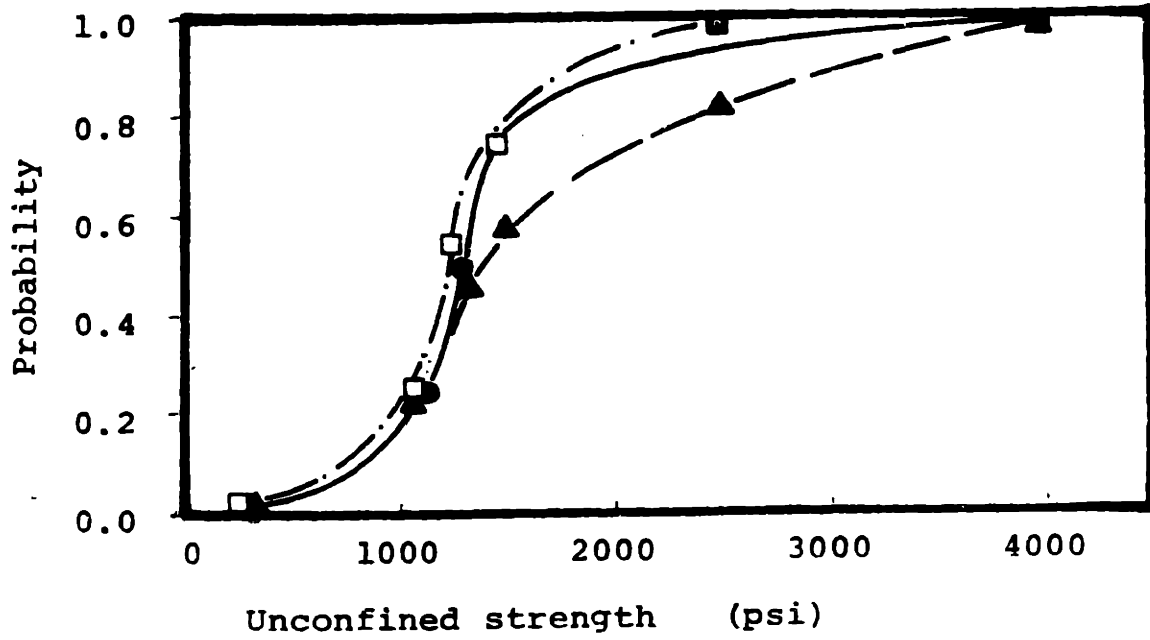


Figure A-1

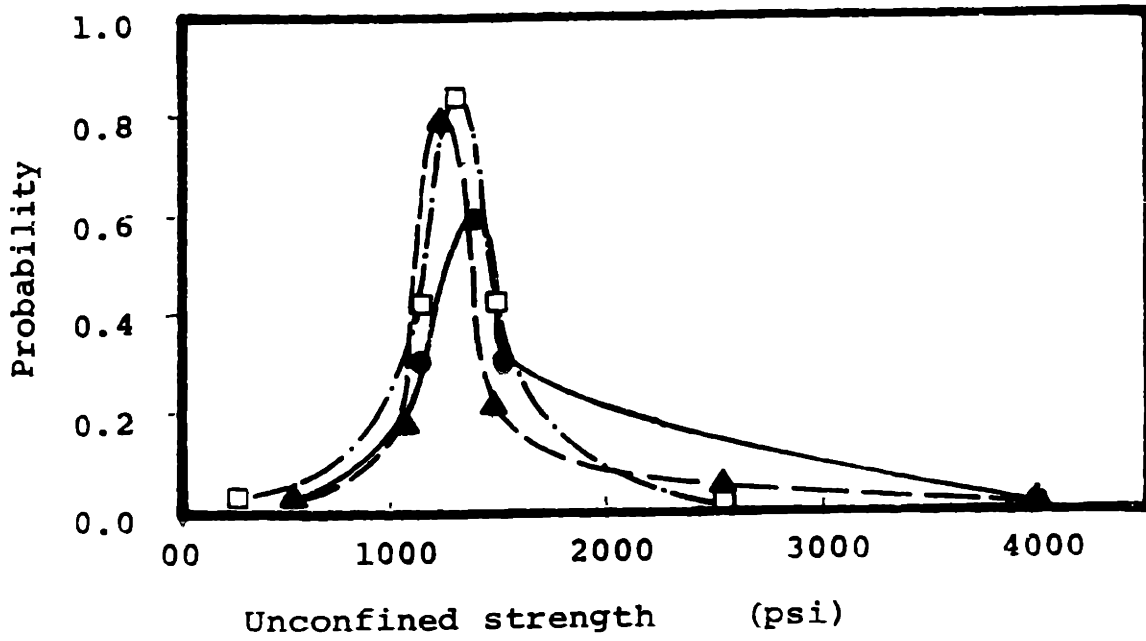
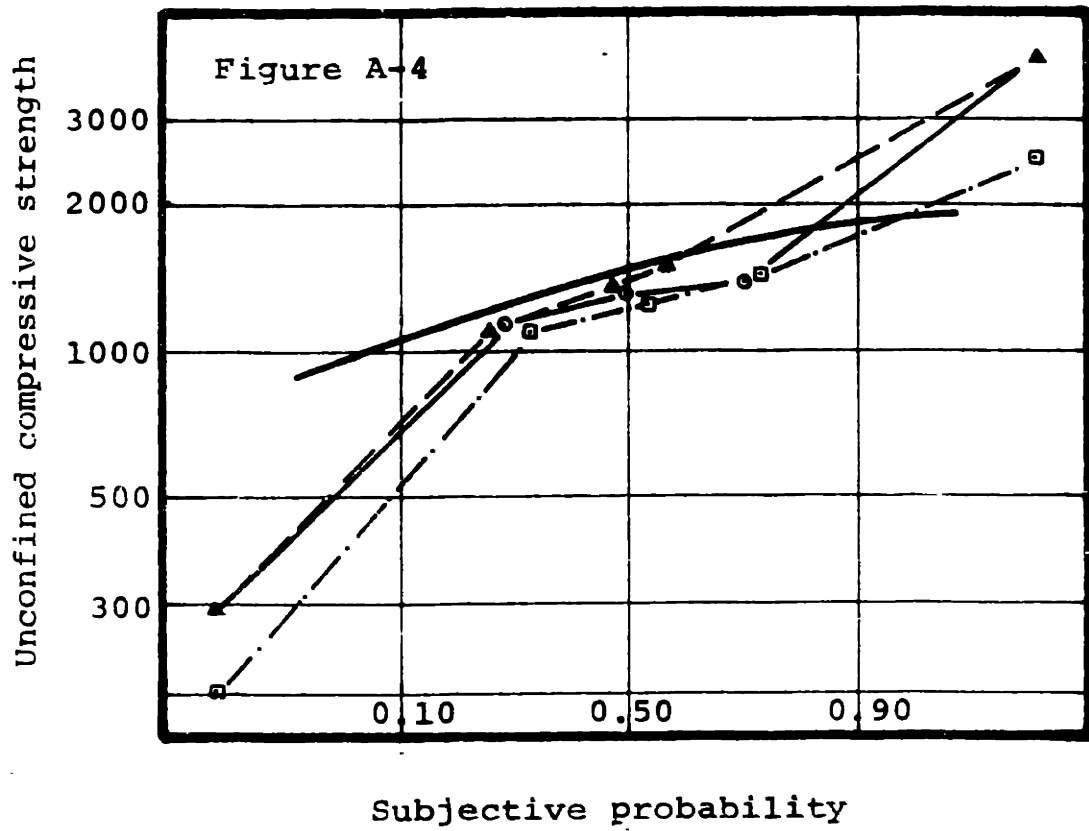
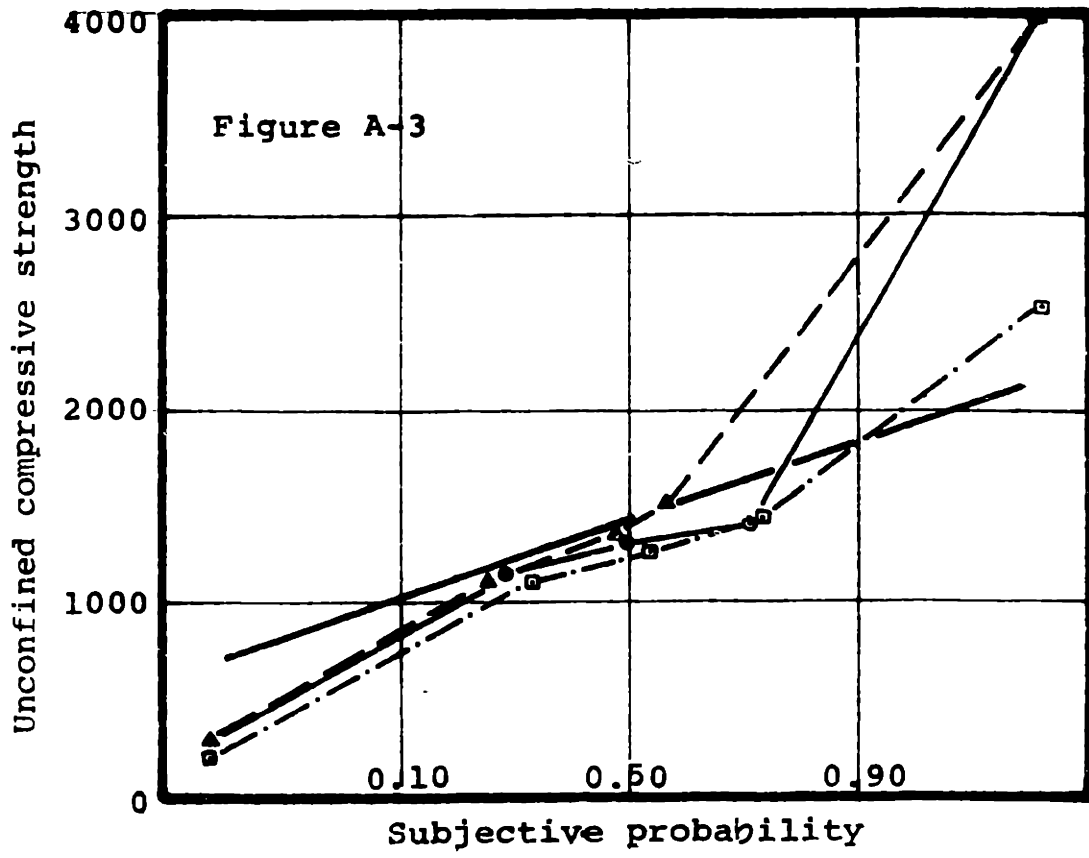


Figure A-2

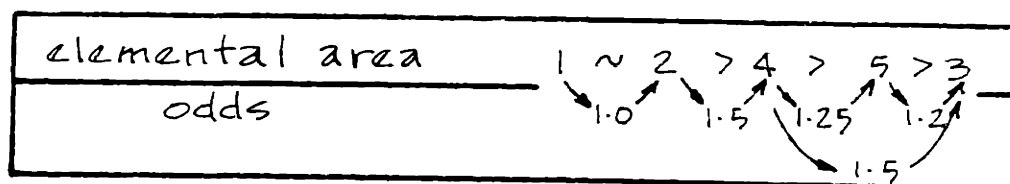


## Assessment #2: Fault Location

Uncertain Quantity: The second UQ was the location of faulting in a calcareous terrain, to be used in conjunction with allocating search effort.

Procedure: To simplify assessment the site was partitioned into five elements of approximately one square mile each (Figure A5). The subject was offered a series of bets involving each combination of two elemental areas. The bets were structured such that if he chose the element containing a fault, he won, and if not, he lost. Relative probabilities were assessed by determining the odds required to make him indifferent between choosing one or the other element in each pair.

Results: The preferential ordering of probabilities that elemental areas contain faults was



which translates into the probability mass Function of Figure A6.

Were the subject to allocate effort to the five elemental areas in search of faults (assuming the existence of a fault in each area to be equally important), he should allocate the most effort to areas 1 and 2, and the least effort (if any at all) to area 3 (see Chapter 6 for details of search theory).



Figure A-5

Topographic map of site

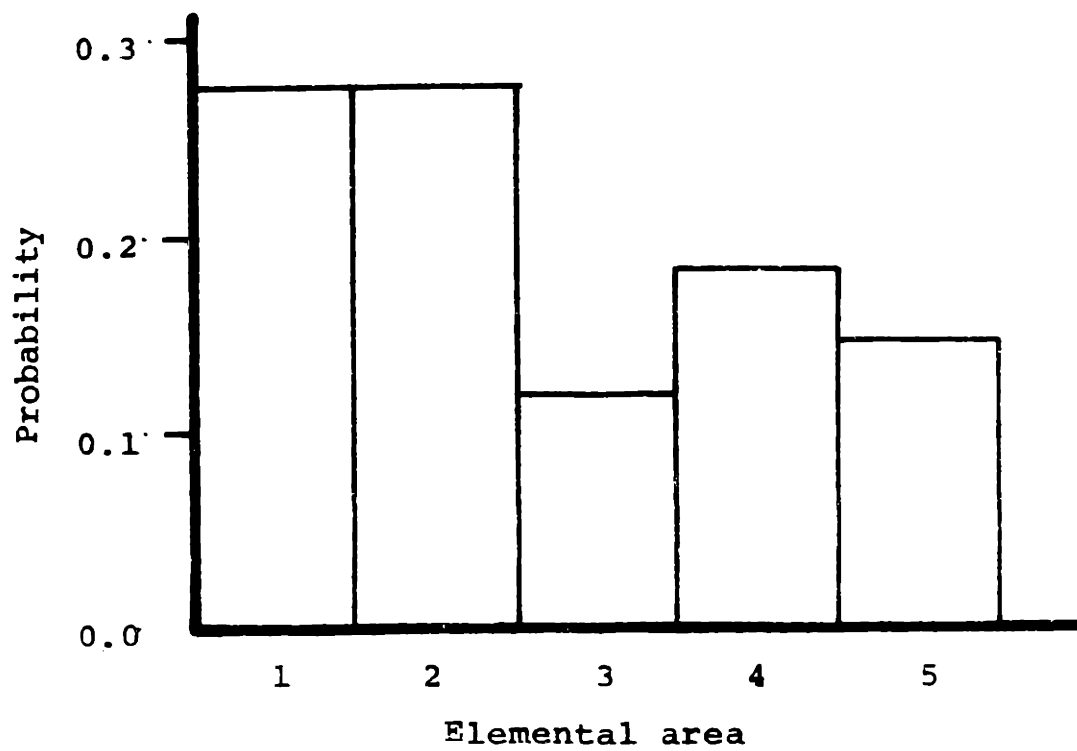


Figure A-6

## CHAPTER 3

SITE EXPLORATION: PROBABILITYSection 3.1 Site Exploration

A site exploration program is a plan of action for obtaining information on local geology. Ideally, it should result in an association of perfectly known engineering properties (i.e., strength, deformation, and permeability) with every point in the space explored.

On the other hand, in reality the resources available for exploration are limited: expenditures must be commensurate with the proposed construction and the consequences of using imperfect information. Generally, given these limitations, one tries to obtain as much information as possible about four aspects of site geology:

1. The geological nature of deposits and formations
2. The location, thickness, and material composition of all formations.
3. The engineering properties of formations affecting performance of the proposed structure.
4. The ground water level and its fluctuations.

All information resulting from site exploration may be characterized as being either of geometric relationships or of material properties; information on the former is the more important (particularly for the design of structures in rock). Precise information on material properties becomes important only when geometric relationships are well known (e.g., see Terzaghi and Peck's (1967)

discussion of heterogeneous vs. homogeneous sites, or Terzaghi's (1929) discussion of geological details in dam construction).

Since exploration is a highly subjective undertaking depending on judgement (which we take to mean experience and knowledge of geology), it has long been handled by rules-of-thumb. Consequently, although textbooks and journal articles devote considerable attention to the detailing of sampling devices, few accounts of exploration strategies are published.\* When engineers do devote attention to the organization of exploration, it is commonly described as consisting of three temporal stages: Reconnaissance, preliminary exploration, and detailed investigation (Fig. 1).\*\*

#### Reconnaissance:

Reconnaissance involves a general review of local geology whose purpose is to obtain qualitative estimates of

1. geologic nature of major formations at the site,
2. engineering properties of major formations,
3. possibility of adverse geological conditions.

This is accomplished by collecting published documents (e.g., regional geologic and topographic maps, records of nearby construction, airphotographs) and possibly by on-site inspections.

---

\*For exception to this rule, see Knill and Jones (1965) or Fookes (1969).

\*\*One might also include a fourth stage, construction observation and post-construction monitoring, although this is usually not considered as part of exploration.

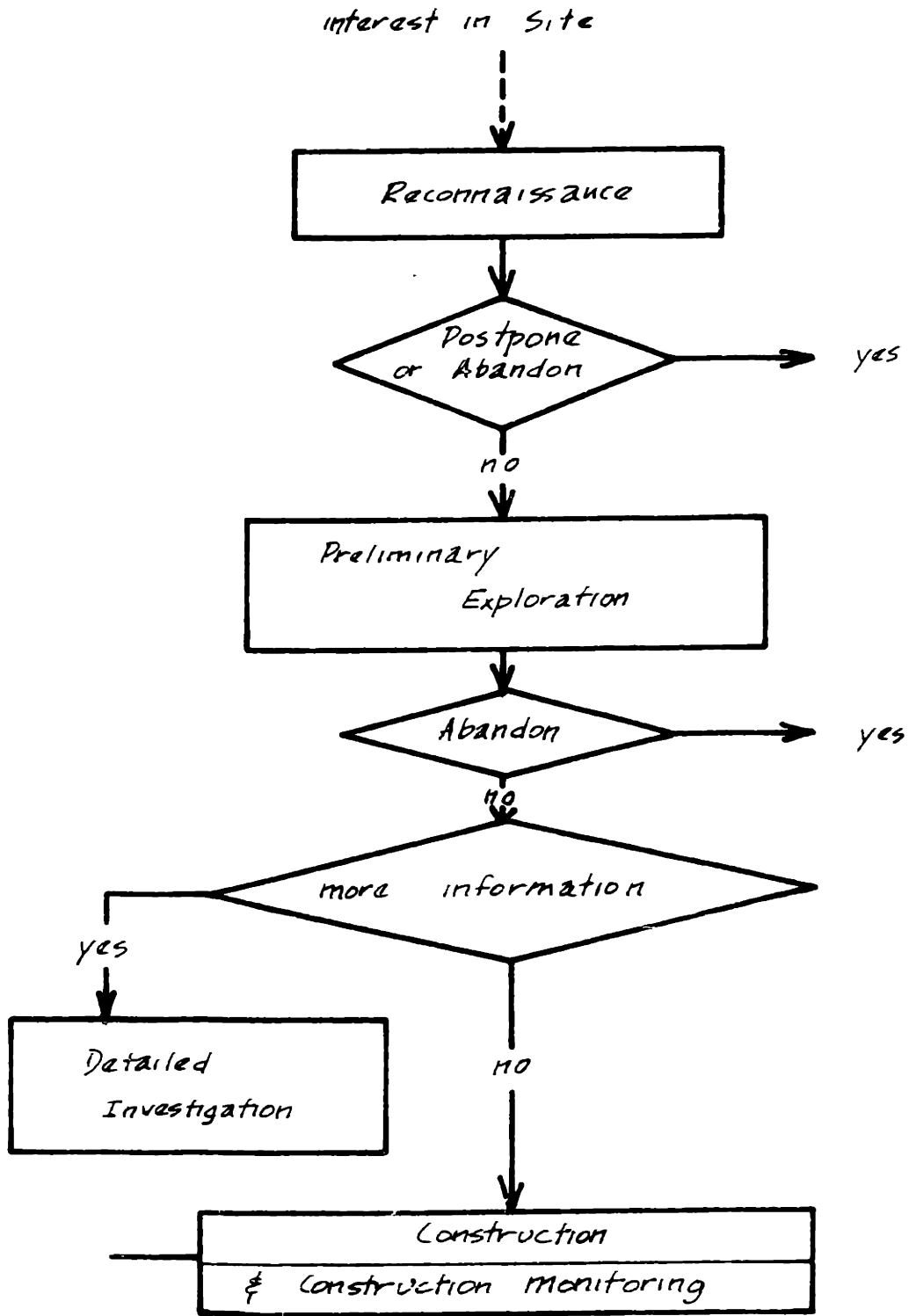


Fig. 1

Classical Description of Exploration Process

### Preliminary Exploration:

Preliminary exploration is oriented toward collecting the first quantitative data from the site. The location and general geometry of major formations is determined by a limited number of borings, by field mapping, and perhaps by geophysical surveys. Initial quantitative estimates of engineering properties are made, and the type and possible location of adverse geological details are hypothesized. Joint surveys are usually begun.

### Detailed Investigations:

In detailed investigation one seeks accurate information on geometry and material properties through intensive boring programs, detailed field mapping, and additional geophysical surveys. Samples are taken for laboratory use, and in-situ testing is often performed to obtain better estimates of engineering behavior.

Although these stages are temporal ones, their boundaries are not distinct: Reconnaissance overlaps preliminary exploration, and preliminary exploration overlaps detailed investigation. In reality, exploration is conducted sequentially, with all preceding results evaluated before subsequent effort is allocated. The functional difference between the classical stages, therefore, lies in the sort of information they seek rather than their place in time or their sequential dependence. Reconnaissance seeks qualitative information on regional geology and geomorphology, preliminary

investigation seeks patterns (i.e., general geometries) of site geology, and detailed investigation seeks precise material properties and precise locations of adverse anomalies.

### Section 3.1.1 Nature of Uncertainty in Site Exploration

The conclusions an engineer draws from the records of exploration--for example, the patterns he recognizes, the maps he constructs, the similarities or dissimilarities he infers to exist across the site--contain much more than the records themselves: Hypotheses are the primary, useable result of exploration. Uncertainties in site exploration, therefore, manifest in the degree to which hypotheses are confirmed by exploration data; that is, in the degree to which the predictions of hypotheses are consistent with field observations.

Hypotheses arise through inductive reasoning, and so, if one is to believe Hume, must always be non-logical. One may think of their development as consisting of two parts. The first is the discovery of hypotheses, which Charles Pierce (see, 1970) called "abduction," and the second is the assignment of a priori degrees of confirmation, which Pierce called "induction proper" (Fig. 2).

Attempts at explaining abduction have led to several schools of philosophical thought. On the one extreme is the view that discovering hypotheses is the unpredictable act of genius which can never be placed in a theoretical framework; on the other extreme is the view that all possible hypotheses are mentally enumerated but only the most plausible retained; between them is the view of Poincaré

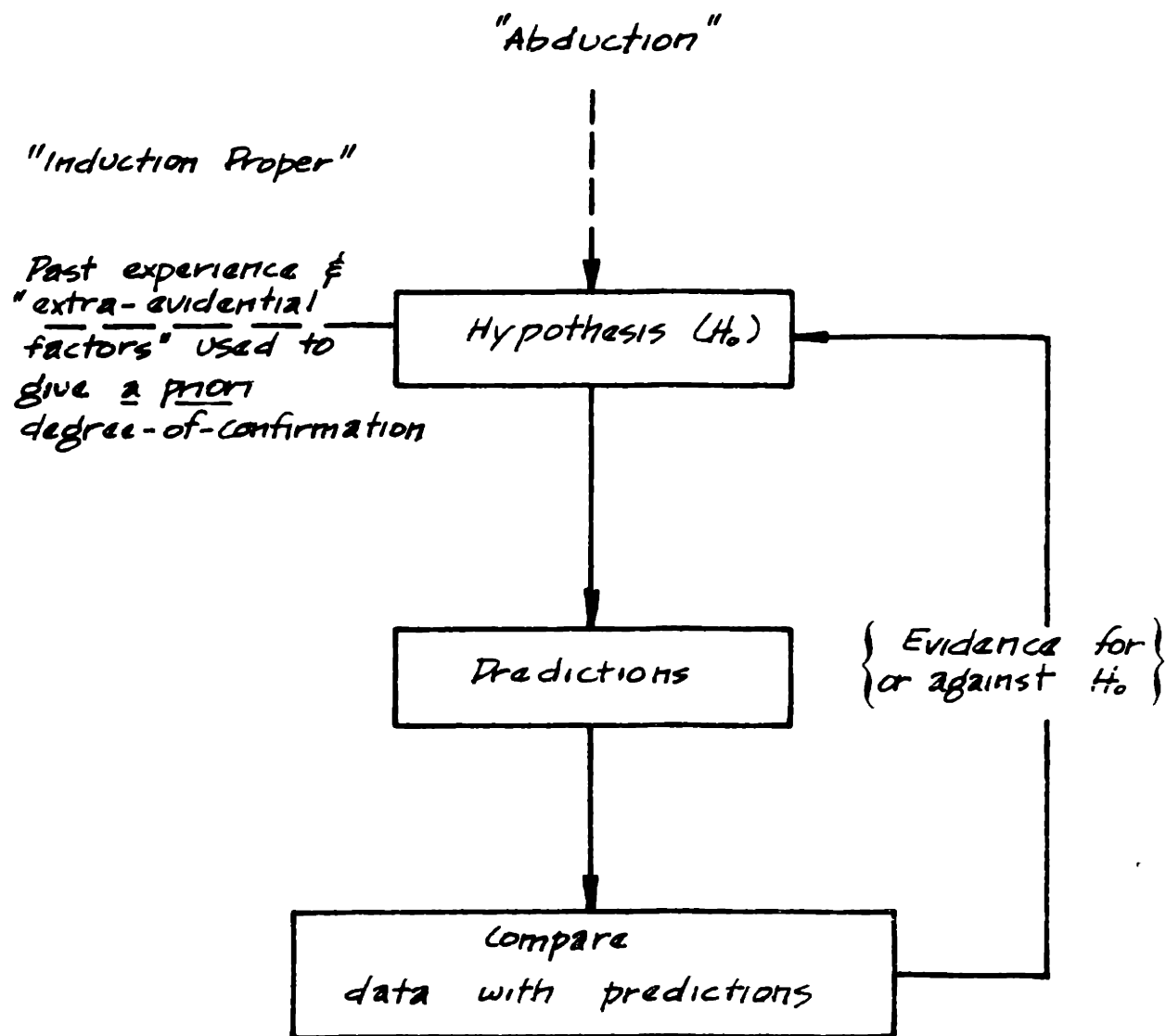


FIG. 2

Induction and Confirmation of Hypothesis

(1914) who postulated a demon guarding the threshold between conscious and unconscious, an "esthetic referee" who admits into the conscious only those hypotheses having "esthetic appeal."

The assignment of a priori degrees of confirmation is similarly subjective; their level is influenced by many factors beyond past observations (i.e., "extra-evidential" factors). Some particular extra-evidential factors are the following (see, e.g., Watanabe, 1969; Salmon, 1966).

1. One hypothesis is often believed in more than another for esthetic reasons--it is simpler, more elegant, or just more to the engineer's liking.
2. The credibility of a hypothesis diminishes in the presence of one that is more credible.
3. Of two hypotheses equally well confirmed by past observations, the one more compatible with other hypotheses about the site will be given more credibility.

So the uncertainties of site exploration are associated with the development of hypotheses, and these in turn are associated with inductive reasoning. Inductive reasoning is highly subjective, and consequently the uncertainties of site exploration are highly subjective. One has experience and knowledge of geology which causes him to suspect conditions not directly manifest in exploration data, and to correlate physical behavior with the results of geological processes. The uncertainties which one associates with his suspicions and correlations cannot be objectively evaluated from the records of exploration.



### Section 3.1.2 Consequences of Uncertainty

Because the geological characteristics of a site are never perfectly known, all design decisions involve risk. In considering alternative actions in design one must consider the probabilities of contingencies and the consequences of their occurrence.

One engages in exploration because by reducing uncertainty it reduces risk; and by reducing risk, reduces expected cost. If the reduction in expected cost does not equal or exceed the cost of exploration, then one is advised not to explore. If, on the other hand, by increasing exploration expenditures one can reduce the total expected cost (i.e., the sum of the expected cost in design and the cost of exploration), he should then expand his exploration program. A trade-off between exploration and risk becomes apparent.

The balancing of risk and exploration costs is nothing new; engineering geologists have long searched for ways of reducing risk. For example, Terzaghi's "observational method" is simply a way of reducing risk by reducing the consequences of unfavorable outcomes (i.e., by taking remedial action prior to failure).

In considering uncertainty and risk we must touch on another point which is essential to the concept of optimal decisions :  
A decision which leads to an unfavorable outcome is not of necessity a bad decision. Decisions under uncertainty, by definition, involve risk, and events which we hold to be improbable do sometimes occur-- otherwise, there would be no such thing as decision under uncertainty.

Say, for example, that we had very limited time to search for a suspected fault. We might concentrate our effort at its most probable locations, and in so doing maximize the probability of finding it (see Chapter 6). However, perhaps the fault actually exists where we least expect it to, and so is not found. Was the decision bad? No, the decision was good, only the outcome was bad.

### Section 3.1.3 Summary

Site exploration is commonly thought to consist of three temporal stages: Reconnaissance, preliminary exploration, and detailed investigation. The primary results of exploration are highly subjective hypotheses which transcend the physical record of exploration.

Because knowledge of local geology is never perfect, design decisions of necessity incur risk. Exploration is a process through which uncertainties can be lessened and hence risk reduced. When the expected benefit of reducing risk does not exceed the cost of exploration, the latter should not be undertaken.

### Section 3.2 Probability Theory In Site Exploration

The applicability of probability theory to site exploration lies in its facility for handling two questions:

1. How much effort should be expended in exploring a site and how should it be allocated to return the most information?

2. What inferences should be drawn from exploration data and what are the uncertainties in those inferences?

The first question involves statistical decision models; the second involves statistical inference.

### Section 3.2.1 Decision Models

To approach the first question quantitatively we must develop an analytical decision model of the site exploration process. Models are ordered sets of assumptions about complex systems, and an engineer uses models even if he makes decisions intuitively (we might call these "mental" models). The advantages of analytical models are that: 1. All simplifying assumptions are explicit, hence open to criticism and revision; 2. actions consistent with, and logically following from the engineer's preferences and judgemental feelings can be derived analytically.

Two problems which must be addressed in constructing any model are what we shall call the "boundary" and the "sub-model independence" problems. Real world processes depend on numerous complicated, interacting variables, and a model, whether it be mental or analytical, cannot account for all of them: Somewhere limits must be set on the variables considered. Determining limits is sometimes called the boundary problem (e.g., see Noren, 1969). For example, decisions made long before specific sites are identified may affect exploration, so may the possibilities of post-construction monitoring, and so may the financial situations of the engineer and client; yet one does

not normally consider these variables when constructing strategies for exploration. In addition, one faces the problem of independence between sub-models. When sub-models are independent, global optima are simply the aggregates of sub-model optima, hence we may treat each sub-model optimization in isolation. When independence does not hold, global optimization can only be achieved by considering interaction; thus complicating both the model and the computations. As before, an advantage of analytical models is that assumptions dealing with independence (whether they be assumptions of independence or of the functional forms of dependence) are explicit and open to scrutiny. In making intuitive decisions an engineer may openly assert that sub-models are dependent, yet base his action on mental models considering components (perhaps subconsciously) in isolation.

A decision model for exploration must consist of sub-models which group problems having similar analytical bases, and which may be treated using similar analytical tools. Furthermore, to increase tractability the sub-models should be as mutually independent as possible, and possess more-or-less precise boundaries. The common, intuitive model of exploration, based on three temporal stages, does not satisfy these requirements; instead, a problematic view of exploration is required.

We propose a five part, problematic model of site exploration based on what we consider to be the five major aspects of exploring rock masses: Reconnaissance, pattern recognition and reconstruction,

search, joint surveys, and in situ mechanical testing. The model we propose is in no way a replacement for the common, temporal model; it is merely another way of looking at the same process. We have ignored decisions made prior to site selection (i.e., we treat only the problem of exploring one site), and after the start of construction. These boundaries correspond to those normally accepted as defining the exploration process, and mark natural junctions beyond which factors other than description of geology become increasingly important. The five sub-models are not entirely independent, nor are their boundaries absolutely precise; yet the facility of treating specific problems within each sub-model by similar analytical tools seems to outweigh these disadvantages.

We will now individually define each sub-model, or problem class composing the overall exploration model. Although certain of these problem classes may apply to all site exploration, in this study we were concerned only with the exploration of rock masses.

1. Reconnaissance:

Problematically, reconnaissance is the stage in which previously existing records (such as published maps and airphotographs) are reviewed, on-site inspections are possibly conducted, and initial hypotheses are formed about local geology. It includes precisely the same endeavors as the "reconnaissance" stage of the three stage temporal model.

## 2. Pattern Recognition and Reconstruction:

Under pattern recognition and reconstruction are grouped those problems which treat the mapping of local geology. Patterns are groups of objects which can in some way be treated alike, and pattern recognition is the decision process of classifying objects according to pattern. We have defined the term reconstruction to mean the process of predicting unobserved portions of objects on the basis of observed portions and pattern recognition.

A typical problem of pattern recognition and reconstruction is profile mapping. In profile mapping one first classifies the materials and structures observed (recognition), then interpolate between and extrapolate beyond them to produce sectional maps.

## 3. Search:

Search treats problems of finding anomalous conditions suspected on the basis of judgement and local geology. For example, an engineer usually suspects the existence of solution cavities in limestone terrains, and at some point during exploration allocates effort to find them, should they exist.

## 4. Joint Surveys:

Rock masses contain fractures called joints, along which there has been extremely little or no movement. Joints are usually aligned to sub-parallel sets, and influence the engineering properties of

rock masses. Frequency of jointing, measured in numbers of joints per unit of rock, is referred to as "intensity."

Joint surveys are sampling programs whose purpose is to determine the distribution of joint orientation and intensity. This is the one problem area in site exploration to which the applicability of statistical techniques has long been recognized (e.g., see Müller, 1933).

#### 5. Engineering Properties Testing:

Ultimately, an engineer requires estimates of engineering properties (strength, deformability, and permeability). He obtains information about these properties by taking specimens for laboratory testing, and by conducting tests in situ. Since the results of laboratory testing are of limited value in the design of structures in rock, in situ testing is the primary concern in this phase of exploration.

We must repeat that this problematic model is not a replacement for the classical three-stage temporal model; it is but another way of viewing the same process. Fig. 3 shows how the two models might relate in time.

#### Section 3.2.2 Inference

The second application of probability theory to exploration is that of inference: What conclusions do we draw from given

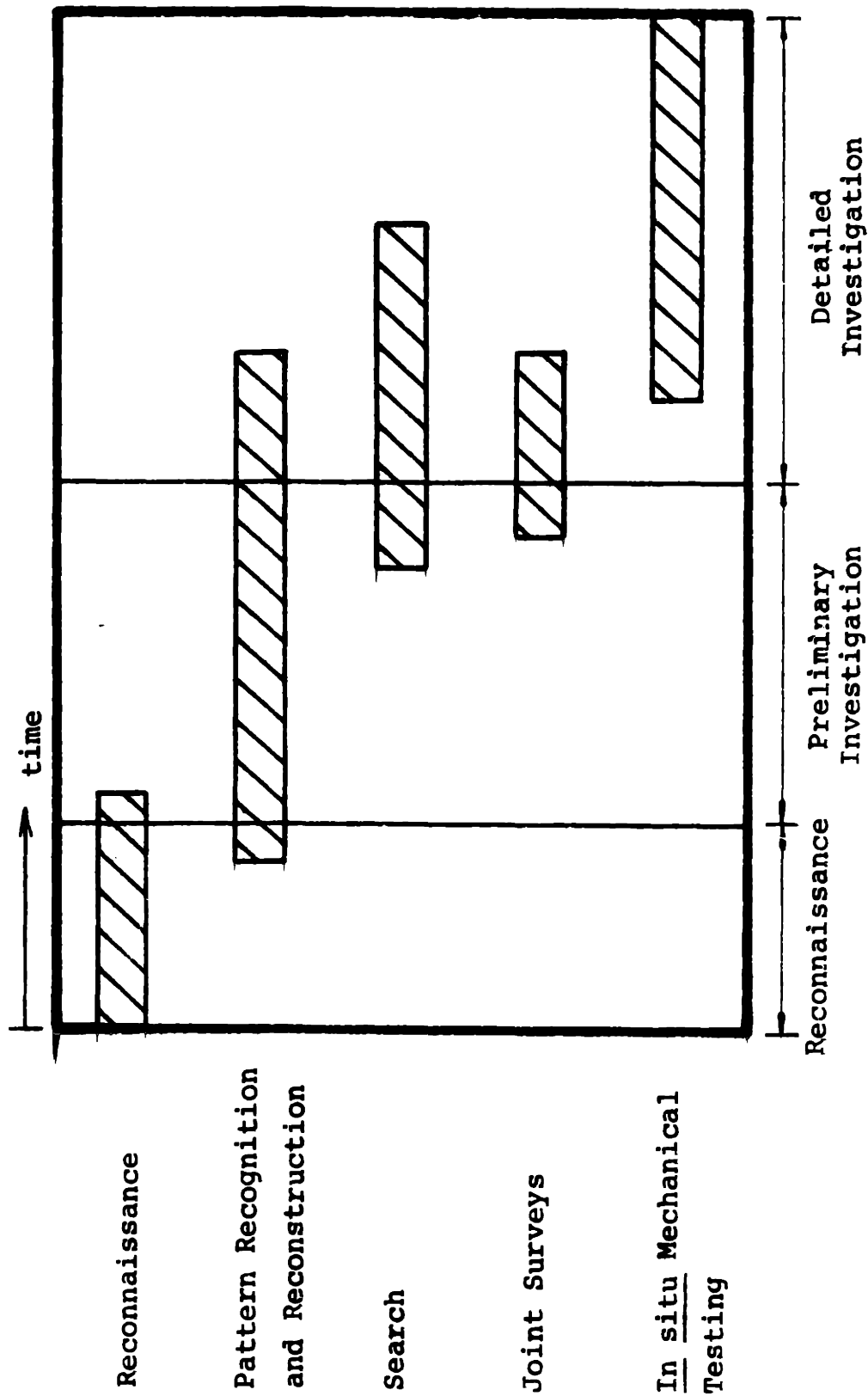


Figure 3: Problematic and temporal organizations of exploration in rock masses (lengths are solely descriptive)



exploration data? Inference and strategy decisions are not separate, for the choice of optimal actions depends on the inferences which can be made from resulting data.

The use of statistical theory in inference is widely recognized, and inference rather than strategy optimization first led the profession to interest in applying probability theory to exploration. Even today, most work on the application of probability theory to site exploration treats problems of inference (e.g., "On the basis of  $n$  measurements what is the average joint set orientation, and is this different than the average orientation at location  $x$ ?"), rather than strategy optimization.

The techniques of statistical analysis are aids which an engineer may employ in drawing inferences. They do not simply output conclusions from some enormous magnetic memory. Ultimately, every inference or conclusion an engineer makes depends on the hypotheses he subjectively induces. If he does not induce the correct hypothesis it will never be considered.

Fig. 4 depicts a situation, common in site exploration, in which observations could be the manifestation of more than one physical reality. Statistical analysis will not keep an engineer from reaching wrong conclusions. He must know from his experience to consider the degree-of-confirmation given hypothesis  $H_a$  as well as that given hypothesis  $H_b$ .

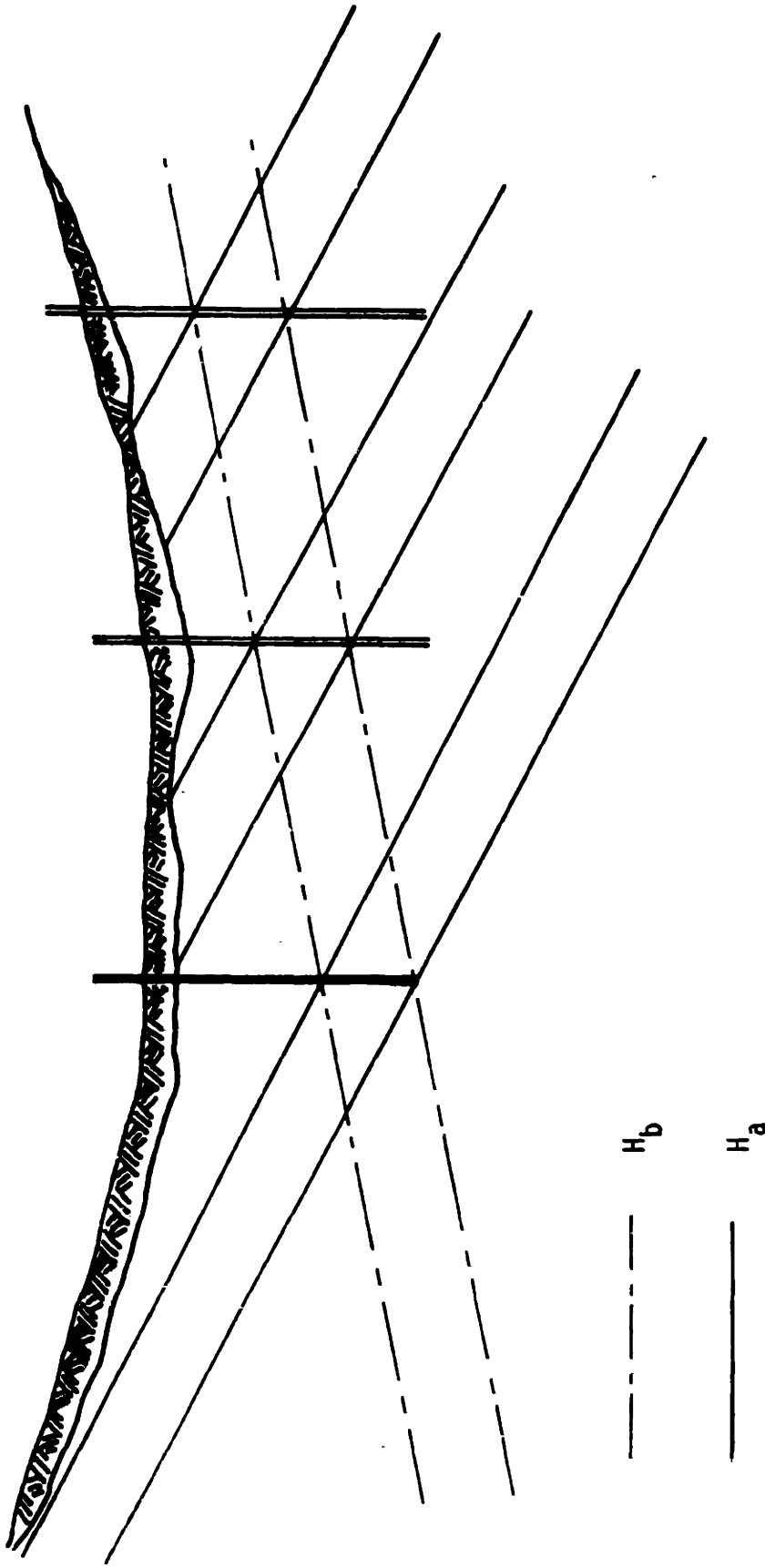


Figure 4

Alternate Hypotheses Explaining The Same Observations

### Section 3.2.3 Bayesian Methods

Judgement is critically important in site exploration for as we have said, the tangible results of exploration are meager compared to the inferences made. Furthermore, drawing inferences about a site is a one-time event. Speaking of exploration probabilities in terms of frequencies (e.g., "if we had 100 sites like this one we could expect a fault to exist at 60 of them") is meaningless and misleading. Therefore, site exploration essentially necessitates a subjectivist degree-of-belief approach to probability.

With minor exceptions (e.g., Folayan, et al, 1970; Turkstra, 1970) published applications of probability theory to exploration have dealt with classical sampling theory, confidence limits, numbers of samples required for given levels of significance, and similar relative-frequentist concepts. In addition, published accounts have considered material properties almost exclusively. These are not the real problems of site exploration. The real problems involve strategy, optimization, inferences about geometries, degree of belief that certain adverse conditions exist undetected, and the like. The amount of effort one is justified expending in exploration has little to do with arbitrarily set confidence limits. One seeks confidence limits and associated sample sizes not because they are innately desired, but because they are the sorts of outputs classical analysis gives. Fundamental advances in the application of probability theory to exploration will only come when one

includes judgement and addresses problems of strategy optimization.

Acceptance of subjective degree-of-belief probability leads to the use of Bayesian decision theory (Chapter 1). Bayesian decision theory is a normative model for behavior under uncertainty; an aid to decision making which structures alternatives, incorporates observations, and relates judgement to other parts of exploration decisions. It is not a replacement for judgement.

One might think of decision theory in somewhat the same way as he does elastic theory. We do not expect the prediction of elastic theory to be precisely realized in a real structure because the assumptions underlying its specific application do not precisely correspond to the real world. Yet, elastic theory identifies critical considerations, indicates the sensitivity of design to particular parameters, and generally structures design problems.

Engineers are, above all else, pragmatists. Although, philosophically, one might assert that degree-of-belief probability is the only proper way (probabilistically) to treat exploration problems, practically, one can identify classes of problems which are satisfactorily, and more simply, treated using relative-frequentist techniques (joint surveys is one of these). We have not hesitated to use classical methods where they seemed justified.

### Section 3.2.4 Size of Project to Which Probability Theory is Applicable

Decisions are not always important enough to warrant a great amount of attention. If the maximum loss incurred by making a poor decision is small, or if formal analysis cannot change an elected course of action, there is no sense wasting time with decision models. However, as decisions become more important, considering them analytically becomes worth our time.

There is no way a few precise, all-inclusive rules can be stated for when and when not to use probability in site exploration. Inference and decision models are only applicable when they meet the two criteria stated above. A few examples will illustrate this point.

There is no reason to use decision theory in locating borings for a moderate-size, midtown office building. No matter the findings, 3 or 4 borings will be placed, roughly at the corners of the site, because this procedure is easy to specify, and a legal safeguard.

Using decision theory in exploring a dam site is certainly appropriate. The cost of decision analysis is small compared to the savings it might lead to by increasing the efficiency of effort allocations, and to the consequences of large uncertainties in design. Furthermore, the range of exploration programs is wide and analysis may have a major effect on elected actions.

Between these extremes, whether or not to use analysis is less clear, and depends on special circumstances. Analysis seems

appropriate in exploring a proposed highway right-of-way through geologically heterogeneous, mountainous terrain in which, say, fractured and faulted zones are expected. Yet, in exploring the same right-of-way through an adjacent lacustrine valley analysis seems inappropriate (since it probably won't alter the elected action, e.g., "place one boring every 2000 ft.").

Using decision theory to decide whether or not to use decision theory leads one into a never ending process which might just be as well handled by intuitively considering the two criteria discussed above.

#### Section 3.2.5 Summary

Probability theory finds application to site exploration in treating problems of strategy and problems of inference. These problems are not independent.

The application of probability theory requires a problematic description of the exploration process for which we have suggested a five part model: reconnaissance, pattern recognition and reconstruction, search, joint surveys, and mechanical testing. This is not a replacement for the common temporal model, merely another way of looking at the same process.

Because exploration decisions and inferences are "one-time" events depending heavily on judgement, subjectivist-degree-of-belief probability is the most appropriate probabilistic approach.

Pragmatically however, one can identify classes of problems which are satisfactorily and more easily treated by relative-frequentist methods.

The size exploration to which analytical methods are applicable depends on two considerations: the marginal expected cost of poor decisions, and the possibility of affecting elected courses of action.

### Section 3.3 Goals of Present Work

We address in this work the question of whether or not probability theory, beyond sampling statistics, is applicable to site exploration in engineering geology; and if it is applicable, to what problems of exploration can it be applied? Although engineers allude to the appropriateness of its use, no one had previously undertaken a systematic, comprehensive evaluation.

This is a venture into a relatively untouched subject and of necessity the work is incomplete. An implicit goal has been to construct a broad foundation upon which future work can be based.

Satisfying this twofold goal required that three intermediate goals be considered. The first was to devise a problematic approach to exploration which defined groups of problems having similar analytical basis. This we have already discussed.

The second, was to develop measures of effectiveness for effort allocations. If one cannot measure effectiveness, he certainly cannot optimize allocations. Once measures were developed they could be

related to utility, and Bayesian decision theory used to determine optimal strategies.

The third was to obtain mathematical tools for optimizing effort allocations and for drawing inferences. In this endeavor we made as much use as possible of work in other disciplines, since there seemed no reason to spend valuable time developing analytical tools already in existence. In most cases these tools had to be modified or extended to solve our problems, and in some cases new derivations were required.

### Some Limitations

We have treated each problem class in this work as if it were independent. For example, we have spoken of allocating search effort as if search were the sole purpose. In real explorations, of course, we allocate effort to accomplish several goals simultaneously. Information from borings is used to construct maps, search for anomalies, and test material properties. We fully recognize the limitations imposed by assuming independence, but introductory research must start somewhere, and this simplification allowed a more comprehensive inquiry.

We have considered only geometric problems of exploration, reflecting a conviction to the importance of these problems and limits to the amount of material any one investigation can include. Strategies for in situ testing present problems of interest both



practically and academically, but would have required more time than was available for inclusion here.

There seems to be a feeling prevalent among orthodox engineering geologists that one working with probability theory should be able to discuss in an hour all that is required for them to use that theory in practice. We find this an untenable belief. No one would expect to apply other branches of mathematics competently with so little introduction, and there is no reason to consider probability theory any differently. This feeling has probably been given weight by the fact that one may apply such numerical tools as the finite element method with little knowledge of the mathematics used in formulating programs. However, one must note that such techniques treat problems and theories (e.g., elastic theory) with which the engineer is very familiar. Generally, the engineer is not familiar with probability theory. The philosophical basis of probability alone is of great importance, and must be comprehended before applying the theory to practical problems. Because of our belief in the importance of this philosophical basis we have devoted the entirety of Chapter 1 to an introductory review.

Papers on the applications of probability theory, even ones specifically written for the uninitiated, are often overwhelming in their use of mathematics and technical wording. We have made a conscious effort in this work to use plausible reasoning and examples of the methods discussed; we have done so at the risk of sacrificing

mathematical rigor and of sometimes presenting examples which seemed contrived. Such a policy was judged appropriate for the purpose at hand. In the future, it will be appropriate for other workers to provide a more rigorous foundation for the work presented here.

In any work presenting new ideas and applications, one would like to test his abstract thoughts through case studies. However, every project must stop somewhere, postponing work one would like to have done. As we shall emphasize in the next chapter (4), a logical continuation of this work is to apply the more refined of the methods presented here to field cases.

## CHAPTER 4

### CONCLUSIONS AND RECOMMENDATIONS

#### Section 4.1 Conclusions

To investigate the use of probability theory in the exploration of rock masses one requires a model of the exploration process. We have proposed a five part problematic model which groups together specific exploration problems having similar analytical bases. These five problem classes are: reconnaissance, pattern recognition and reconstruction, search, joint surveys, and mechanical testing.

Probability theory seems applicable to four of the five classes: reconnaissance, being a purely inductive undertaking seems outside the realm of logical analysis. In this thesis we have considered the three classes treating geometric information, and have left the fourth (i.e., mechanical testing) to future work.

The conclusions of this study are divided into two parts: general, dealing with overall conclusions; and specific, dealing with each problem class individually.

#### Section 4.1.1 General Conclusions

##### 1. General Applicability

Probability theory, beyond sampling statistics, is generally applicable to optimizing strategies and drawing inferences in site

exploration. Techniques now exist which may be modified or extended to treat many problems of exploration. However, since exploration is founded in inductive reasoning, some specific problems (particularly those associated with mapping) will require considerable theoretical effort before solutions become available.

## 2. Importance of Judgement

Because exploration is highly dependent on judgement, and involves "one-time" events, it essentially necessitates a subjectivist degree-of-belief approach to probability. Current use of probability theory in exploration seems to be heading in the wrong direction by using relative-frequentist techniques, and by not addressing problems of strategy.

If one admits the use of subjectivist degree-of-belief probability, Bayesian decision theory can be used to structure and optimize decisions involving risk. This theory is a normative model of behavior, and serves only as an aid to the decision maker. It is not a replacement for judgement.

Quantifying judgement for use in Bayesian decision theory is not a trivial task. Since subjective probabilities are not necessarily intuitive even to the decision maker, they must be inferred from his behavior in simple situations involving risk. Empirical evidence suggests that our techniques for doing this result in rough approximations of the decision maker's true judgemental feelings.

### 3. Dependence of Sub-Models

Effort is allocated in site exploration to simultaneously gather information on several facets of site geology. Although the work contained here considers problem classes independently, we see no prohibiting reason why an interacting model (i.e., one which does not assume independence) cannot be developed. Obviously, such a model would be an order of magnitude more complex than those discussed here, and might require multi-dimensional utility functions. We discuss this further in Section 4.2.

### 4. Sequential Decisions

Advantages can be gained in exploration by using sequential strategies, but optimizing these strategies is difficult. There exist no solutions at present other than directly enumerating actions and consequences of each stage in the sequence (i.e., dynamic programming or preposterior analysis), and for processes of more than a few stages and actions, direct enumeration is infeasible. Fortunately for the cause of site exploration, search is one of the few processes for which sequential decisions can be optimized analytically.

#### Section 4.1.2 Specific Conclusions

##### 1. Pattern Recognition and Reconstruction

Pattern recognition and reconstruction is dependent on inductive reasoning, and analytical models are generally difficult to formulate. Given certain prior information, practical techniques are now

available for handling recognition problems; however, reconstruction models are only developed for the simplest cases (i.e., "proximity"-type rules). Further development of reconstruction models awaits satisfactory methods for handling trending bodies.

A significant problem in reconstruction is overcoming computational difficulty. While the rules discussed here are straightforward, they lead to lengthy computations when strategies are optimized.

We have proposed a system of quantitative mapping which associates an ordered n-tuple of probabilities with each point in the space mapped, and have introduced a measure-of-effectiveness based on Shannon's entropy function. Although this measure does not contribute appreciably to computation, its relationship to design costs and other utilities is not clear. At present we have no way of relating uncertainty in geological mapping to utility in design and construction, with the exception of cases in which the design decisions to be based on mapping information are specifically known prior to exploration.

Reconstruction models developed sufficiently for field use will require considerable work, both theoretical and mathematical. However, the value of those models is also considerable: effort allocations for field mapping cannot be optimized without them. For short-term benefits, though, research effort would be spent more wisely on problems other than reconstruction.

## 2. Search

Due to heavy emphasis on search and detection for military purposes, solutions exist which can be adapted to fit most exploration problems. In contrast to pattern recognition and reconstruction, theories of search are developed to the point that field applications are in order.

In this thesis we have considered four measures-of-effectiveness for search, and have shown that each leads to the same optimal allocation in any specific case.

Use of grid patterns for search is straight-forward, although overlapping geometries complicate computations for small grid-spacing to target-size ratios. Tables and charts of probability-of-a-find as a function of grid spacing and target size exist for limited combinations of grid and target shapes (one can show solutions to be insensitive to the latter).

When the probability density function (pdf) of target location is not uniform, strategies can be devised which offer a greater probability-of-a-find than do grid patterns. Optimal non-sequential (which we have called "single-stage") allocations can be determined using simple graphical methods. The assumptions upon which these methods are based are acceptable for most exploration problems.

Fortunately, search is one of the few sequential decision processes which can be optimized analytically. Optimal sequential strategies of search are myopic; that is, they specify to search the location at each stage with the greatest current ratio of probability-

of-a-find to cost.

The sensitivity of optimal allocations and resulting probabilities-of-a-find to the prior pdf of location have not been investigated, although the author suspects low sensitivity for minor changes.

Comments throughout this work concerning difficulties in assessing prior subjective pdf's apply equally to search.

### 3. Joint Surveys

The importance of using statistical methods in joint surveys has been recognized for several decades. Nevertheless, the use of these methods has generally been haphazard, and most workers have not taken full benefit of existing statistical theory. Joint Surveys is one problem class of exploration in which the use of relative-frequentist methods is appropriate, and techniques and sampling plans presented in this thesis may be used in practice with little or no additional development.

Ruth Terzaghi's widely known work, and other investigations of joint sampling, follow directly from statistical sampling theory, and we can use that theory to generalize and extend these previous studies (e.g., by developing weighting factors for entire joint surveys rather than for each outcrop; and by developing self-weighting plans for sampling individual outcrops). A sampling plan for joint surveys is suggested. From both statistical and empirical considerations, sample sizes in the range 100 to 150 should be used. The similarity between populations sampled (e.g., joints at the ground surface) and populations whose characteristics are desired (e.g.,



joints at the depth of a proposed gallery) is purely a geological question and cannot be treated statistically.

Several statistical techniques can be used for testing the hypotheses that no preferred orientation exists and that two or more joint populations (e.g., from different parts of a site) are, in fact, the same.

Mathematical models can be used to obtain smooth, continuous summaries of observed data; however, there is no theoretically correct distribution for joint populations and extrapolation beyond the data set (as is now sometimes done) is a dubious undertaking. A fitted analytical model must conform to the topological constraints of orientation. Since orientation is defined on a unit hemisphere, use of the bivariate normal distribution (which is defined on an infinite plane) does not satisfy these constraints -- despite trends in the literature toward its use. At present the mathematical properties of hemispherical functions are not well known. The spherical analogue to the normal distribution is presented and statistical tests associated with it are discussed.

#### Section 4.2 Recommendations for Additional Study

A goal of this thesis has been to establish a foundation for the study of probability models in exploration. Frankly, we are overwhelmed by the breadth and promise of future investigation, and in this section we suggest fertile avenues of study.

Like the previous section, this one is divided into general and

specific recommendations, but additionally, we have added a subsection in which intriguing topics related to, but not specifically part of the present investigation are suggested.

#### Section 4.2.1 General Recommendations

##### 1. In Situ Mechanical Testing

In situ mechanical testing is the one problem class of exploration not considered in this thesis. Since large sums of money are expended for such testing, the optimization of effort allocations may lead to large savings.

Conclusions drawn from in situ testing are directly utilized in specific design decisions, and the uncertainty in these conclusions is directly manifest in expected utility. Therefore, the problems of constructing a decision theoretic model which plague mapping endeavors should not hinder the development of analytical techniques for allocating in situ testing effort. Furthermore, one suspects that widely applicable heuristic rules may result from a study of in situ testing, which may be used when precise analysis is not justified (e.g., should one perform a few large tests, correlate index properties to them, and measure the latter at many locations; or simply perform several large tests?).

Two considerations, which have been of limited importance in this thesis, will be of major concern in investigating in situ testing. One is that allocations must be made over testing method as well as space and time; the other is that mapping continuous

attributes and inferring their value at unobserved locations will be required for spatial sampling optimization.

## 2. Interacting Models

With the development of techniques for analyzing mechanical-testing strategies and inferences, we will be in a position to begin work on interacting models; that is, models which do not assume independence among the four problem classes. The central problems will be constructing utility functions related to information on all of the four aspects of exploration, and developing mathematical tools for optimization.

## 3. Utility Functions

We allude throughout this work to utility functions relating exploration information and costs to design costs, risk of failures, the subjective value of professional reputation, and the like. Investigation of these functions, their assessment, and relationship to uncertainty about local geology is needed whether or not quantitative decisions are made.

## 4. General Applications

This thesis presents techniques for rationally handling many specific problems of exploration. Profitable future work of both professional and academic interest would be to apply available techniques to case studies. Feedback from case studies will indicate deficiencies either dismissed or overlooked in theoretical analyses, and hopefully support the use of general probabilistic methods in exploration.

## Section 4.2.2 Specific Recommendations

### 1. Pattern Recognition and Reconstruction

The theory of analytical mapping requires considerable general work before becoming of practical use. However, advances in two areas would further the theory immensely. First, the development of trend rules for inferring geological classifications at locations not observed. Second, a method for directly relating utility to uncertainty in mapping.

Recognition theory is well advanced, and application to problems involving large numbers of classification decisions (e.g., borehole logging by remote sensing) would be interesting.

Stochastic process models are receiving increasing attention in soil mechanics with specific reference to material properties. As yet these models have not been used to answer problems of exploration (i.e., those of strategy and inference), but have been used to treat design problems exclusively. Sophisticated methods exist for predicting spatial variables at unobserved sites and functions of spatial variables over regions. These should be tested by predicting actually observed field data, and applied to practice. One such application is to ground water level problems and the impact of constructed facilities.

### 2. Search

We could neither foresee nor treat all of the specific problems of search arising in practice, but the general theory of search

can be easily modified to handle most practical problems. Future work should concentrate on applications since this avenue will lead to quickest refinement. Successful application would also be convincing evidence for the general use of probabilistic methods, thus spawning interest in the profession and generating feedback from orthodox practitioners.

A theoretical question of importance, but not considered here, is what we have called "partial information observations"(p.i.o.). This problem is sure to arise in application so must at some time be treated. A p.i.o. is one which, although not finding a target, gives direct evidence of its location. By direct evidence we mean beyond that contained in knowing that a target does not exist at some location.\*

For example, in searching for serpentinized zones, the encounter of other alteration products leads one to believe he is "getting warm." The p.i.o. problem shares much with reconstruction and may be difficult to rationalize for the same reasons. Solutions of this problem will first require a measure of evidence contributed by p.i.o.'s, and strategy formulation might be aided by response surface techniques.

Mathematical proof that total cost is a unimodal function (i.e., has a unique minimum) of the number of exploration stages in sequential

---

\*By Bayes' Theorem, information that targets do not exist at certain locations causes changes in the probability density function, and thus is indirect evidence of location.

search would mean that computations could be terminated once a local minimum was found.

### 3. Joint Surveys

Generally, effort would be more wisely spent on exploration problems other than joint surveys, since work on the statistics of sampling joint populations is showing diminishing returns. Until design procedures which make use of joint orientation distributions are refined, there is little need for further theoretical work. What is needed, however, is the application of statistical sampling techniques to practice.

To facilitate application, computer programs should be developed which would: (1) analyze sampling data by computing weighting factors and various sample statistics, and (2) compute moments of orientation distributions directly from bore hole data.

Unless one uses oriented cores, the only measure of joint orientation obtained from a boring is the angle of intersection of the joint plane and boring axis. Given three non-collinear borings, the strike and dip of a single joint plane can be uniquely determined. At present, engineers assess the average orientation of a joint set by using the single plane solution in conjunction with the average angle of intersection with each boring. Unfortunately, since the expression for expected orientation is a function of the expected cosine of angle of intersection (rather than the cosine of the expected angle) this practice does not give the correct average orientation. A program to compute mean and variance from the distri-

butions of angle of intersection would save the engineer a time consuming task.

If analytical models of joint orientation are ever to be widely applied -- making the assumption that more sophisticated design methods are developed -- mathematical properties of the asymmetric hemispherical normal distribution must be investigated. Furthermore, we should know more of the empirical behavior of orientation distributions. Are hemispherical normal distributions good models? How do the "tails" of real distributions behave? Empirical information can be obtained by taking large samples which conform to the requirements of statistical sampling.

#### Section 4.2.3 Topics of Related Interest

##### 1. Search for Construction Sites

Siting large facilities is, and will continue to be an issue of intense public interest involving technical and social problems spanning many disciplines.

Strategy decisions for locating sites are predicated in search theory (e.g., "which regional areas should be searched, which specific sites investigated, and when should search be stopped?"); however, the utilities one deals with have several attributes -- some certain (e.g., transportation distances of raw materials and products), and some uncertain (e.g., environmental, political, and economic impacts). Forming decision models for search and selection

would be intriguing and is of immediate importance.

## 2. Engineers' Judgement and Intuitive Decision Making

One uses judgemental feelings, or what can be called subjective probabilities and utility, in all decision making, analytical or intuitive. The techniques of statistical decision theory allow us to infer apparent subjective probabilities and utilities from past decisions, and hence are tools with which to study judgement. By quantifying apparent judgemental feelings we can critically evaluate the predicates of intuitive decisions.

Of questions which might be addressed, two are of most importance. First, as a whole, are engineers' feelings about local geologies externally valid? We have limited evidence suggesting they are not. We have also evidence from experimental psychology suggesting consistent trends of misassessment (as inferred from behavior) for people in general.

Second, what is the structure of utility functions under which decisions are now made? Whose utility is maximized?...the client's? ...the engineers'?...the user's? What weight is given to the engineer's reputation? ... to inconvenience for the user? ... to schedule slow-downs? Questions of utilities involve questions of ethics; unless we can assess the predicates of decisions we cannot assess the degree to which normative maxims (e.g., ethics) manifest in practice.

## 3. Heuristic Rules

With the passage of time, soil engineers working predominantly



within a single geographical area repeatedly encounter similar projects -- both in terms of geology and facility type. For example, the design of buildings in most urban areas fits this description, because cities are usually located on broad depositional basins. Since strategy decisions for allocating exploration effort in each new case depend on similar subjective probabilities and utilities, generally applied heuristic rules usually develop. Although the use of decision analysis for each new project is infeasible, one comprehensive study could be undertaken which might result in quantitatively founded heuristic rules for allocating exploration effort. These would include facility size, location within the depositional basin, and availability of nearby construction records as parameters; and would serve as guides for projects otherwise not justifying decision analysis.

## CHAPTER 5

GEOLOGICAL MAPPING: PATTERN RECOGNITION AND RECONSTRUCTIONSection 5.1 Introduction

Perhaps the first task facing the engineer or geologist after reconnoitering a site is to "map" local geology. Conventionally, this is accomplished by observing the geological nature of a finite number of locations (unless bedrock is exposed everywhere, which is seldom the case), and inferring contacts between formations. There are two ways probability theory can aid this process. First, it can be used to make inferences about the geology of unexplored locations (and quantify the confidence that should be placed in inferences); and second, it can be used to optimize the way effort is allocated (i.e., "how do we obtain the most information per dollar?"). We would also like to use statistical decision theory to help answer the question, "how much exploration is enough," but as we proceed we shall see that unless certain well-defined information is available about the consequences of mapping errors, we will not be able to quantitatively treat this decision.

The development of a comprehensive theory of quantitative mapping is a monumental undertaking, but the rewards are enticing. Such a theory would afford us measures of potential confidence to be placed in projections, techniques for economizing expensive effort

allocations, and facility for rational design decisions based on mapped information. Such a theory might also usher in the use of machine constructed geological maps, including favorable subsequent allocations of effort as output. Haralick and Dinstein (1971) have already developed a program which clusters geographical phenomena and produces area classification maps directly from raw satellite photographs.

This chapter organizes the general task and points out specific problems, philosophical and practical. The task is much too large to hope for a total solution here; instead we pursue an approach to the problem upon which future work can hopefully be based.

The attributes which geological maps portray are of two types, discrete and continuous. A normal geological map (such as those published by the U.S.G.S.) assigns to each site location one of several discrete classes, which are commonly geological formations or lithologies. A second type of map, and one of importance to engineers, assigns a value of some continuous attribute to each location. Commonly mapped continuous attributes are ground water level, and index properties. The first part of the chapter is devoted to mapping discrete attributes, while we delay discussing continuous attributes until the latter half.

## Section 5.2 Discrete Attributes

Discrete maps are constructed by observing and classifying geology at a finite number of points (either in a plane or in space),

and inferring the geological nature of points which have not been observed (Figure 1). As such, it is very much an inductive process: we recognize patterns in site geology, and based on these patterns and prior experience, estimate geometries that are hidden to us.

We will call these two problems recognition and reconstruction, respectively. Their differences can be demonstrated by considering a problem outside of site exploration. Viewing a photograph of a person's face, a recognition problem would be,

Given the size of the mouth and separation of the eyes, whose face is it?

while a reconstruction problem would be,

Given that we've decided whose face this is, what does the rest of the body look like, and does (s)he have 10 or 12 toes?

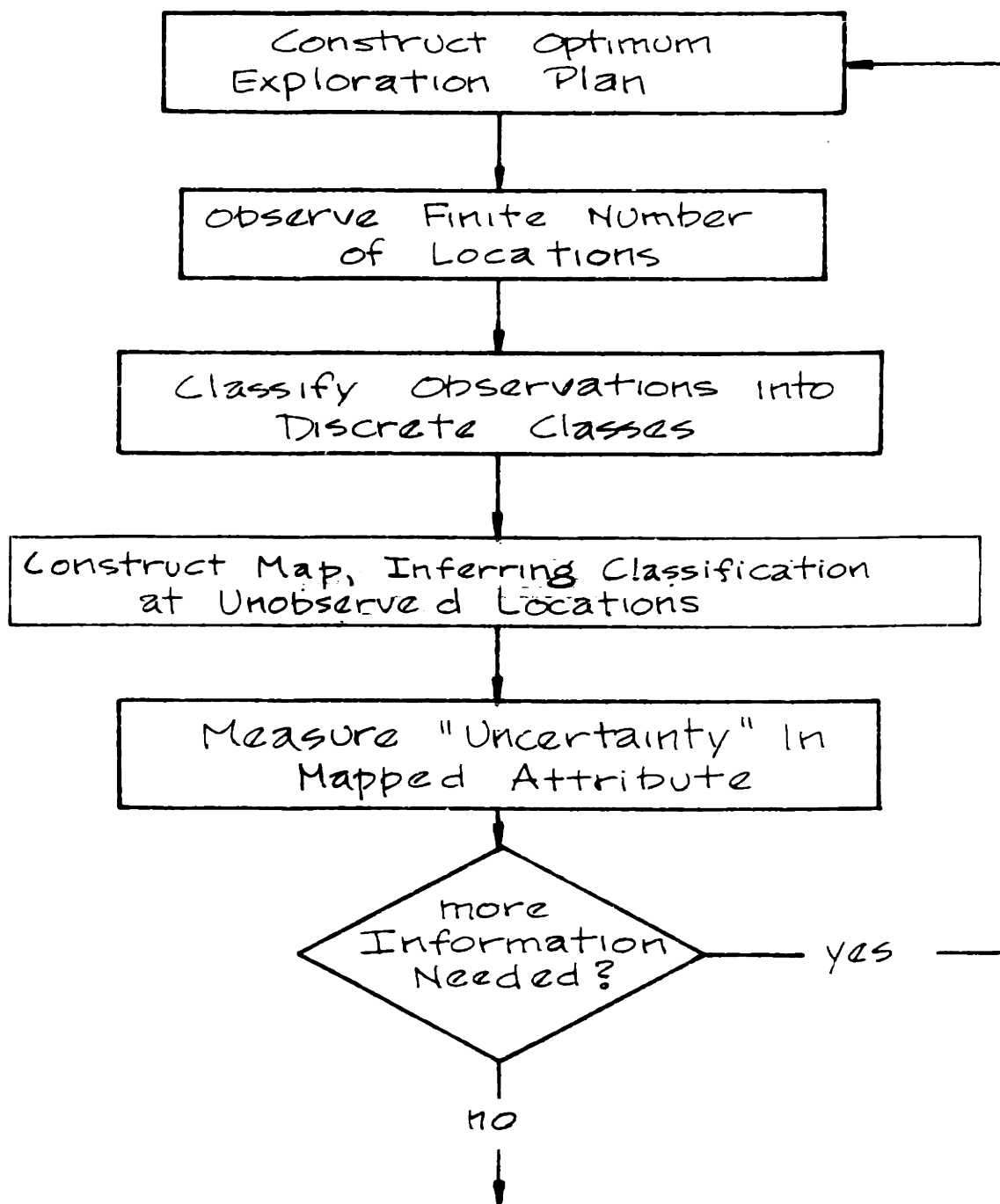
The literature, particularly that of artificial intelligence, abounds with work on pattern recognition, but contains essentially no work on reconstruction; and the latter is a more difficult question. We will consider recognition first since it occurs first in exploration, and since it is more tractable than reconstruction; then we shall discuss approaches to reconstruction.

### Section 5.2.1 Classification

Proper classification of geological bodies into classes of interest to the engineer is a necessary part of preconditioning (i.e., preparing) exploration data for pattern reconstruction. Some classification tasks are obvious and may be accomplished by inspection

Fig. 1

## Discrete Mapping Flow Chart



(e.g., logging a core run); others are not so obvious and can be aided by the use of analytical techniques (e.g., logging a drill hole on the basis of geophysical logs). Concisely, the classification problem is this: Given measurements of a limited number of physical properties on a particular body, to which geologic class should the body be assigned such that some decision criterion is optimized (e.g., most likely class, maximum expected utility, etc.)?

Figure 2 shows data from a geophysical core run in a sedimentary (sandstone, shale and limestone) formation. Assigning stratification by lithology to this run is an example of the classification problem.

When we begin discussing rules for inferring classifications at unexplored locations the problem of deciding what type formation we are dealing with (i.e., depositional or structural mode) becomes important. Since this is essentially just a problem of classification also; we will discuss it here in addition to the point observation problem.

### Basic Theory

Classification may be considered in two parts. First, certain characteristics of the entity to be classified, called factors, are isolated and measured. This is called "feature extraction." Second, the relative likelihoods of these factor measurements are evaluated, conditioned on the true classification, and the entity classified (the likelihood is just the probability of the observed factor value, given the true classification -- see chapter 1). This we will call the

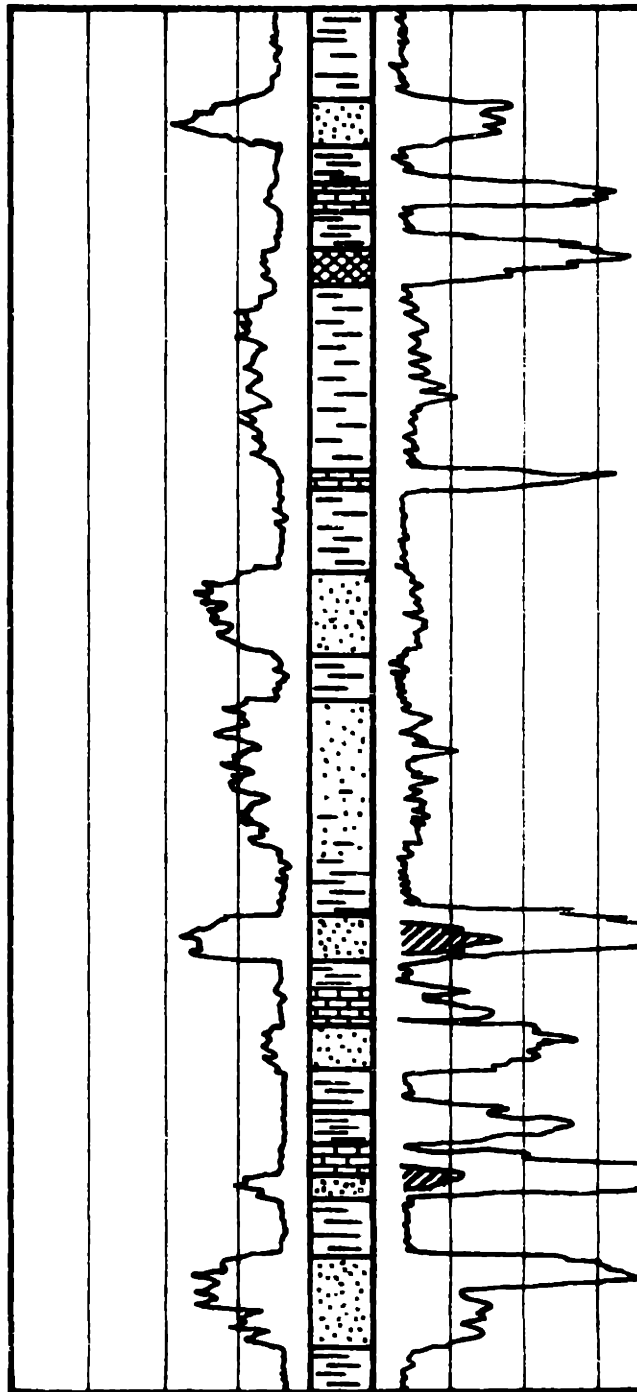
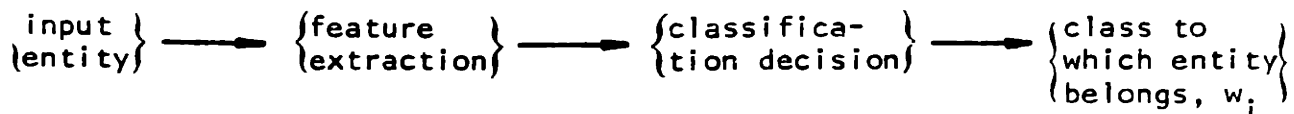


Figure 2

Geophysical log in sandstone shale-  
limestone formation (from Krumbain and  
Sloss, 1963)

"classification decision" (Figure 3).

Figure 3



Let there be  $n$  classes to which an entity may belong,  $w_1, \dots, w_n$ ; and  $m$  factors measured for classification,

$$\underline{x} = (x_1, \dots, x_m)$$

(where  $\underline{x}$  is the vector of the measurements). The prior probability of an input pattern belonging to class  $w_i$  is  $p(w_i)$ ,  $i = 1, \dots, n$ , and the conditional probability (likelihood) of the measurement vector  $\underline{x}$  given the class  $w_i$  is  $p(\underline{x} | w_i)$ ,  $i = 1, \dots, n$ .

The process of classification is simply one of statistical decision: classify an input entity, based on the prior probabilities and likelihoods of observation, such that a desired criterion (e.g., expected utility) is optimized. The problem may be expanded to include costs associated with feature extraction without altering the decision theory basis.

### Development of the Theory

Once a set of factors has been isolated and measured (we will return to the problem of which factors to measure) the immediate step is to evaluate the probability of the entity belonging to class  $w_i$ ,  $i = 1, \dots, n$ .



$$p(w_i | \underline{x}) = \frac{p(w_i) L(\underline{x} | w_i)}{\sum_i p(w_i) L(\underline{x} | w_i)} \quad 1$$

Adopting a Bayesian decision approach, we then wish to classify the entity such that expected utility is maximized.

If the disutility of each type misclassification is the same, minimizing the probability of misclassification maximizes expected utility. This probability is minimized by assigning the entity to the most probable class (i.e.,  $\max_i p(w_i | \underline{x})$ ). Since the denominator of the right hand side of 1 is constant, equivalent criteria are,

$$\max_i ( p(w_i) L(\underline{x} | w_i) ) \quad 2$$

$$\max_i ( \log p(w_i) + \log L(\underline{x} | w_i) ) \quad 3$$

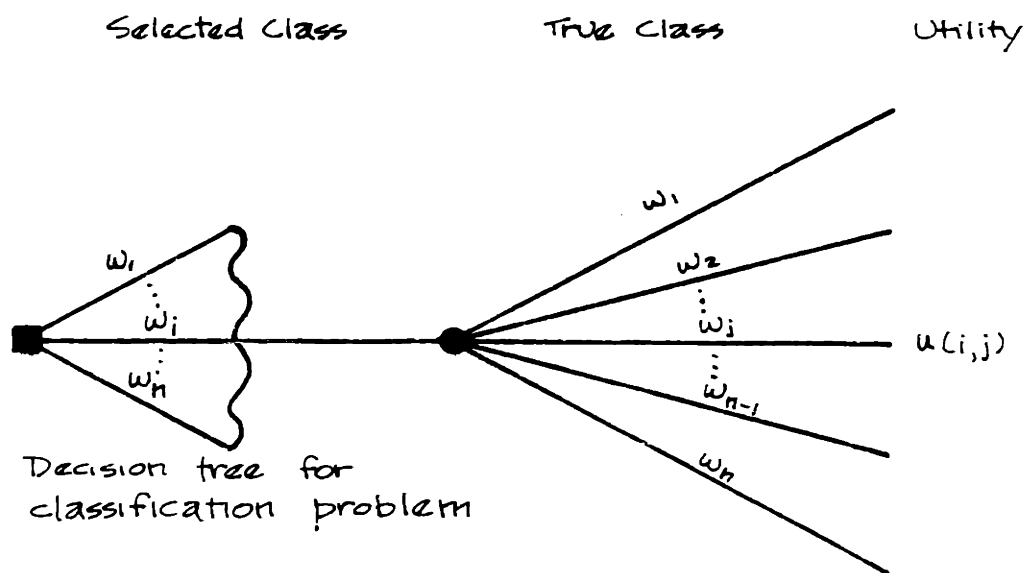
the latter being convenient when the likelihood is exponential.

If the disutility of each type misclassification is not equal, then the Bayes decision is: Decide  $w_i$  such that

$$E[\text{utility}] = \max_i [ \sum_j p(w_j | \underline{x}) u(i, j) ] \quad 4$$

where  $u(i, j)$  is the utility of deciding  $w_i$  when the actual class is  $w_j$  (Figure 4).

Fig. 4



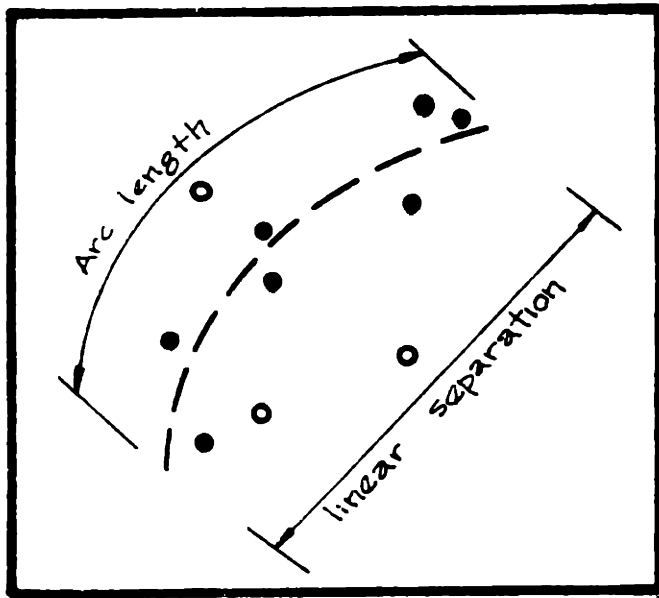
The basic theory of classification is uncomplicated, but before going to elaborations a simple example is in order.

Shoestring sands are elongated sandstone deposits formed by the burying of ancient streams ( $w_1$ ) or off-shore bar deposits ( $w_2$ ). The mode of deposition is quite important in exploration because the plan shape of the two deposits is different: Bar deposits resembling a low order polynomial arc and stream deposits being sinuous. Although there are several features which could be extracted for classification, for clarity we will consider only the ratio of arc-length along known locations of the body to the linear separation of its end points (Figure 5). We will call this ratio the "straightness-index" (S.I.).

From empirical data on shoestring sands the frequency distribu-

tions of straightness-index for  $w_1$  and  $w_2$  could be evaluated as schematically shown in Figure 6.

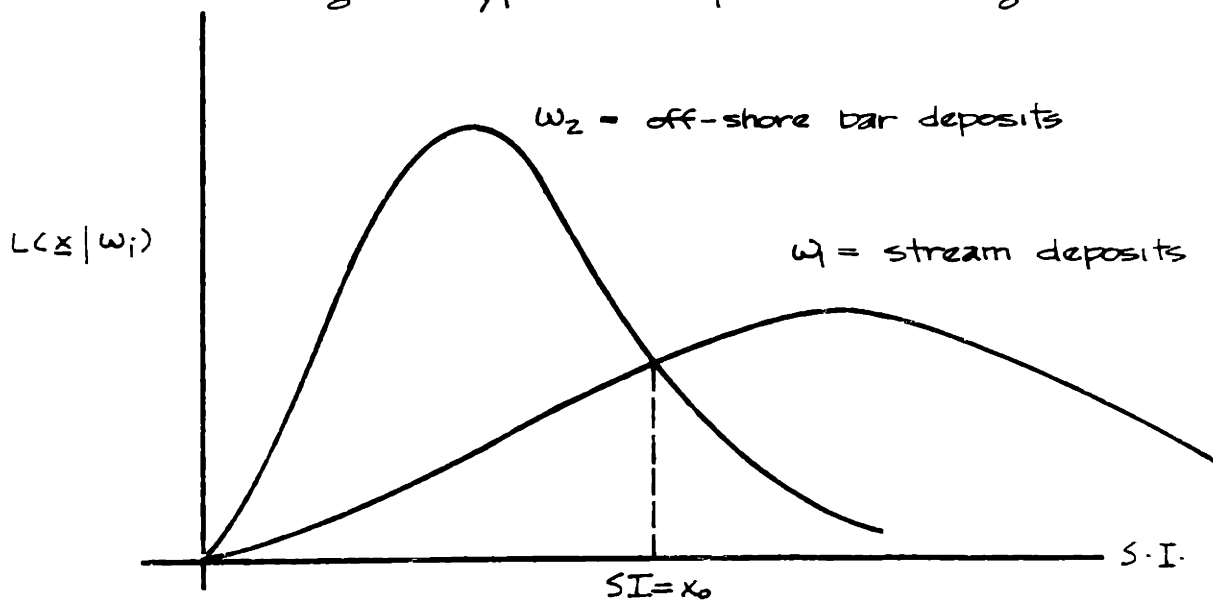
Fig. 5 : Definition of "Straightness Index"



- known locations of body
- locations of other formations

$$S.I. = \frac{\text{arc length}}{\text{linear separation}}$$

Fig. 6: Hypothetical pdf's of straightness index



Following our classification procedure, assuming the utilities  $u(i,j)$  to be all equal, the approximate S.I. of the body in question is measured and the body is assigned to the class which maximizes  $p(w_i) L(\underline{x} | w_i)$  (equation 2). Assuming the prior probabilities equal, the decision rule reduces to (Figure 6),

$$\left\{ \begin{array}{l} \text{decide } w_1 \text{ if } x = \text{S.I.} > x_0 \\ \text{decide } w_2 \text{ if } x = \text{S.I.} < x_0 \end{array} \right.$$

when the utilities of misclassification are not equal, the decision rule reduces to (equation 4) (Figure 4),

$$\left\{ \begin{array}{l} \text{decide } w_1 \text{ if } E(\text{utility} | w_1) > E(\text{utility} | w_2), \text{ or,} \\ \quad [ p(w_1) L(x|w_1) u(1,1) + p(w_2) L(x|w_2) u(1,2) ] \\ \quad > [ p(w_1) L(x|w_1) u(2,1) + p(w_2) L(x|w_2) u(2,2) ] \\ \text{decide } w_2 \text{ otherwise} \end{array} \right.$$

Figure 7 shows the results of oil drilling experience in the Silverville sand, McKean County, Pennsylvania. The straightness-index of the Silverville sand between the points marked in Figure 7 is 1.12.

Figure 8 shows empirically determined cdf's and pdf's of straightness index for both bar and channel type deposits based on a limited number of observations (Baecher, 1971).

Let  $w_1$  and  $w_2$  be defined as before, and assume  $p(w_1) = p(w_2) = 0.5$ . From equation 1,

$$\begin{aligned} p(w_2 | \text{S.I.} = 1.12) &= \frac{p(w_2) p(\text{S.I.} = 1.12 | w_2)}{p(w_2) p(\text{S.I.} = 1.12 | w_2) + p(w_1) p(\text{S.I.} = 1.12 | w_1)} \\ &= 0.83 \end{aligned}$$

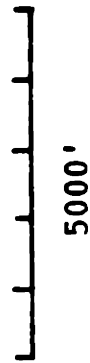
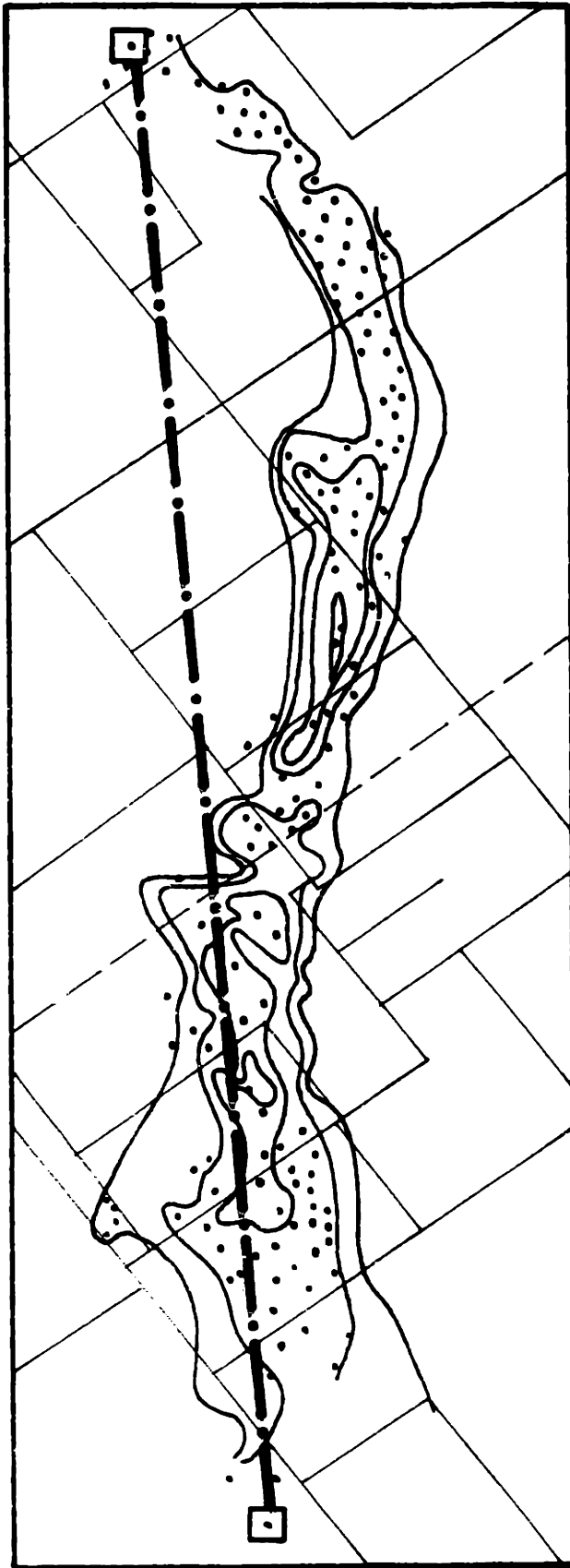


Figure 7: Isopach map of the Silverville sand, Music Mountain oil pool (after Levorsen, 1967)

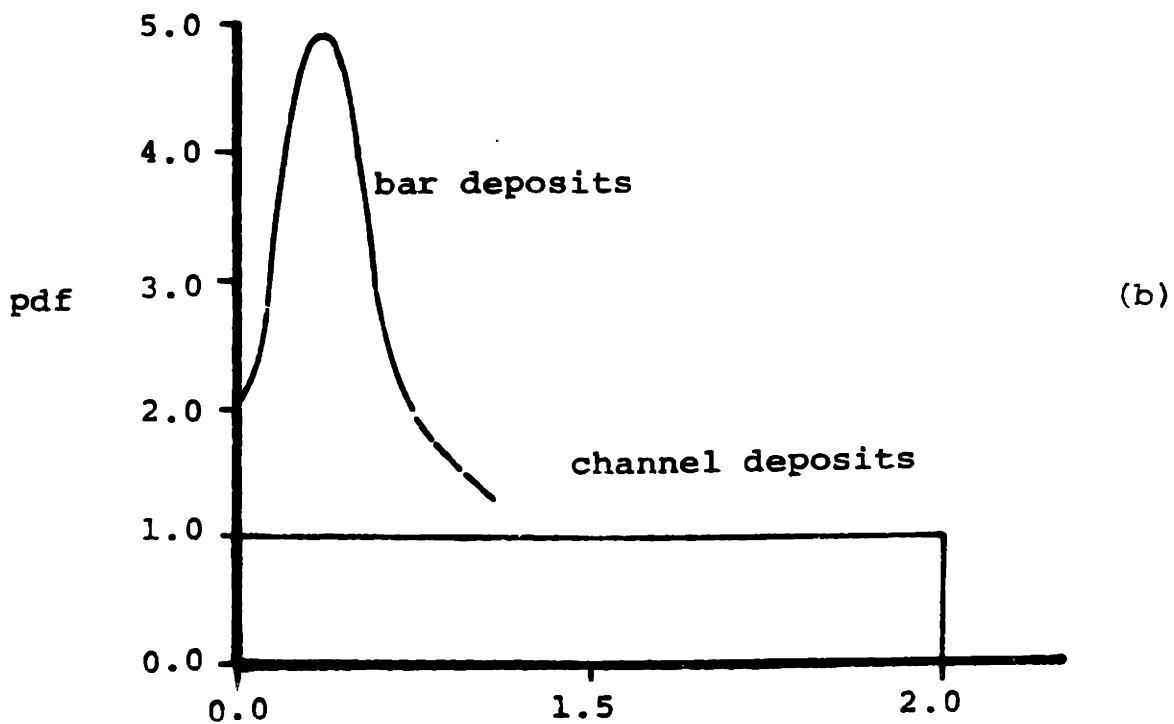
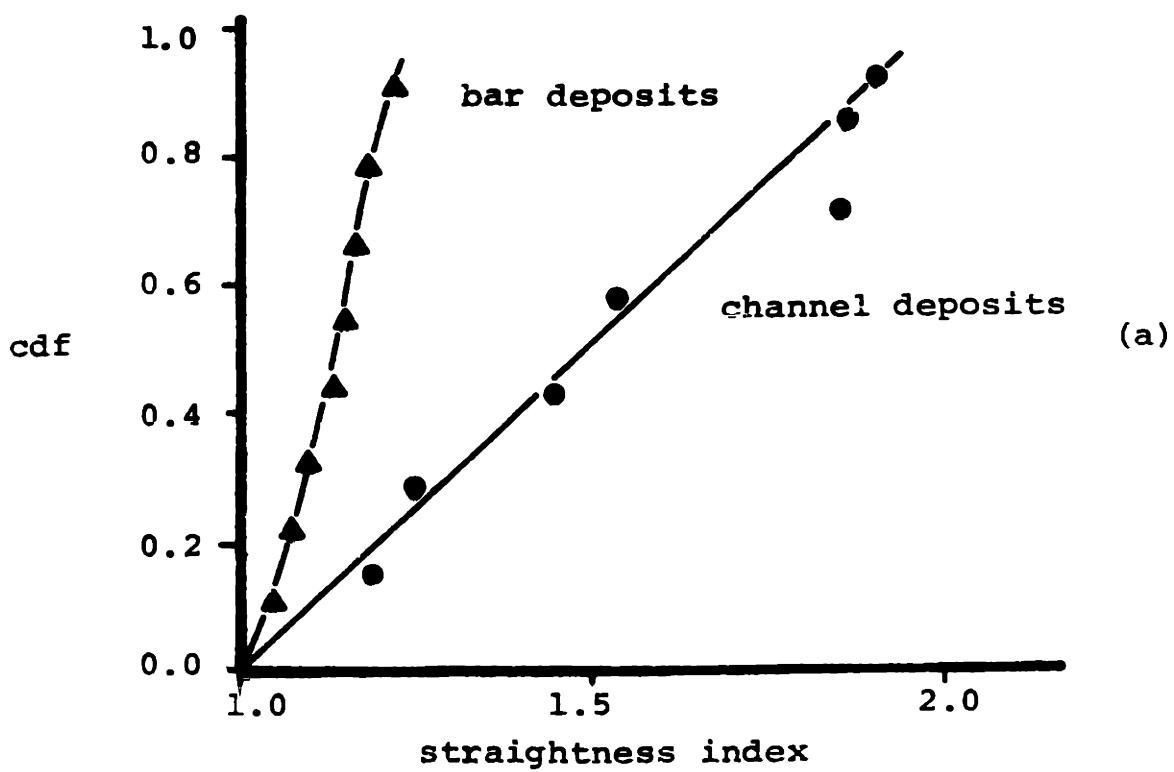


Figure 8: probability distributions for straightness index based on a limited number of empirical observations

$$p(w_1 | s.l. = 1.12) = 1 - 0.83 = 0.17$$

therefore, decide  $w_2$ , the body is bar type.

Were another shoestring deposit in the vicinity of the Silverville known to be bar-type, the prior probabilities might no longer be uniform. Assume  $p(w_1) = 0.25$  and  $p(w_2) = 0.75$ , then,

$$p(w_2 | s.l. = 1.12) = 0.94$$

$$p(w_1 | s.l. = 1.12) = 0.06$$

which is a considerably higher probability than with uniform priors, as should be expected.

### Feature Extraction

Posterior probabilities of classification (e.g.,  $p(w_i | \underline{x})$ ) depend on the factor or sets of factors measured: some sets lead to greater expected deviations of the likelihood ratio (i.e.,  $L(w_i | \underline{x})/L(w_j | \underline{x})$ ) from unity than do others. A set of factors which leads to large expected deviations is said to have high resolution; one which leads to small expected deviations is said to have low resolution. Other things equal, factors having high resolution are favored to factors having low resolution, because they lead to lower probabilities of misclassification, and fewer factor measurements when classification is performed sequentially.

The use of multiple factors increases resolution, but also increases cost. Sometimes, instead of using the best subgroup of

factors from a large number of possible factors, "dimensionality reduction" is used -- which simply refers to forming linear combinations of the original factors, thereby reducing their number to a convenient size (Fu, 1970; Casey, 1965, as reported by Duda, 1970). The linear discriminant function, used for classifying in classical statistics (Fisher, 1936) is a dimensionality reduction.

The way factors are selected in practice (and even if statistical methods are not used, classification decisions must still be based on factor evaluations) is mostly intuitive or empirical -- selecting factors by comparing their resolution on past data (e.g., see Krumbein and Sloss, Chapter 12; Nagy, 1968). Several measures of "goodness" for evaluating factors have been proposed (e.g., entropy (Lewis, 1962), variance ratios (Miller, 1962)); however, none of the measures can be shown to produce "optimal" factors except in special situations. For now, the best approach to factor selection for site exploration is a heuristic one, in which the opinions of experienced field personnel are solicited.

### Partitioning the Factor Space

When the number of items to be classified becomes large, evaluating equation 1 for each of them becomes infeasible. If, however, a map of the factor space (the space over which the vector  $\underline{x}$  is defined) could be partitioned into regions, such that at every point  $\underline{x}_0$  in a region the decision  $w_i$  was the same, classification could be accomplished simply by noting where a vector  $\underline{x}$  plots in the



partitioned space.

Consider classifying strata on the basis of self-potential and resistivity logging in a sandstone formation with interbedded shales (e.g., Figure 2). The factor for a particular stratum would be

$$\underline{x} = (\text{self-potential, resistivity})^t$$

Given sample data from logs for which a physical core was recovered (i.e., "calibrating" data), we first need to evaluate the conditional probability densities of the factors for the two lithologies. Much of the difficulty in determining the factor's conditional probability densities rests with the assumptions made at the outset. We can simplify the task considerably by assuming the factors to be statistically independent. The validity of this assumption is questionable, of course, but it reduces the two-dimensional problem to two one-dimensional problems. If we also assume that the density functions are members of particular families of distribution (e.g., normal, beta, etc.), the problem reduces to estimating the parameters of the distribution. Again, this is a questionable assumption, but it greatly simplifies our computations; once empirical evidence is built up, better assumptions can be made. We'll assume that the marginal distributions of the factors

$$x_1 = \text{self-potential}$$

$$x_2 = \text{resistivity}$$

are independently normally distributed with means  $m_1$ ,  $m_2$ , and variances  $\sigma_1^2$ ,  $\sigma_2^2$ , respectively. That is,

$$p(\underline{x} | \text{shale}) = (1/2\pi\sigma_1\sigma_2) \exp\left\{-\frac{1}{2}\left[\frac{(x_1 - m_{1sh})^2}{\sigma_1^2} + \frac{(x_2 - m_{2sh})^2}{\sigma_2^2}\right]\right\} \quad 5$$

$$p(\underline{x} | \text{sandstone}) = (1/2\pi\sigma_1\sigma_2) \exp\left\{-\frac{1}{2}\left[\frac{(x_1 - m_{1ss})^2}{\sigma_1^2} + \frac{(x_2 - m_{2ss})^2}{\sigma_2^2}\right]\right\} \quad 6$$

Equations 5 and 6 can be rewritten in matrix notation by defining the covariance matrix

$$\underline{V} = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$$

or,

$$p(\underline{x} | sh) = \frac{1}{2\pi|\underline{V}|^{1/2}} \exp\left\{-\frac{1}{2}(\underline{x} - \underline{m}_{sh})^t \underline{V}^{-1} (\underline{x} - \underline{m}_{sh})\right\} \quad 7$$

and,

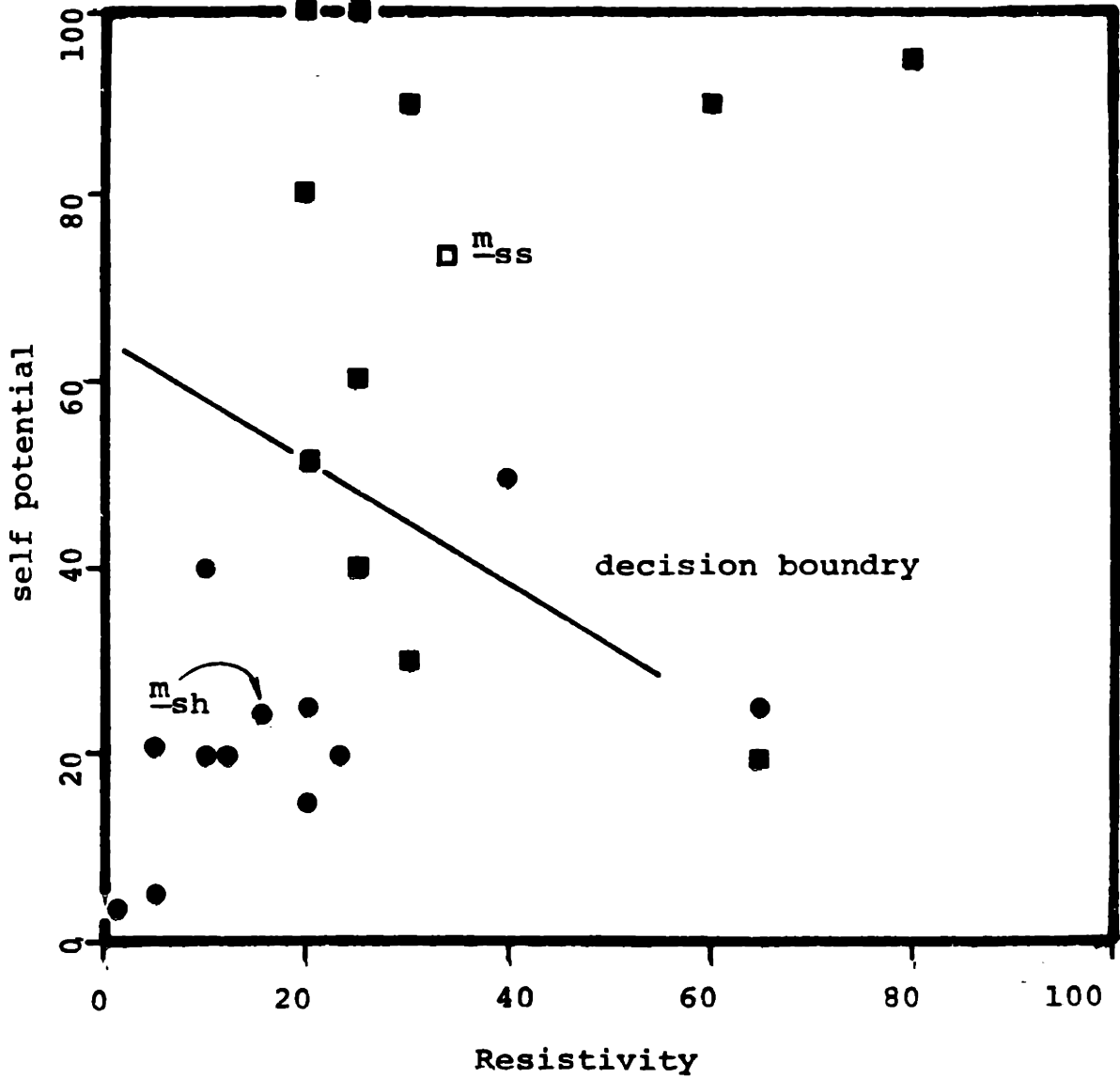
$$p(\underline{x} | ss) = \frac{1}{2\pi|\underline{V}|^{1/2}} \exp\left\{-\frac{1}{2}(\underline{x} - \underline{m}_{ss})^t \underline{V}^{-1} (\underline{x} - \underline{m}_{ss})\right\} \quad 8$$

where  $\underline{m} = (m_1, m_2)$ .

If we assume the prior probabilities of shale and sandstone to be equal, and the utilities of misclassification to be constant, a decision rule based on equation 2 reduces to selecting the class which maximizes  $p(\underline{x} | w_i)$ :

$$\left\{ \begin{array}{l} \text{decide "shale" if} \\ \quad (\underline{x} - \underline{m}_{sh})^t \underline{V}^{-1} (\underline{x} - \underline{m}_{sh}) < (\underline{x} - \underline{m}_{ss})^t \underline{V}^{-1} (\underline{x} - \underline{m}_{ss}) \\ \text{decide "sandstone" otherwise} \end{array} \right.$$

Shown in Figure 9 is a plot of empirical data taken from "calibrating" borings for a region under exploration (from LeRoy, 1951). From this data the estimated marginal density parameters are



$$\underline{m}_{sh} = (15.3, 24.4)$$

$$\underline{m}_{ss} = (33.3, 73.7)$$

$$\underline{Y} = \begin{pmatrix} 243 & 0 \\ 0 & 433 \end{pmatrix}$$

Figure 9

$$\underline{m}_{sh} = (15.3, 24.4)$$

$$\underline{m}_{ss} = (33.3, 73.7)$$

$$V = \begin{bmatrix} 243 & 0 \\ 0 & 433 \end{bmatrix}$$

Combining these estimates with the decision rule above gives the linear partition shown in Figure 9. This is a line passing through  $1/2(\underline{m}_{sh} + \underline{m}_{ss})$  and normal to  $V^{-1}(\underline{m}_{sh} - \underline{m}_{ss})$ . Strata whose factor vectors plot above and to the right of the decision boundary are classified as sandstone; those below and to the left are classified as shale.

This decision rule is based on questionable assumptions about the nature of the conditional probability densities of the factors; however, better boundaries can be obtained only when better (i.e., more realistic) assumptions can be made. Better assumptions can be made by evaluating empirical data in light of the statistical approach.

### Sequential Methods

The use of sequential methods in classification problems has received considerable attention, particularly from Fu (1968a, b; 1970). We shall consider two approaches here: direct application of Wald's sequential probability ratio test (SPRT), and the Bayesian decision approach. Dynamic programming techniques will not be introduced, although several applications have been proposed in the literature

(Fu, 1970; Fu, Chien, and Cardillo, 1967; Cardillo and Fu, 1967; Chien and Fu, 1967).

### SPRT

Consider a two-class problem in which the possible classes are denoted  $w_1$  and  $w_2$ . If, as before, there are  $m$  factors which can be measured, a sequential procedure is one which evaluates the current evidence for classification after each factor is measured, and decides whether to terminate the process by classifying the pattern or to continue by measuring more factors.

One vehicle for making this decision is Wald's sequential probability ratio test (Wald, 1947). After each stage,  $k$ , the probability ratio

$$\lambda_k = \frac{P_k(\underline{x} | w_1)}{P_k(\underline{x} | w_2)} \quad 9$$

is computed, which is the ratio of the likelihood of observing  $\underline{x}$  given that the pattern belongs to  $w_1$ , to the likelihood of observing  $\underline{x}$  given that the pattern belongs to  $w_2$ . When these likelihoods are the same,  $\lambda_k = 1.0$ . Factor measuring is continued as long as  $\lambda_k$  lies within specified bounds  $A$  and  $B$ , and is terminated if  $\lambda_k$  exceeds these bounds. The decision rule is

$$\left\{ \begin{array}{ll} \text{continue measuring if} & B < \lambda_k < A \\ \text{decide } w_1 \text{ if} & \lambda_k \geq A \\ \text{decide } w_2 \text{ if} & B \geq \lambda_k \end{array} \right.$$

The bounds A and B depend on the tolerable magnitude of the probabilities of misclassification. The larger the allowable probabilities, the closer A and B are to 1.0, and the fewer factors that need measuring. If  $e_{ij}$  is the tolerable probability of deciding  $w_i$  when actually  $w_j$  is correct, then (Wald, 1947)

$$A = \frac{1 - e_{21}}{e_{12}} \quad 10a$$

$$B = \frac{e_{21}}{1 - e_{12}} \quad 10b$$

Use of these bounds assures that the probabilities of misclassification will be at least as small as  $e_{12}$  and  $e_{21}$ . Furthermore, it has been shown (Fu, 1969) that the SPRT is an optimal strategy for the two class problem, in that, for a given  $e_{12}$  and  $e_{21}$  no strategy exists with error probabilities at least as low as  $e_{12}$  and  $e_{21}$  and with a lower expected number of factor measurements before classification. The SPRT differs from a Bayesian approach in that it specifies tolerable errors instead of maximizing expected utility, and as such the cost of considering additional factors does not enter the analysis (see Chapter 1).

For greater-than-2-class problems the generalized sequential probability ratio test (GSPRT) can be used (Fu, 1969). The generalized probability ratio at any stage  $k$  for any class  $w_i$  is

$$u_k(x|w_i) = \frac{p_k(x|w_i)}{\sqrt[n]{\prod_{j=1}^n p_k(x|w_j)}}$$

which is compared with the stopping boundary for the  $i^{\text{th}}$  class

$$A(w_i) = \frac{1 - \alpha_{ii}}{\sqrt[n]{\prod_{j \neq i} (1 - \alpha_{ij})}} \quad 12$$

If  $u_k(\underline{x} | w_i) < A(w_i)$ , the hypothesis that the pattern belongs to class  $w_i$  is rejected, and by progressive elimination the final classification is made. When the number of classes equals 2 the GSPRT reduces to the SPRT. Optimality of the GSPRT has never been proved (FU, 1969).

A significant difficulty with the SPRT and GSPRT is that the required number of factor measurements may be quite large, particularly if the  $e_{ij}$ 's are chosen to be very small. To avert this problem modifications of the GSPRT have been suggested. The simplest proposes to follow the GSPRT until a maximum of  $N$  factor measurements have been made. If no decision has been reached prior to the  $N^{\text{th}}$  measurement, the most probable class is chosen according to equation 1. This truncated procedure utilizes the advantages of evaluating the current data after each measurement, while recognizing the realistic requirement that only so many factors are available for measurement (either for cost or physical reasons).

Fu (1968a) has suggested using varying stopping boundaries which grow closer to 1.0 as additional factors are measured, offering some control over the number of measurements as well as over the allowable error probabilities. Elaboration is given in Fu (1968a, 1970) and Anderson (1960).

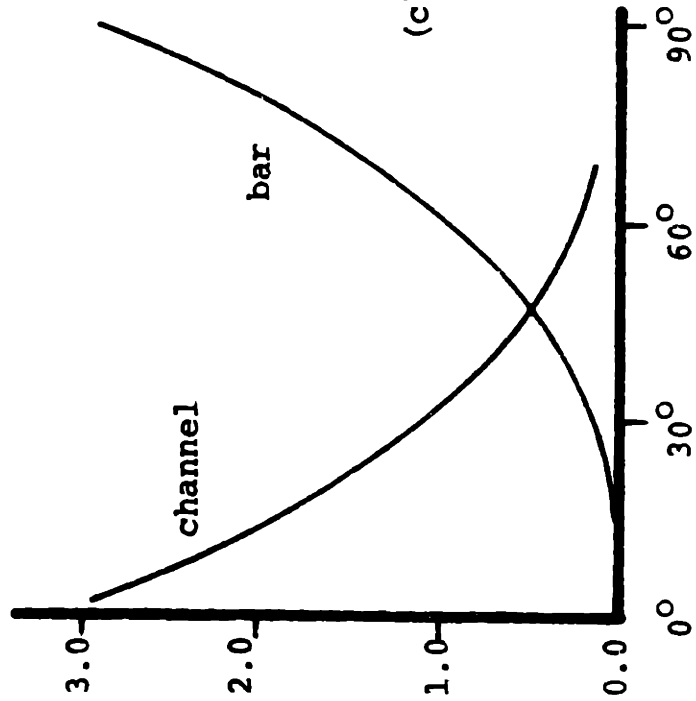
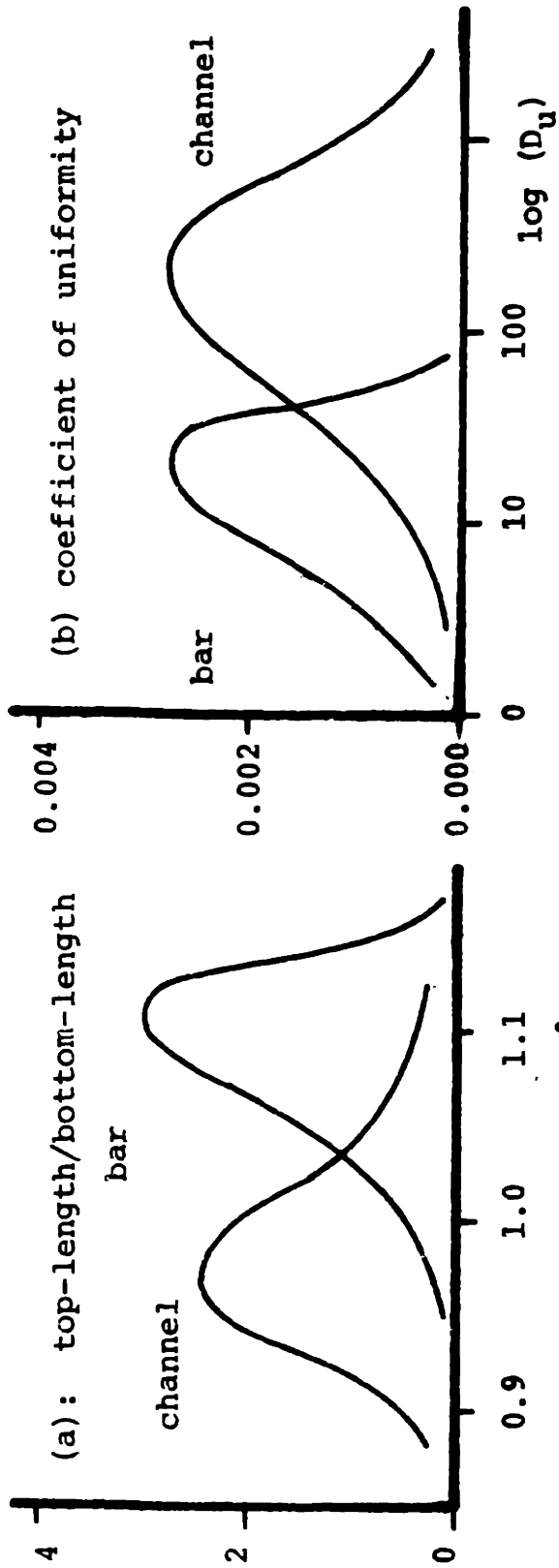


FIGURE 10: CONDITIONAL PDF'S OF FACTORS FOR SHOESTRING SANDS



This time, however, we will consider a sequential procedure (using Wald's SPRT) in which after each factor is measured and evaluated, a decision is made to either classify upon current information or measure additional factors.

We will consider four factors, requiring increasing amounts of effort to measure,

1. Coefficient of uniformity of the grain size distribution:  
 $D_u = D_{60}/D_{10}$
2. Average angle of cross bedding (in the horizontal plane) with respect to the axis of the body
3. Straightness-index
4. Ratio of cross-sectional top surface arc-length to bottom surface arc-length

Hypothetical pdf's of the last three factors are shown in Figures 10a, b, c.

If error probabilities of 5% have been chosen,  $e_{12} = e_{21} = 0.05$ :

$$A = \frac{1 - e_{21}}{e_{12}} = 19$$

$$B = \frac{e_{21}}{1 - e_{12}} = 0.053$$

The sequential procedure then results in:

Stage 1: Factor,  $D_u = 30$

$$\lambda_k = \frac{(0.0014)}{(0.0025)} = 0.56$$

Decision: Collect information on an additional factor

Stage 2: Factor, Crossbedding angle =  $50^{\circ}$

$$\lambda_k = 0.56 \left( \frac{0.3}{0.5} \right) = 0.34$$

Decision: Collect information on an additional factor

Stage 3: Factor, Straightness-index = 1.12

$$\lambda_k = 0.34 \left( \frac{1.0}{4.8} \right) = 0.07$$

Decision: Collect information on an additional factor

Stage 4: Factor, Top arc-length/Bottom arc-length ratio

$$\lambda_k = 0.07 \left( \frac{0.8}{1.5} \right) = 1.05$$

Decision: Classify as  $w_2$  (bar deposit)

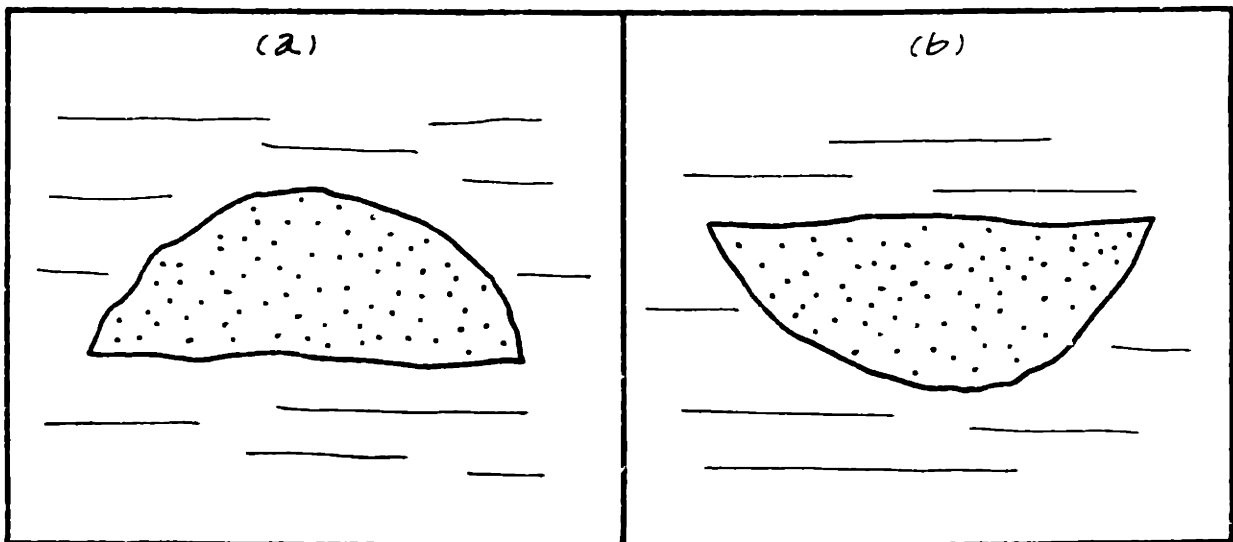
### Bayesian Approach

Instead of treating sequential classification by means of a SPRT we may use a Bayesian decision approach, which allows us to weigh the cost of factor measurement against the utility of misclassification. To simplify our calculations and discussion we will use a factor which can assume only two discrete values rather than the continuously distributed factors of the last example. Our decision is this: We are considering expending additional effort to investigate the cross-sectional shape of a shoestring. This gives us information on the depositional environment, but a cost is associated

with the additional exploration.

Say, the body may be shaped either as in Figure 11a, or 11b, and that if the body was formed by off-shore deposition there is an 80% chance that the cross-section will be as "a," while if the body was formed by stream deposition there is a 70% chance that it will be as "b". If the prior probability is 60% that the body was formed by offshore deposition, should we determine the cross-sectional shape or classify upon information already in hand?

Fig. 11



A decision tree for this decision is shown in Figure 12. The expected probability of correct classification associated with continued exploration is

$$\begin{aligned}
 E \left[ \text{pr}(\text{correct classification}) \mid \text{continue} \right] \\
 &= 0.63 (0.81) + 0.37 (0.76) \\
 &= 0.79
 \end{aligned}$$

The expected probability of a correct classification associated with not determining the shape is 0.6. If the utility of increasing the probability of correctly classifying the body from 0.6 to 0.79 is greater than the exploration cost  $C$ , then the new factor should be measured; otherwise, a decision should be made based on the current information.

Example:

A shoestring deposit is known to be water bearing and is to be drilled for water supply. Were the depositional mode of the body known, its position in the vicinity of interest could be guessed with a success ratio of 60%; however, were the incorrect depositional mode assumed, a success ratio of only 20% would be realized.

Drilling to the depth of the sand costs about \$3,000, but the net present worth of a producing well is estimated to be \$20,000. Should \$2,000 be spent to determine the cross-sectional shape, or should the body be classified upon existing information?

The expected value of the decision to evaluate cross-sectional shape is (Figure 12)  $\$8882 - 2000 = \$6882$ ; the expected value of the decision to classify upon current information is \$7120. Therefore, a well should be drilled without exploring cross-sectional shape.\*

---

\*For simplicity we have considered only one drill/do-not-drill decision. Actually, because success ratio depends on depositional mode, a drilling which does not intersect the sand is information which can be used to update probabilities of each mode. Therefore, in a real application one must consider the possibility of second or further drilling attempts and the changing probabilities of correct classification.

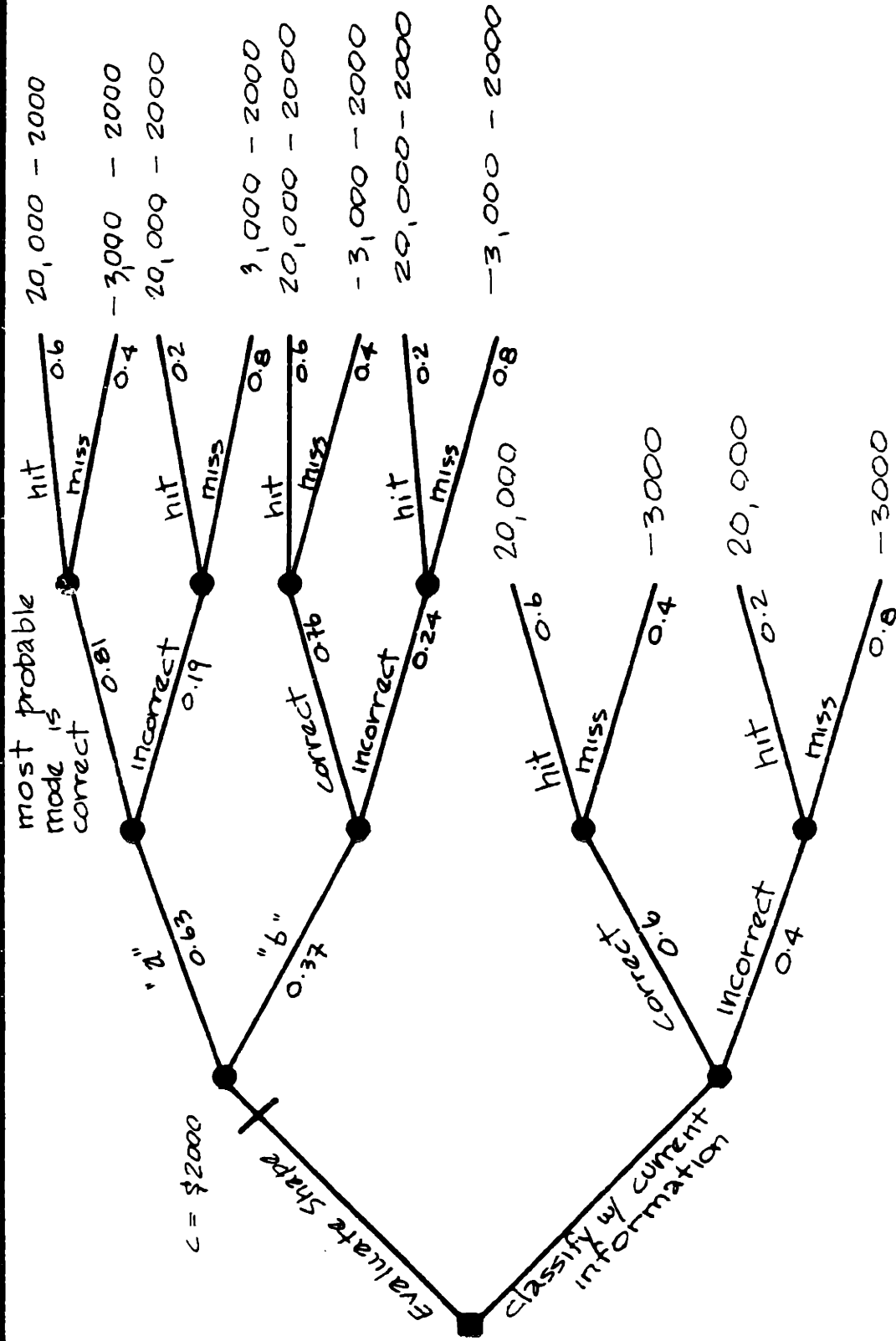


Figure 12: Decision Tree for Evaluating Body Cross-Sectional Shape <sup>65</sup>

We have implicitly assumed in treating the decision on each factor in isolation that the conditional likelihoods of all the factors are independent, that is

$$\begin{aligned} L(x_1, x_2 | w_i) &= L(x_1 | x_2, w_i) L(x_2 | w_i) \\ &= L(x_1 | w_i) L(x_2 | w_i) \end{aligned} \quad 13$$

which is not always the case. When the factors are dependent the sequential decision must be made enumerating all the possible combinations of outcomes and computing the expected utility of each branch of the decision tree (i.e., preposterior analysis).

Suppose the cross-sectional shape and the orientation of cross-bedding were not independent. For example, a stream deposited body whose profile was shaped as shown in Figure 11a might be more likely to have cross bedding parallel to its axis than a stream deposited body whose profile was shaped as shown in Figure 11b. For the sake of example assume that the conditional joint pmf's of profile shape and cross bedding orientation are as shown in Figure 13a. We'll take the utility of a branch proportional to the expected probability of correct classification were that branch followed. From Figure 13b the expected utility of evaluating factor  $e_1$  is  $(0.85k - C_1 - C_2)$  or  $(0.85k - C_1)$ , whichever is larger, where  $k$  is the proportionality constant relating utility to correct classification and  $C_i$  is the cost of evaluating  $e_i$ . The expected utility of classifying without evaluating  $e_1$  is  $k(0.6)$ . If the expected utility of measuring  $e_1$  is greater than the expected utility of

Probability mass functions for profile shape and cross-<sup>167</sup>  
bedding orientation

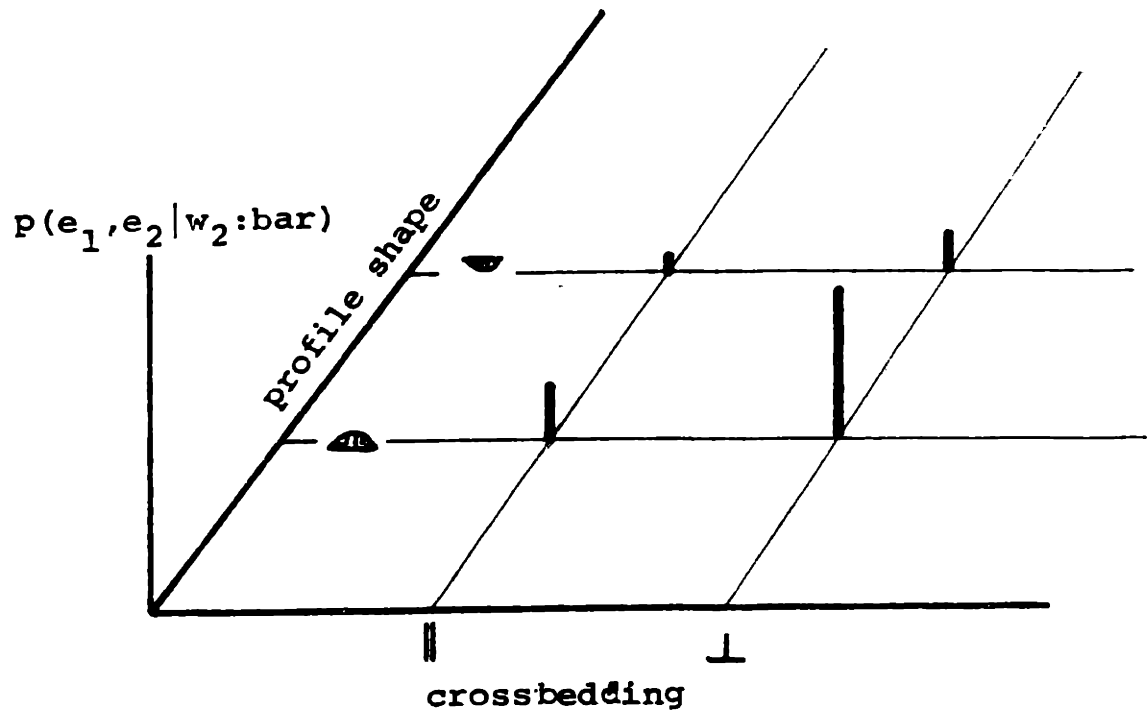
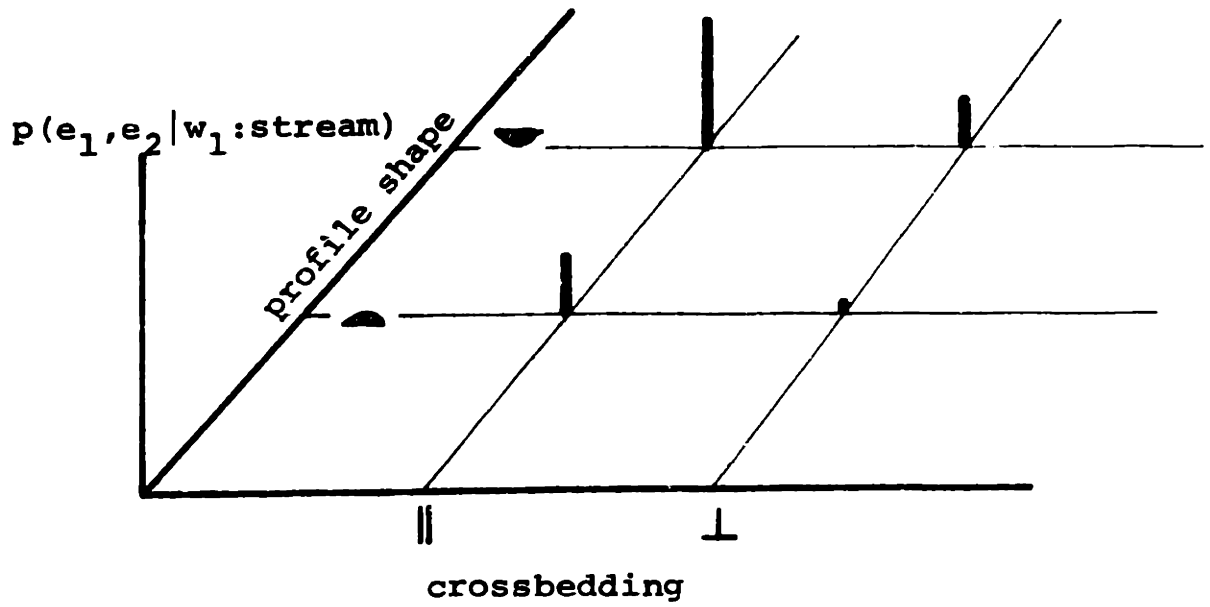


Figure 13a

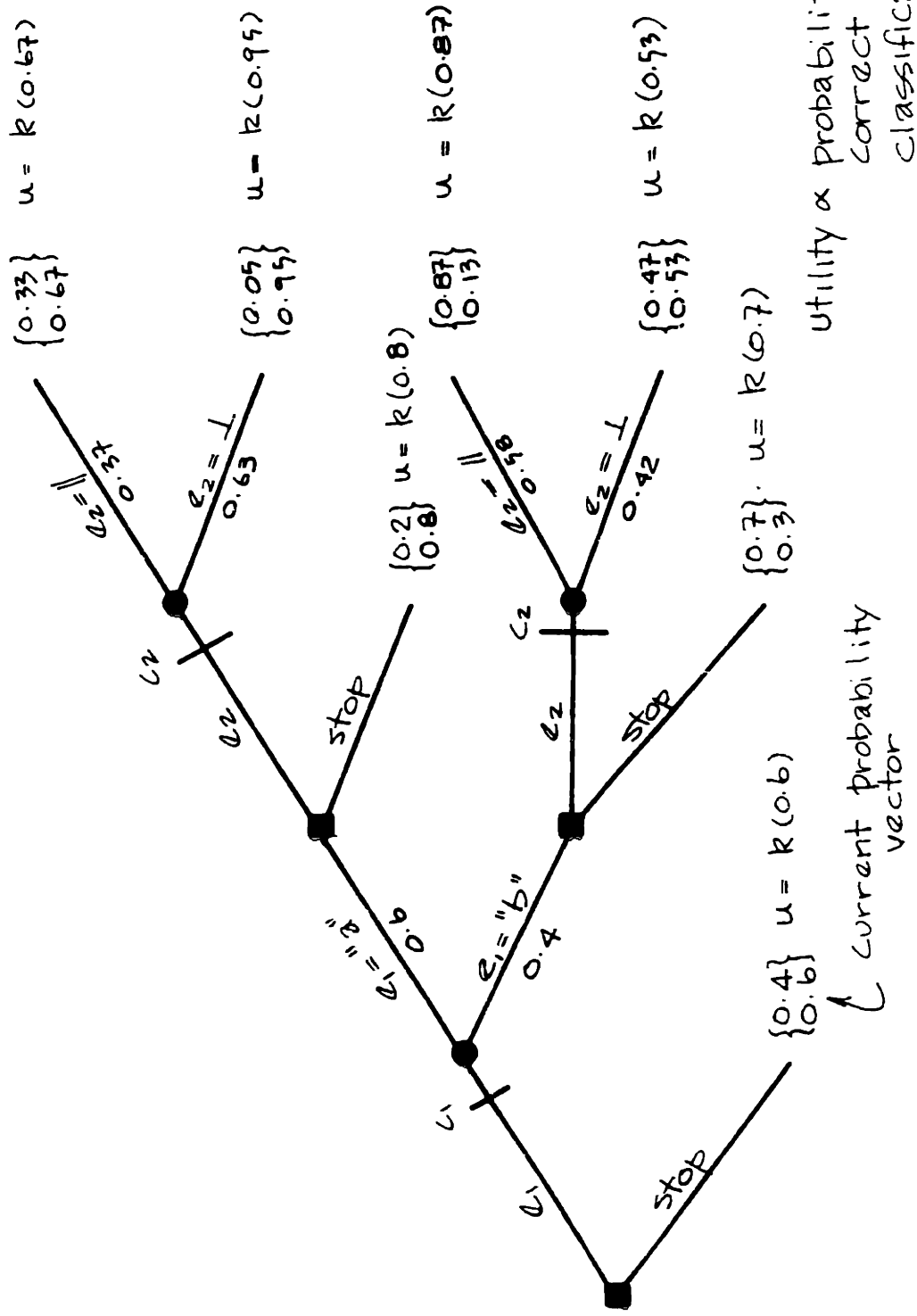


Figure 13b: Decision tree for sequential factor measurement



stopping,  $e_1$  should be evaluated. If  $e_1$  is evaluated, again a decision is to be made whether or not to measure  $e_2$ ,

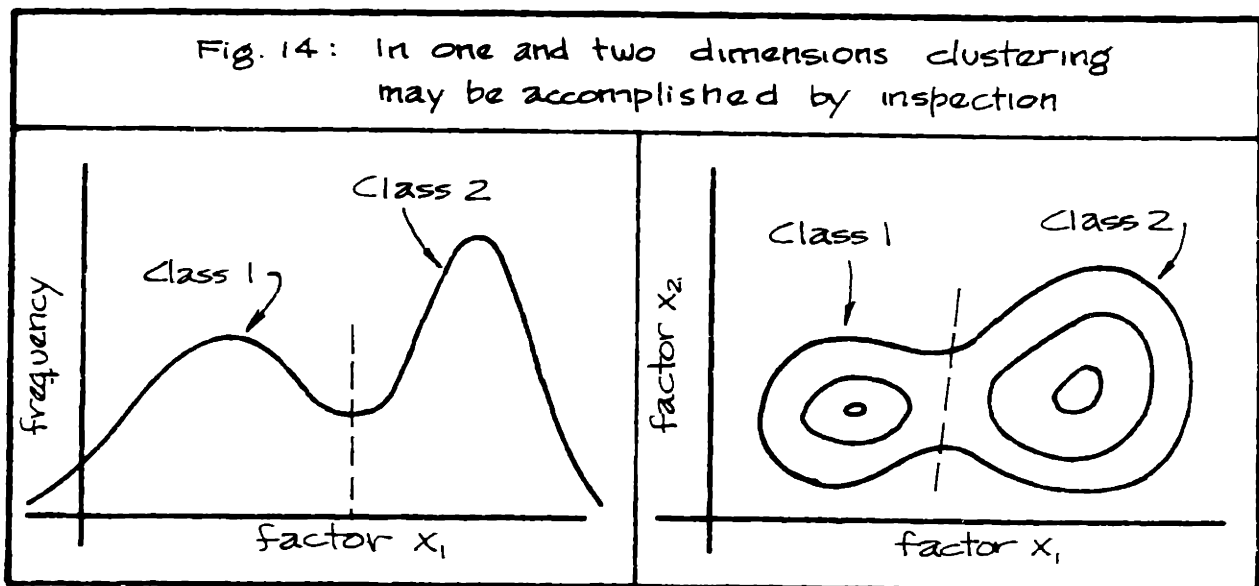
$$\text{if } e_1=a, \text{ measure } e_2 \text{ if } (0.85k - C_2) > 0.8k$$

$$\text{if } e_2=b, \text{ measure } e_2 \text{ if } (0.72k - C_2) > 0.7k$$

Preposterior analysis may be used for any number of sequential decisions, although the number of branches increases rapidly with increases in the number of decisions.

### Clustering

When encountering a new problem the number of classes into which data should be grouped and their locations in the factor space may be unknown. In one or two dimensions the number of classes can be adequately determined by inspecting a plot of factor space vs. frequency of observation (Figure 14). However, when the number of factors becomes larger than two, visual inspection is of little use.



Were we to project the data upon lines or planes, visual inspection could then be used to isolate modes or clusters in the empirical data. The problem is to find directions on which clustering in the projected data is apparent (e.g., line A-A in Figure 15 is a line with high resolution for the particular empirical data, while line B-B is not).

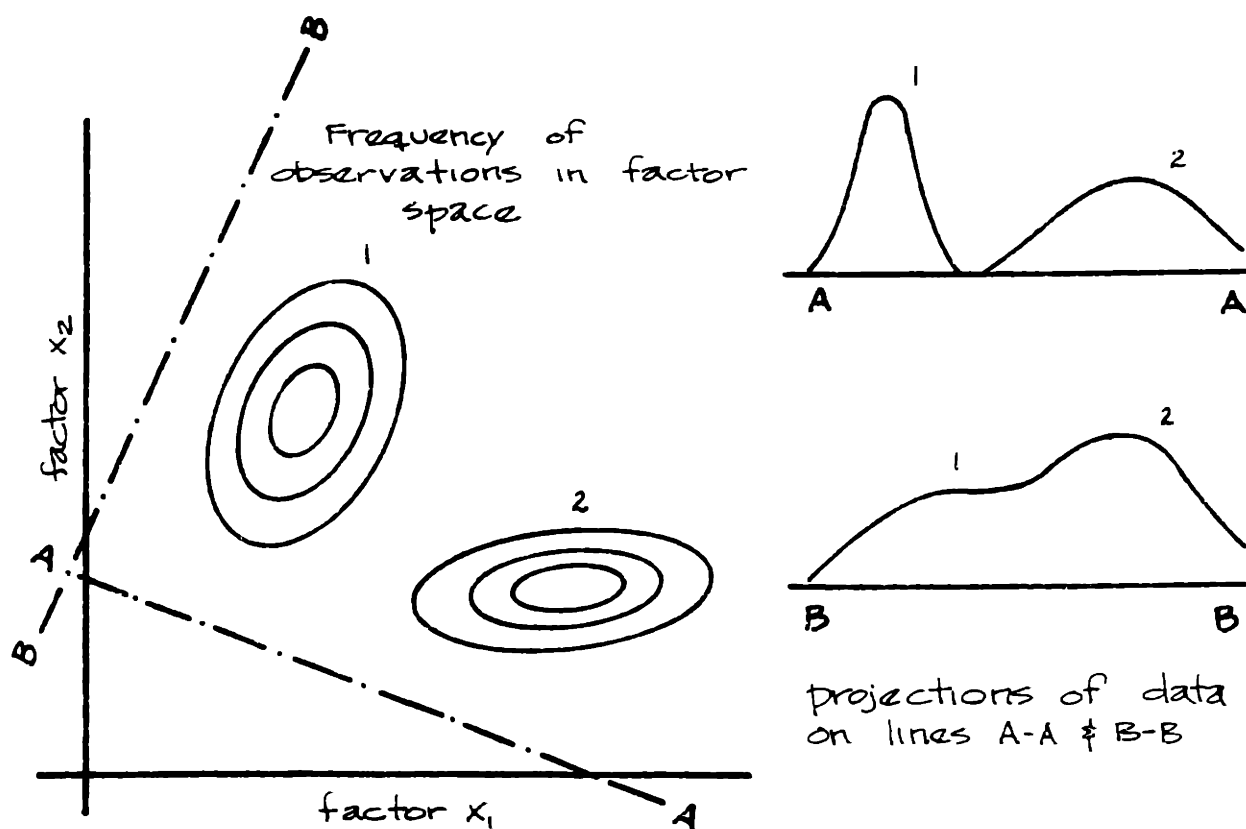


Fig. 15

Mattson and Dammann (1965) present a method for seeking modes or clusters in multidimensional data by a projection technique. We wish to find a line along which the projected density has the maximum

variance. Since variance is analogous to moment of inertia (i.e., they are both second moments of distributions), we might think of this as being the line along which the projected data has the greatest moment of inertia.

Let there be  $k$  samples in a data set, and let the  $i^{\text{th}}$  factor value of the  $j^{\text{th}}$  sample be  $x_{ij}$ . The value (or location) of the  $j^{\text{th}}$  sample when projected onto a line  $S$  equals the scalar product of  $\underline{x}$  and  $\underline{w}$ , where  $\underline{w}$  is a vector along line  $S$ ,

$$s_j = \underline{x} \cdot \underline{w} = \sum_{i=1}^m x_{ij} w_i \quad 14$$

and  $m$  is the dimension of the factor space (i.e., the number of factors measured on each sample). The average location  $\bar{s}$  for the entire data set is

$$\bar{s} = \frac{1}{k} \sum_{j=1}^k s_j \quad 15a$$

$$= \sum_{i=1}^m \bar{x}_i w_i \quad 15b$$

where  $\bar{x}_i$  is the average value of the  $i^{\text{th}}$  component of the observed factor vectors. The variance along line  $S$  is proportional to

$$\text{Var}(s_j) \propto \sum_{j=1}^k (s_j - \bar{s})^2 \quad 16$$

which can be made as large as desired by allowing the  $w_i$  to be large.

If, however, we seek to maximize 16 subject to the constraint that  $\sum w_i^2$  is small (i.e., maximize  $\xi$  where

$$\xi = \frac{\sum_{j=1}^k (s_j - \bar{s})^2}{\sum_{i=1}^m w_i^2} \quad 17$$

Mattson and Dammann have shown that the problem reduces to finding the eigenvector associated with the largest eigenvalue of the matrix

$$\{a_{ij}\} = \left\{ \sum_{r=1}^k (m_{ir} - \bar{x}_i)(x_{jr} - \bar{x}_j) \right\} \quad 18$$

which is proportional to the data covariance matrix.

The procedure for clustering data with no prior class information then is the following:

1. Compute the largest eigenvalue of the data set covariance matrix, and determine the eigenvector corresponding to this eigenvalue.
2. Project the multivariate data set onto the eigenvector to obtain a one dimensional density distribution.
3. Partition the data set at any nodes in the projected density which approach zero probability density (e.g., Figure 14a). If no such nodes exist, repeat the process using the eigenvector corresponding to the second largest eigenvalue. If all of the eigenvectors have been considered with no partitioning, conclude that only one cluster exists.
4. To further partition the data set, isolate the data in each partitioned subregion and repeat the procedure for each.

Consider the 9-cluster two dimensional example shown in Figure 16 (from Mattson and Dammann, 1965). Following the procedure just outlined, lines of maximum projected variance are shown in Figure 17a, and the projected densities in Figure 18a. The data set is partitioned, as shown, at the minimum node. The partitioned data sets are now treated in isolation and the process repeated, resulting in the second partitioning (Figures 17b and 18b). The process is

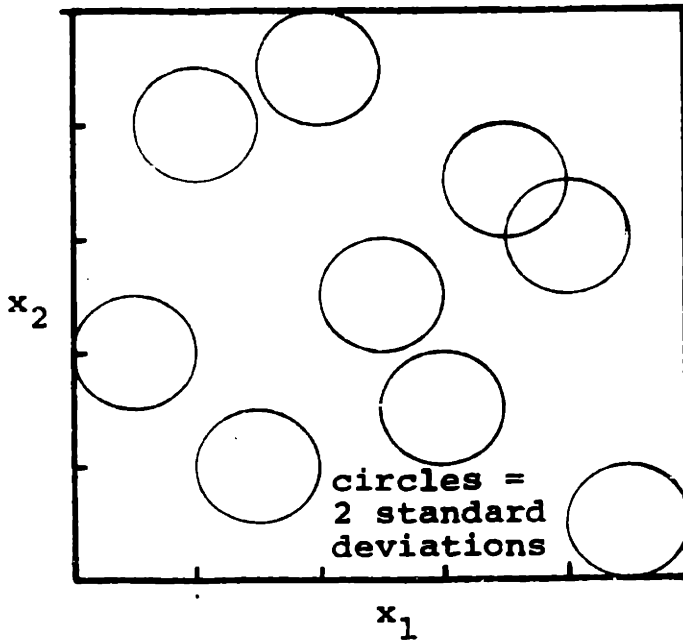
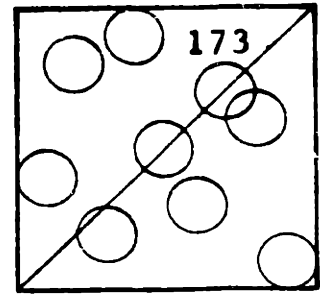
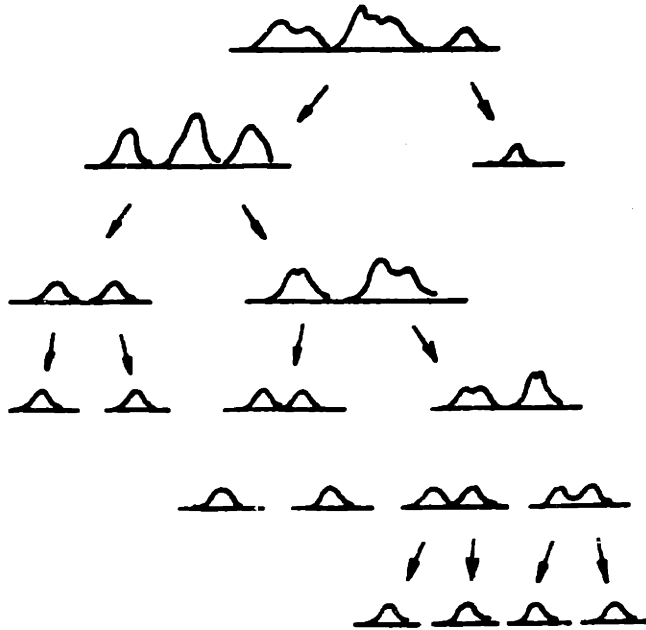


Figure 16

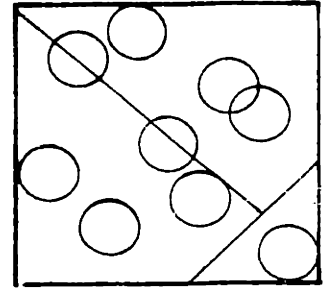
9-cluster, 2-dimensional data set (after Mattson and Dammann, 1965)

Figure 17: projected lines of maximum variance

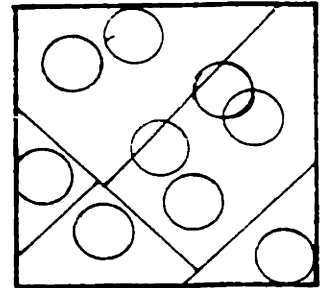
Figure 18: projected data



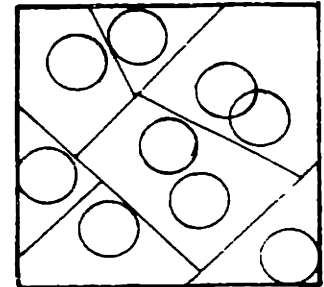
a



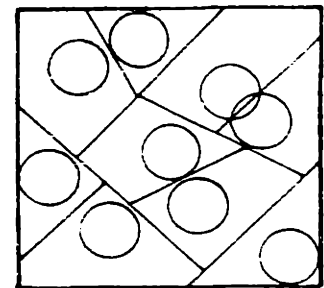
b



c



d  
e



continued (through Figures 17e and 18f) until all 9 clusters are separated.

### Section 5.2.2 Probabilistic Mapping and Rules for Inference

The construction of geological maps seems at first to be almost outside the realm of quantitative analysis, since it is presently a highly subjective "art." However, if we are ever to make rational decisions based on the probable geological nature of sites and on allocating effort to reconstruct geometric patterns of site geology, we will require an analytical approach. In this section we introduce the concept of a probabilistic map which associates with each site location an ordered  $n$ -tuple describing the probabilities of possible geological classes. To the author's knowledge this approach to mapping (i.e., entirely in terms of an  $n$ -tuple function function of location) is novel and one must emphasize that an approach to quantitative mapping is being presented here, not a completed theory. Considerable additional inquiry must be made before these methods will be of practical use.

Certain theoretical and computational difficulties exist and will become apparent as the discussion proceeds.

#### Quantitative Probabilistic Mapping

A probabilistic mapping scheme based on uncertainties in knowledge of site geology should satisfy four requirements:

1. It should be simple to conceive and use.
2. One should be able to assign subjective probabilities to its components.
3. It should allow the construction of probabilistic rules for inferring the composition of unobserved locations.
4. Uncertainties in mapping should be readily interfaceable with design decisions.

The scheme we introduce, to associate an ordered n-tuple of probabilities with each point in space, does meet these requirements. The concept is simple; subjective pmf's may be used as priors; inference rules can be constructed (as discussed later in this section); and uncertainties may be directly interfaced with design decisions. Design decisions are based on the probabilities of composition in regions influenced by proposed construction. These probabilities may be read directly from a map of the type proposed, which includes the probabilities of unlikely composition as well as likely ones. Other schemes which might have been proposed, such as assigning each point to the most probable class and associating with it a probability of misclassification, generally do not have this facility.

Consider a site composed of only two geological materials -- a so-called two class region. We can assign to each point at the site an order pair  $(p_i) = (p_1, p_2)$  whose first term is the probability that the point belongs to class 1 and whose second term is the probability that the point belongs to class 2. For observed points this pair would be either (1,0) or (0,1) (assuming for now that true classification can be ascertained by observation). For points other

than those observed the pair  $(p_i)$  would be somewhere between  $(1,0)$  and  $(0,1)$ , subject to the constraint that  $p_1 + p_2 = 1.0$ . Obviously, we may expand the dimension of this assignment to any required size, in which case we would assign an ordered  $n$ -tuple, and may include a term representing the probability that the location belongs to an unenumerated class.

Unless the function which assigns  $n$ -tuples to locations can be approximated by some continuous mathematical model, discretizing the site into elemental areas of homogeneous classification may be required.\* With the aid of electronic computers these elements may be made essentially as small as needed to adequately satisfy the assumption of homogeneous classification, and this approximation should be no worse here than in other discretized methods of engineering analysis.

### Rules for Inference

The way one infers the geologic nature of unexplored locations is not based in deductive logic. One does not begin from premises and observations and deduce necessary conclusions. Instead, these inferences are inductive. One observes previous sites, imagines simple models describing them, and assumes these models apply to the present site.

David Hume (1739) was the first to recognize the full import of

---

\* Throughout this chapter the term "n-tuple" will be taken to mean "ordered n-tuple."



the epistemological problems originating from inductive reasoning. In his philosophy of "skepticism" he in essence said that induction has no rational justification, and the problem of justifying induction, or "Hume's problem", has been the central focus of the philosophy of science ever since (e.g., see Salmon, 1966). One should not conclude that induction is illogical, rather it is "non-logical" (meaning outside of logic).

This places the way we infer geology at unobserved locations on a tenuous base. The inference is composed of two parts, model building and statistical reasoning, and only the latter is susceptible to analytical formulation. The former is simply and totally outside the realm of rational analysis and always will be. We put much more into model building than merely the bare prior observations; it is an imaginative process (see, Medawar, 1968; Brillouin, 1964).

In this section we will suggest models that attempt to describe the way geological formations behave, and apply statistical analysis to using these models. For the present they are simplistic, but because of the nature of model building they will be refined only after considerable empirical evaluation of their predictions is made.

The class existing at a given element is presently inferred on the basis of two considerations. The first of these is the nature of elements close by, and the second is the presence of trends in the local geology. Figure 19 presents three illustrations of the application of proximity and trend considerations. Suppose that the numerals 1 and 2 refer to different geological formations, and that

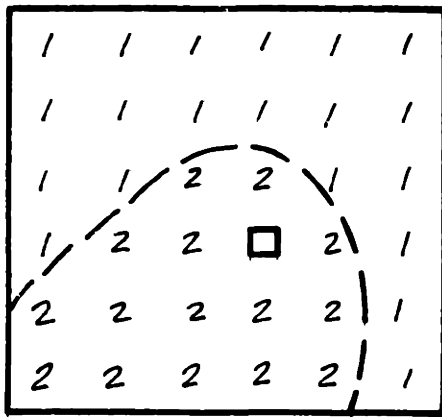


Fig. 19a

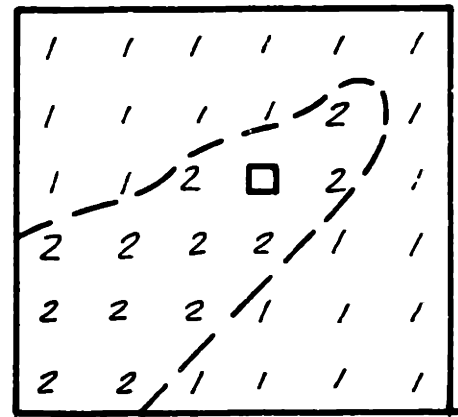


Fig. 19b

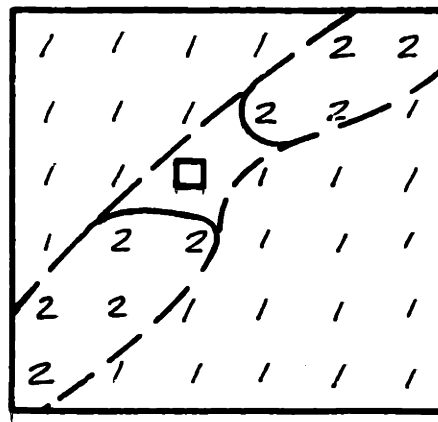


Fig. 19c

a void square locates an unexplored element. The unexplored element in 19a would likely be assigned to class 2, illustrating an application of the proximity rule. The unexplored element in 19b would likely be assigned to class 2 also -- although intuitively this assignment is not as certain as it was in case 1 -- illustrating an application of the trend rule. The assignment of the unexplored element in 19c is not at all clear. Proximity considerations offer evidence that 1 is the most probable class, while trend considerations offer evidence that 2 is the most probable. An intuitive preference for one of these rules,

which existed in cases a and b, does not exist in c. Possible partitionings are shown by the solid and dashed lines.

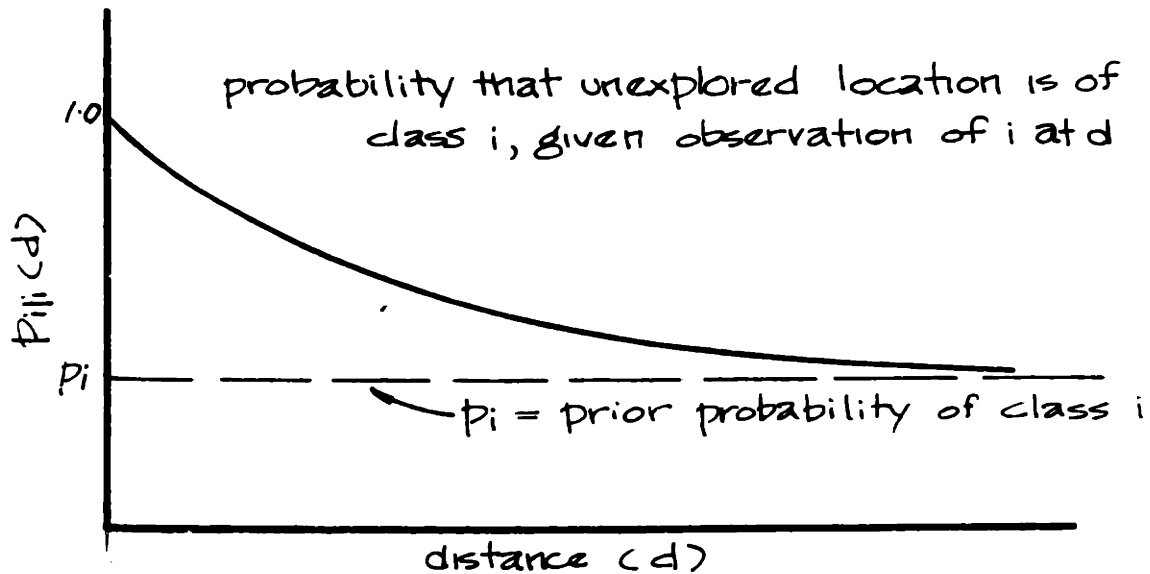
#### Proximity Rules:

Simple nearest neighbor rules appear frequently in civil engineering practice (e.g., assigning areas of influence to rainfall measurements), but usually specify only the most probable classification. The typical nearest-neighbor rule would be: "Assign to any unexplored location the class to which the nearest sampled point belongs." We want more detailed inferences than this: We want a rule that specifies the most probable class, and its probability.

Proximity rules are based on the assumption that locations close together are more likely to be of the same class than locations far apart. Switzer (1967) and Matérn (1960) have suggested the use of a function  $p_{i|i}(d)$  which describes autocorrelation (spatial correlation) between locations in terms of the probability that a location will be of class  $i$ , given that an observation at distance  $d$  is of class  $i$ . The value of this function must equal 1.0 at  $d = 0$ , and the prior probability of class  $i$  at  $d = \infty$ . Matérn has shown that  $p_{i|i}(d)$  must be convex in the vicinity of  $d = 0$ , and Switzer has suggested the form (Figure 20),

$$p_{i|i}(d) = p_i + (1 - p_i)e^{-ad}$$

Fig. 20



Switzer's model assumes an isotropic autocorrelation (i.e., the correlation between locations depends only on their separation distance and not on direction, but equation 19 can be expanded to account for anisotropy,

$$P_{i|i}(d) = p_i + (1 - p_i) e^{-\sqrt{a^2 d_x^2 + b^2 d_y^2}} \quad 20$$

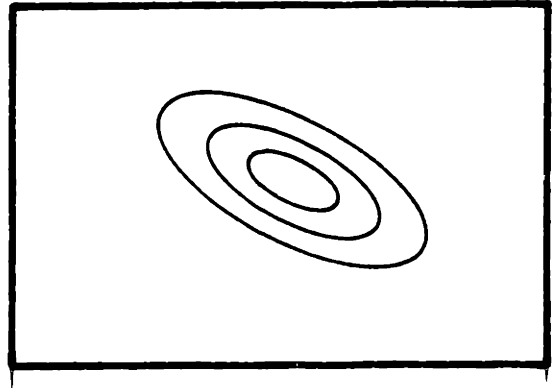
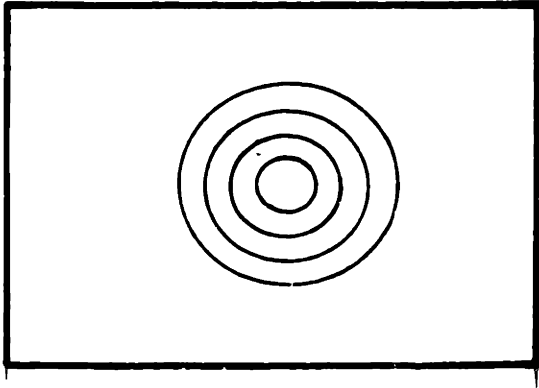
where  $d_x$  and  $d_y$  are the separation distances parallel and perpendicular to the regional preferred orientation (Figure 21).

Fig. 21

Contours of constant  $p_{i|i}(d)$ 

isotropic autocorrelation

anisotropic autocorrelation



By the relationship,

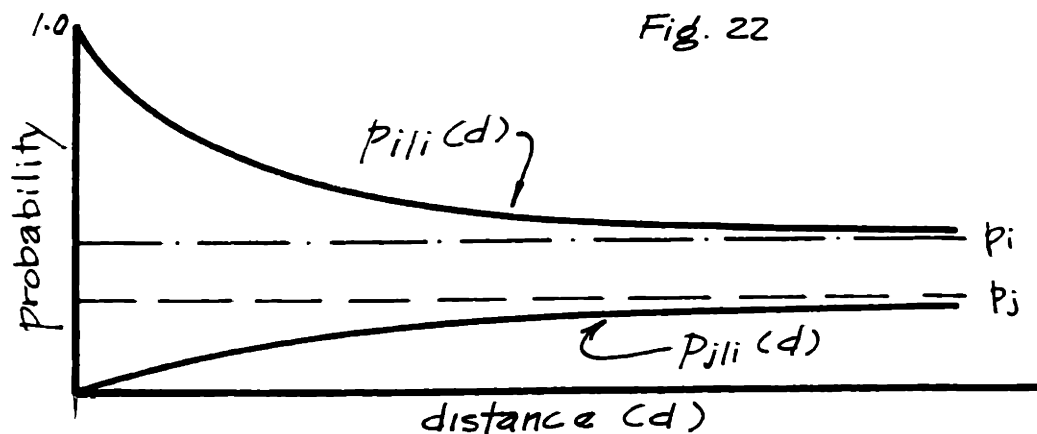
$$\sum_j p_{j|i}(d) = 1.0 \quad 21$$

the functions  $p_{j|i}(d)$ ,  $j \neq i$ , are constrained by,

$$\sum_{j \neq i} p_{j|i}(d) = 1 - p_{i|i}(d) \quad 22$$

or for a two-class region (i and j) (Figure 22),

$$p_{j|i}(d) = 1 - p_{i|i}(d) \quad 23$$



Although previous work has only considered the partitioning problem -- not the determination of probability vectors -- and has only considered the influence of the nearest neighbor, equation 1 can be extended to include the effect of all sample data by application of Bayes' Theorem. Given  $r$  sampled locations,  $z=1, \dots, r$ , each at a distance  $d(z)$  from an unexplored location, the posterior probability of class  $i$  at the unexplored location is

$$P_i' = \frac{P_i L(\text{data} | i)}{\sum_i P_i L(\text{data} | i)} \quad 24$$

$$= \frac{P_i \prod_z P_{\xi | i}(d)}{\sum_i P_i \prod_z P_{\xi | i}(d)} \quad 25$$

where  $\xi$  is the class at sample point  $z$ . Due to the shape of the functions  $P_{\xi | i}(d)$  (Figure 22) there is some critical distance  $d_c$  beyond which the influence of sampled points is negligible.

Let  $A$  be the set of sampled points within a circle of radius  $d_c$ .

Equation 25 can be rewritten as,

$$P_i' = \frac{P_i \prod_{z \in A} P_{\xi | i}(d) \prod_{z \notin A} P_{\xi | i}(d)}{\sum_i P_i \prod_{z \in A} P_{\xi | i}(d) \prod_{z \notin A} P_{\xi | i}(d)} \quad 26$$

$$= \frac{P_i \prod_{z \in A} P_{\xi | i}(d) \prod_{z \notin A} P_{\xi}}{\sum_i P_i \prod_{z \in A} P_{\xi | i}(d) \prod_{z \notin A} P_{\xi}} \quad 27$$

from which the constant term  $\prod_{z \notin A} P_{\xi}$  cancels. So, we need only

consider points within distance  $d_c$ . In the case of numerous sample points this elimination greatly reduces required computation.

An example of proximity rules is shown in Figure 23. From the observations shown determine the probability vector at the cross-hatched element on the basis of the proximity rule.

The proximity rule, for our use, must be considered an empirical model whose parameters are empirically determined for the site in question. Several authors have discussed the possibility of developing random process models which would predict the shape of equation 19, but no realistic model has yet to result.

A random process model is one which describes the generating mechanism for a particular physical phenomena in probabilistic terms, from which is described the theoretically "correct" stochastic behavior of the phenomenon. This is in contrast to an empirical model which simply fits a convenient, smooth analytical function to observed data with no theoretical basis for choosing the particular function. Lack of a random process model needn't concern us as long as empirical data is available from which to construct a function; however, in a given situation we may find need to exchange Switzer's exponential form for some other.

We have assumed in our discussion that the autocorrelation function is stationary. This means that the function  $p_i | i(d)$  is the same for every location at the site (i.e., it depends only on separation distance and direction, not location), which in certain geological contexts may prove to be a poor assumption. One

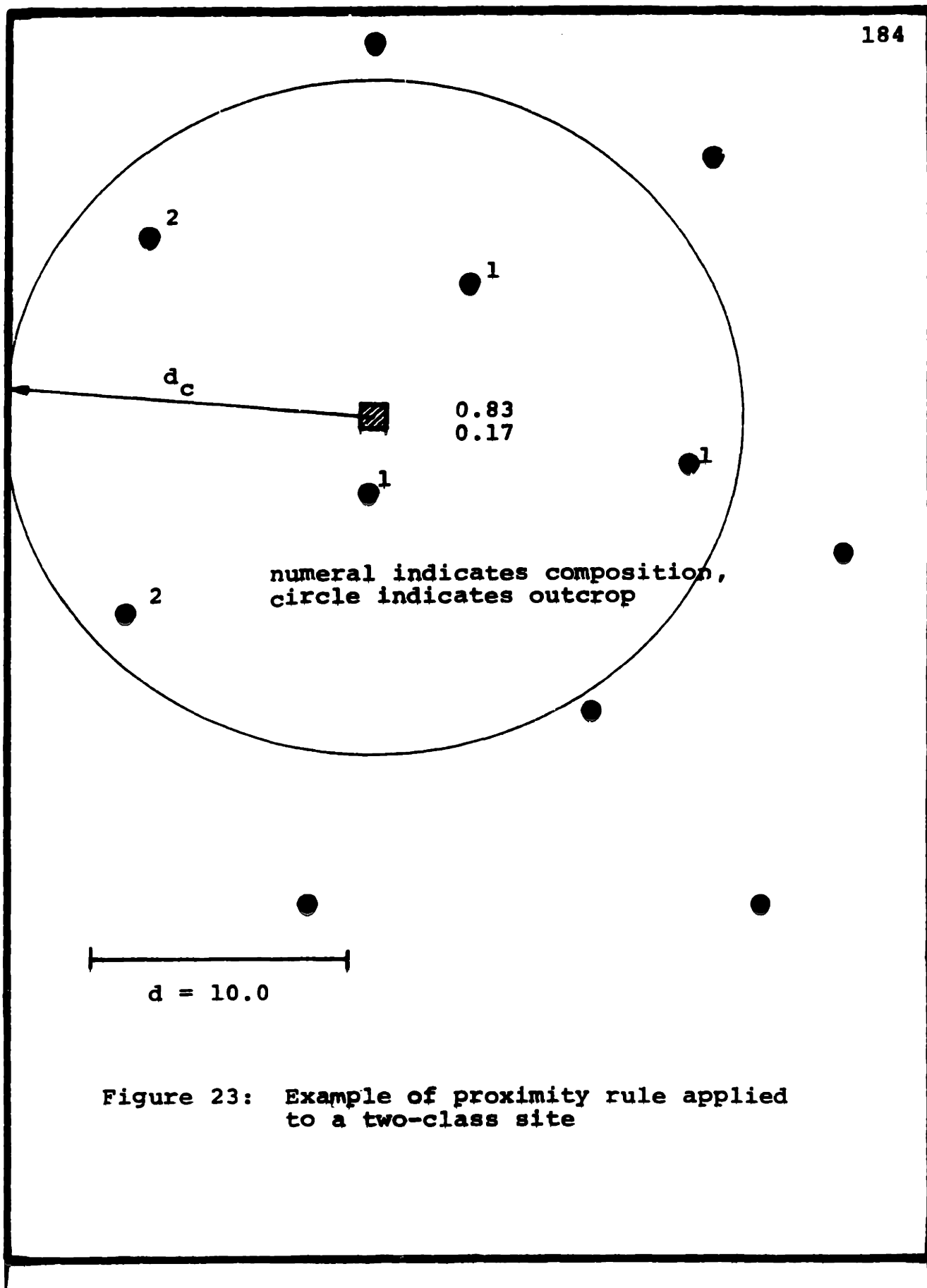


Figure 23: Example of proximity rule applied to a two-class site



Computations associated with Fig. 23

Determine the probabilities of composition for the cross-hatched element of Fig. 23 using proximity considerations.

Proximity Rule Relations:

$$\{P_{i|j}\} = \begin{bmatrix} P_{1|1} & P_{1|2} \\ P_{2|1} & P_{2|2} \end{bmatrix} \quad \begin{array}{l} P_{1|1} + P_{2|1} = 1.0 \\ P_{1|2} + P_{2|2} = 1.0 \end{array}$$

$$\begin{aligned} P_{1|1} &= 1 - (1-p_1) e^{-ad} \\ P_{2|1} &= 1 - [1 - (1-p_1) e^{-ad}] \\ P_{2|2} &= 1 - (1-p_2) e^{-bd} \\ P_{1|2} &= 1 - [1 - (1-p_2) e^{-bd}] \\ \text{suppose: } a &= b = 0.2 \end{aligned}$$

Data:

$z$	$d(z)$	class	$P_{2 1}(d_z)$	$P_{2 2}(d_z)$
1	3	1	0.77	0.23
2	7	1	0.63	0.37
3	12.5	1	0.54	0.46
4	12	2	0.46	0.54
5	13	2	0.46 <sup>5/7</sup>	0.53 <sup>5/7</sup>

Results:

Let,  $L_{i|j} = L(\text{composition } i|j @ d)$

$$\begin{aligned} P_1' &= \frac{P_1 \prod_z P_{2|1}(d_z)}{\sum_i P_i \prod_z P_{2|i}(d_z)} \\ &= \frac{0.5 L_{1|1}}{0.5 L_{1|1} + 0.5 L_{2|1}} \end{aligned}$$

$$\begin{aligned} &= \boxed{0.83} \\ \therefore P_2' &= \boxed{0.17} \end{aligned}$$

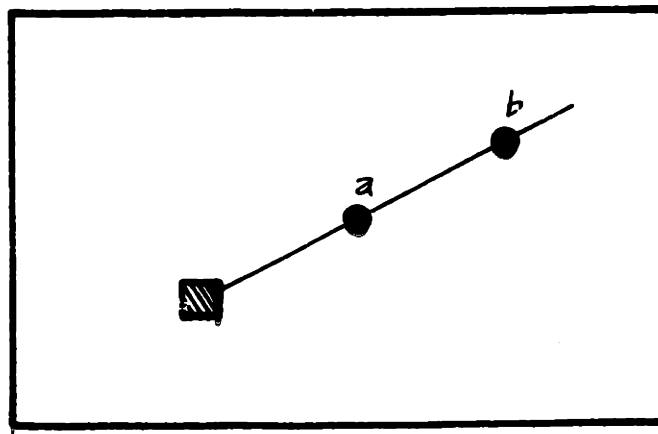
intuitively suspects the concentration of certain geologic bodies may be greater in some regions of a site than in others (i.e., bodies may be "clumped"). For example, altered bodies may occur in clumps due to the structural nature of the region. Non-stationarity would cause parameters of the function  $p_i | i(d)$  to depend on location, and would essentially prohibit the use of proximal techniques, since, unless a realistic random process model existed, no way would exist to empirically determine the trend of the parameters with site location prior to mapping.

Another assumption implicit in our proximity rule is that the observed points are independent. As a first approximation, independence is an admissible assumption, but one can conceive of situations in which it is not. If observations a and b (Figure 24) are made, the joint likelihood given the classification of the unknown location is not merely the product of the individual likelihoods (i.e.,  $L(a,b | c) \neq L(a | c)L(b | c)$ ). The observation of a affects the likelihood of b.

In the present work we will overlook this problem, but future inquiry will need to address it (perhaps by only considering the nearest observation within each of several angular sectors emanating from the unknown location).

Fig. 24

example of non-independent proximal data



Estimating the functional relationships in an actual case may prove difficult, and will always be limited in precision by normal sampling fluctuations. Empirical estimation requires that a near-by region or a region of similar geology has been completely mapped. A method of estimation would be the following: A grid of points is overlaid upon the mapped region, and several concentric circles drawn about each point.\* If the center point is of class  $i$ , the proportion of each circle intersecting material of class  $i$  is an estimate of  $p_i | i(d)$ , where  $d$  is the radius of the circle. Taking many such estimates -- one at each value of  $d$  for each grid point in class  $i$  -- the functional exponents of  $p_i | i(d)$  can be estimated. By

---

\*If regional periodicity is suspected these points may be located randomly.

classical sampling theory (see appendix to Chapter 6) the variance in the estimate of  $p_i | i(d)$  at a given value of  $d$  is proportional to the reciprocal of the number of measurements.

When the method of observation is not a perfect classifier (i.e., when one component of the probability vector of an observed point is not elevated to 1.0) the inference task is made considerably more difficult. The likelihood function (equation 20) must then be the weighted sum of the likelihoods of the possible classes

$L(\text{observed data point} | i @ \text{unobserved location}) =$

$$\sum_j p_j p_{j|i}(d) \quad 28$$

instead of merely  $p_i | i(d)$ , increasing the required computation several fold.

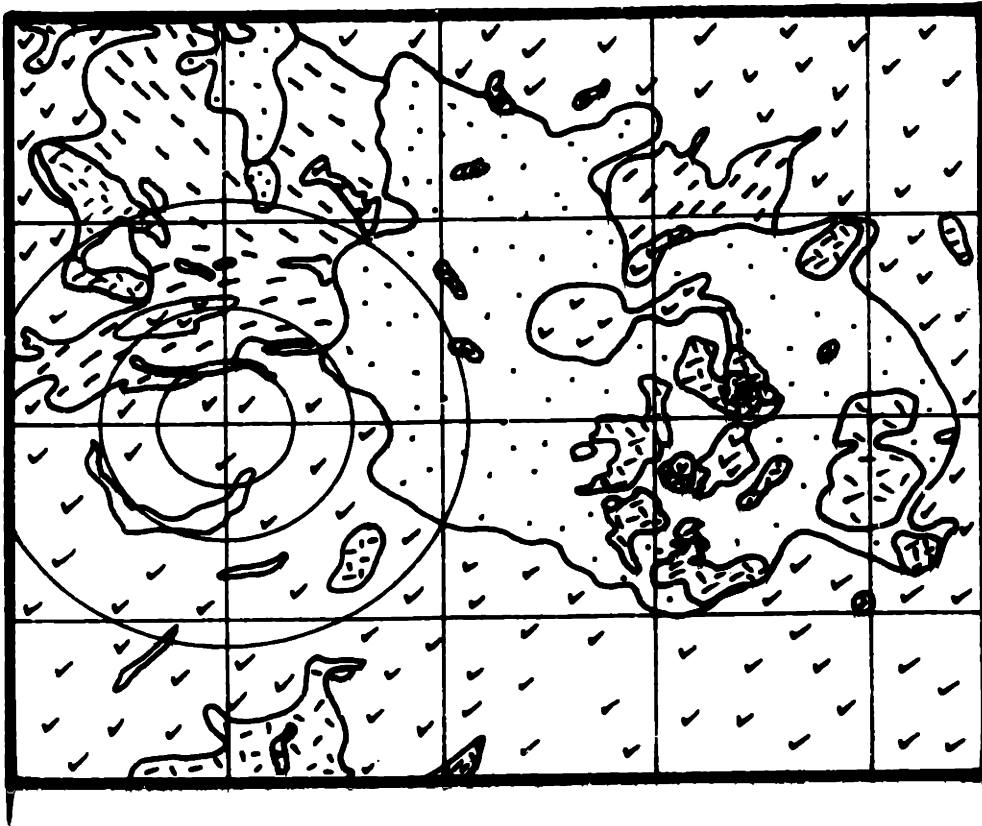
We will leave this refinement of the proximity rule to future work, but it does raise intriguing problems. Subjective priors could be easily evaluated if we treated each elemental area independently; however, if we allow that partial information about one element affects our uncertainty about nearby elements (which is what we are saying if we accept that proximity rules can be used when the observation method is not a perfect classifier), then our assessments of every element must be in perfect compatibility with every other element (the certainty which we assign to the classification of one element affects the certainties of adjoining elements and vice versa) -- a major task of balancing.

**Example: Empirical Determination of the Function  $p_i | i(d)$**

1. Shown in the accompanying Figure is a geol. map of a volcanic area near Cripple Creek, Colo. (After Lovering and Goddard, 1950.)
2. On this map a grid of points has been randomly located.
3. About each point three concentric circles ( $d = 1,000'$ ,  $2000'$ , and  $4000'$ ) have been drawn.
4. The local geology is divided into two classes:
  - class i: "hard" igneous  
(Basalts, granites, Rhyolites,  
Gneiss, etc.)
  - class j: "weak" igneous and clastic rocks\*  
(tuff, sandstone, cong, etc.)
5. The angular ratios of the intersections of each circle within class i to class j are measured, and the fns.  $p_i | i(d)$  and  $p_j | i(d)$  estimated.

---

\*This materials are not necessarily of low strength; this is just a convenient grouping for the sake of illustration.



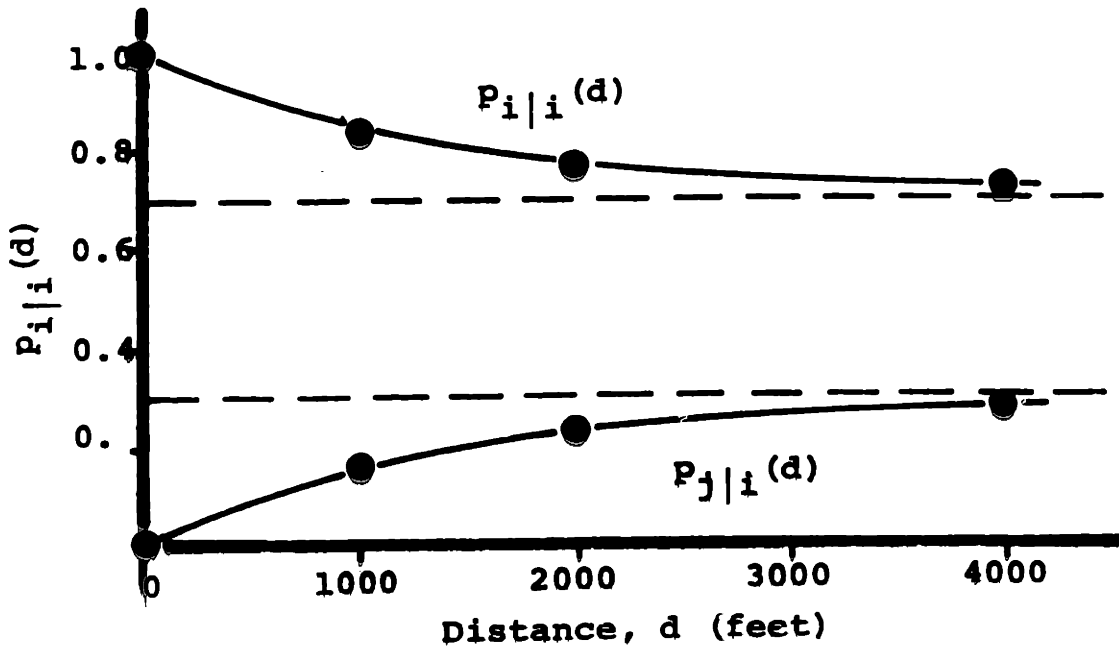
gneiss  
& schist

granite

breccia,  
tuff, ss

phonolite

(after Lovering & Goddard, 1950)



Tabulation of sample data:

Ratio of class  $j$ , given  $j$  at center

	<u>1000'</u>	<u>2000'</u>	<u>3000'</u>
	1.00	1.00	0.91
	1.00	0.85	0.70
	0.67	0.67	0.61
	0.78	0.87	0.81
	0.66	0.45	0.58
	0.95	0.91	0.89
	0.82	0.82	0.78
	0.53	0.50	0.50
	0.97	0.72	0.88
	<hr/>	<hr/>	<hr/>
$\Sigma$ :	738	679	647
Ave. :	0.82	0.77	0.72

Inference based on Bayesian theory, however, has the property that discrepancies between prior distributions converge quickly as the number of observations increases, particularly when the priors are close to being uniformly distributed. As a first approximation priors may be evaluated independently for each element, unless locations of high certainty exist, which may be treated as perfectly certain and the proximity rule of equation 28 applied.

In later inferences, again as a first approximation, observations might be divided into two groups. Those for which the maximum component of the probability vector is greater than some arbitrary value (say, 0.9) might be classified and subsequently used in inference as perfectly certain, while those below might not be used in inferring adjacent probability vectors at all.

### Trend Rules

Evidence of the composition of an unexplored location is contained in the existence of trends in local geology as well as in proximal observations, and trend extrapolation is an integral part of the way maps are usually constructed. So, in addition to proximity rules we require quantitative techniques for the extrapolation of trends and inference of probability vectors at distal locations.

The initial decision that trending should be considered is a difficult one to quantify. Perhaps, in the future, descriptions of the "strength" of trends will be developed as an aid, but these decisions must be made inductively. In this work we will treat the



limited problem of inferring compositions once trending has been accepted.

The procedure we will use to extrapolate trends is to fit mathematical curves to observations, determine the probability distributions of the coefficients of those curves, and extrapolate them to unexplored areas.

The selection of appropriate mathematical curves for specific geologic bodies, as in all empirical curve fitting, can only be made through one's professional knowledge and experience (e.g., faulted structures can often be modeled by straight lines), although in certain cases the theoretical mechanism of genesis may aid the selection (e.g., cone sheets, folding; see Ramsay, 1967; or Busk, 1929). In most curve-fitting problems polynomial functions are selected, unless there exist reasons to believe that other functions would be better fits, since polynomials have simplicity and flexibility.\*

The question of how far a trend is continuous is a classic one in exploration, and indeed in philosophy of science generally (see Teller, 1968). For those reasons discussed, we can never hope to develop satisfactory "objective" rules, although we shall discuss two approaches for particular situations.

---

\* One may take exception to the seemingly heuristic manner which we suggest for choosing trend shapes, but our contention throughout is that site exploration is very much a subjective undertaking, and probabilistic models are valid only in so far as they are based on and make use of professional judgement and experience.

When one reviews geological maps he sees that trend problems may be roughly grouped into three classes. There are "tongue shaped" bodies with large length to width ratios and constant width;\* there are contacts between formations; and there are, for lack of a better name, "wedge" shaped bodies whose width diminishes with length (Figure 25). We shall consider each in succession.

#### Trending Bodies of Approximately Constant Width:

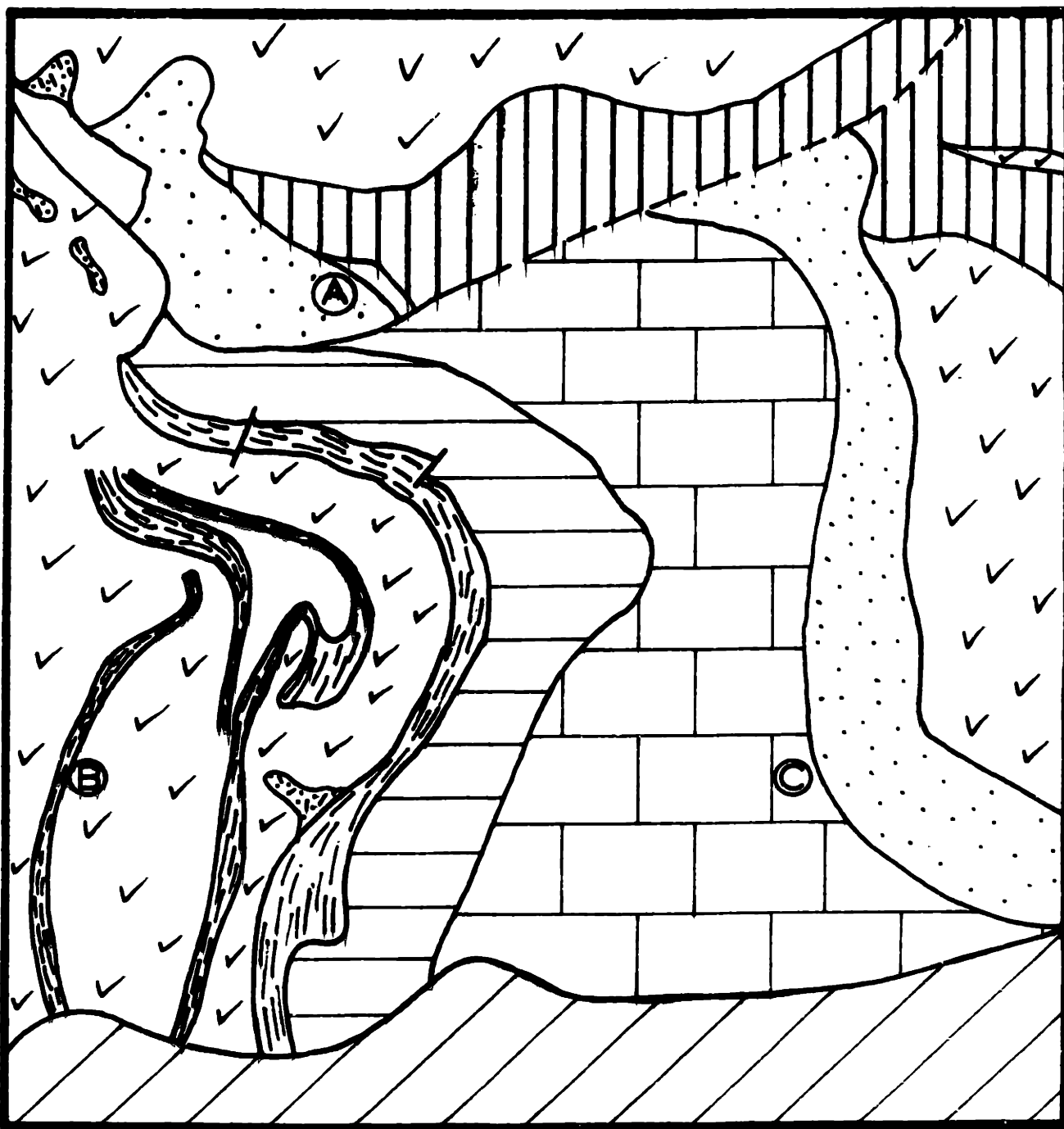
Our problem is to develop an analytical model that describes possible projections of a trend and associates a probability with each. Letting the coordinates of the plane be  $x$  and  $y$ , we do this by subjectively selecting a family of models to approximate the locus of points  $E(y | x)$ , and ascertaining the joint pdf of the coefficients of the model. For example, were we modeling the centerline of a trend on the ground surface by the line  $y = a + bx$  we would use the data already collected to evaluate the joint distribution  $f(a_0, b_0)$ , then estimate the centerline location (i.e.,  $E(y | x)$ ) beyond the data set.

A hypothetical set of exploration data is shown in Figure 26 from which we would like to estimate the probability vectors in the unexplored regions. Assume that we have familiarity with this type deposit or formation and have decided to use a 2<sup>nd</sup> order polynomial

---

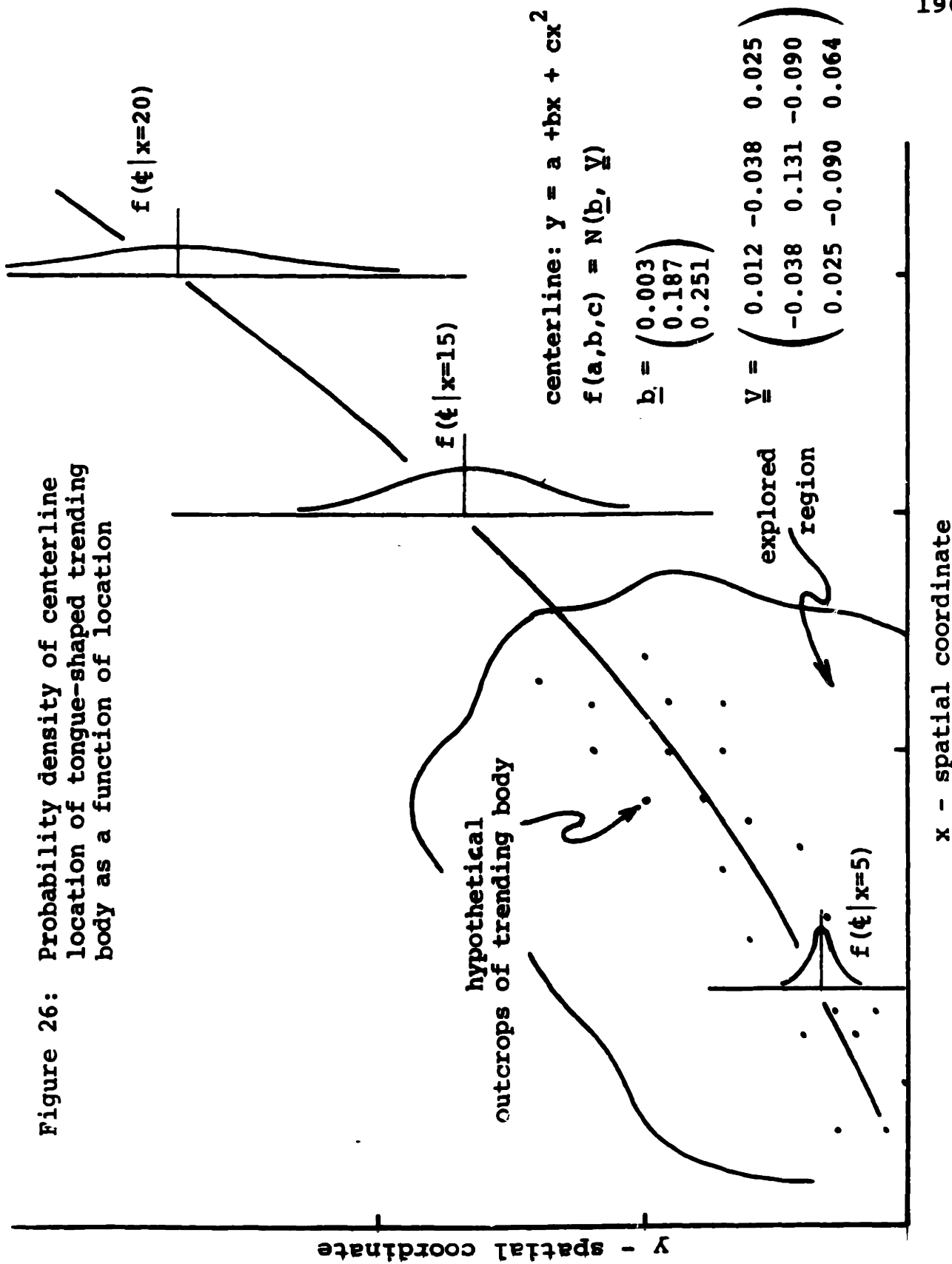
\*Examples of "tongue" shaped trends are the profile shape of sedimentary bedding, and the plan shapes of buried alluvial deposits and serpentinized sheet structures associated with ultramafic intrusions.

Figure 25



Examples of trending bodies: A--"wedge" shaped  
B--"tongue" shaped  
C--contact trend

Figure 26: Probability density of centerline location of tongue-shaped trending body as a function of location



to model the trend,  $y = a + bx + cx^2$ . We now face the problem of determining the joint pdf  $f(a_0, b_0, c_0)$ . Given a set of data and a prior joint pdf (which we shall assume to be uniform), the joint pdf after sampling is, by Bayes' Theorem,

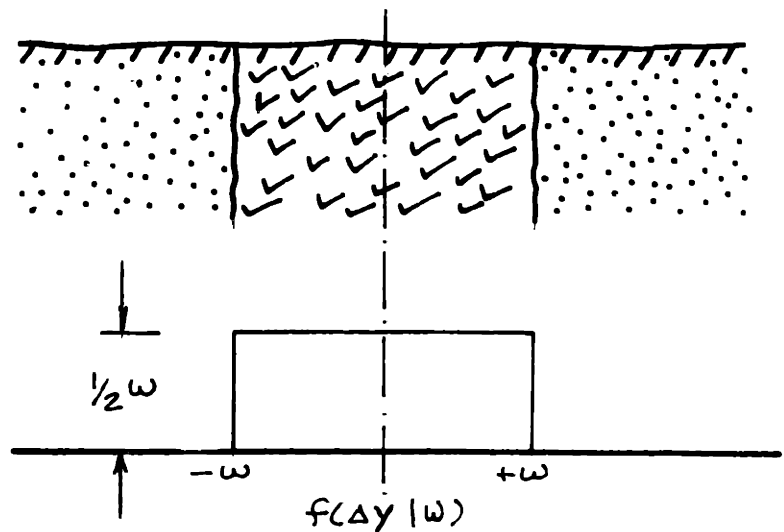
$$f(a,b,c) = \frac{f(a,b,c) L(\text{Sampled data} | abc)}{\iiint f(a,b,c) L(\text{Sampled data} | abc) da db dc} \quad 29$$

The facet of the sample data of importance in evaluating the likelihood function is the discrepancy between the value of  $y_0$  predicted by the trend line and the observed  $y$  value

$$\Delta y = |y_{\text{obs.}} - y_{\text{pred.}}| = |y_{\text{obs.}} - (a + bx_0 + cx_0^2)| \quad 30$$

The expected value of  $\Delta y$  is zero; the variance depends on the probability distribution of the body's 1/2-width,  $w$ , and some random component,  $e$ , which we will assume has zero mean, to describe local erraticness of the centerline location. If we look at a profile view of a trending body we see that the distance  $\Delta y$  is uniformly distributed between zero and  $w$  (Figure 27).

Fig. 27



Since  $w$  itself is a random variable with pdf  $f(w)$ , the marginal distribution of  $\Delta y$  (i.e., the distribution independent of  $w$ ) is

$$f_{\Delta y}(\Delta y) = \int_{\Delta y}^{\infty} \frac{1}{w} f(w) dw \quad 31$$

which equals the likelihood of  $\Delta y$

$$L(\Delta y) = f(\Delta y) \quad 32$$

Fitting a curve to the observations would be made easier if the likelihood function were a normal distribution, because this model is frequently used, and published solutions exist. We will assume that the likelihood is normally distributed only for this reason; this may not at all be a good assumption. A curve fitting exactly as the one here can be made with any likelihood function  $L(\Delta y)$ , but the computations often become intractable if a simplifying assumption is not made.

If there are  $k$  data points  $(x_i, y_i)$  we can define the matrix

$$\underline{x} = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_k & x_k^2 \end{bmatrix} \quad 33$$

and the vector

$$\underline{y} = (y_1, y_2, \dots, y_k) \quad 34$$

With uniform prior probabilities on the coefficients  $\underline{m} = (a, b, c)$ , and variance in the  $\Delta y$  likelihood function equal to  $\sigma^2$ , one can show (Pratt, et al, 1965) that the posterior distribution of  $\underline{m}$  is

multivariate normal with mean,\*

$$\underline{b} = (\underline{x}^t \underline{x})^{-1} \underline{x}^t \underline{y} \quad 35$$

and covariance matrix,

$$\underline{v} = \sigma^2 (\underline{x}^t \underline{x})^{-1} \quad 36$$

If the prior distribution is not uniform, but is multivariate normal with mean  $\underline{b}^0$  and covariance matrix  $\sigma^2 \underline{t}^{-1}$  one can show (DeGroot, 1970) that the posterior mean is

$$\underline{b} = (\underline{t} + \underline{x}^t \underline{x})^{-1} (\underline{t} \underline{b}^0 + \underline{x}^t \underline{y}) \quad 37$$

and the covariance matrix is,

$$\underline{v} = \sigma^2 (\underline{t} + \underline{x}^t \underline{x})^{-1} \quad 38$$

Generally the variance  $\sigma^2$  in the likelihood of  $\Delta y$  is not known a priori, but also assumes a posterior distribution based on the observations, resulting in a joint posterior distribution of  $\underline{m}$  and  $\sigma^2$ . Pratt, et al (1965) suggest that, when the number of observations is large, the marginal distribution of  $\underline{m}$  (i.e., the distribution of  $\underline{m}$  irrespective of  $\sigma^2$ ) can be approximated by the conditional distribution (i.e., the distribution of  $\underline{m}$ , given the value of  $\sigma^2$ ) using the estimate,

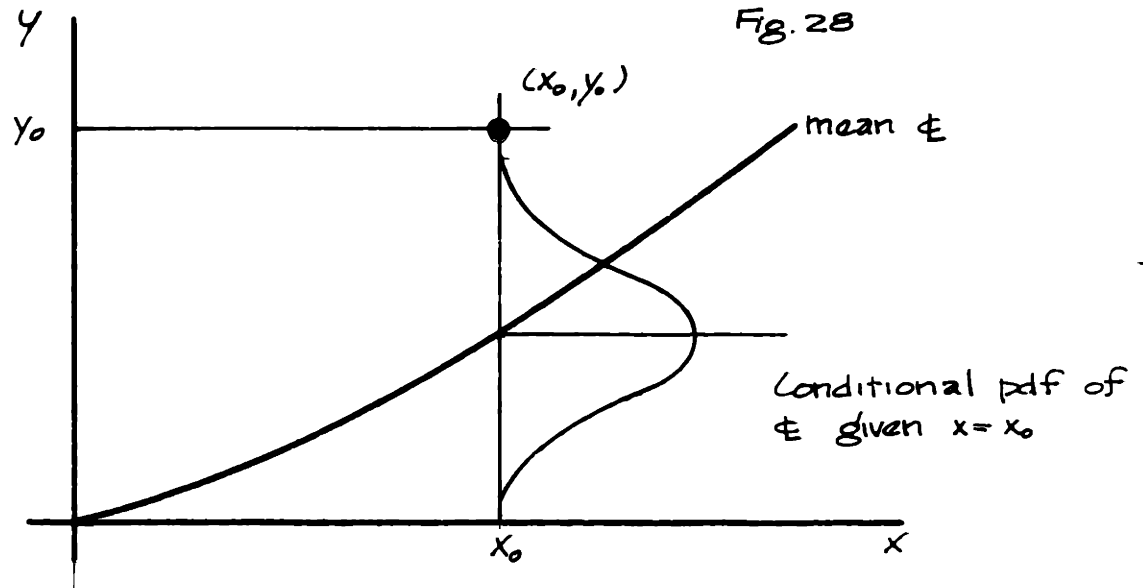
$$\sigma^2 = \frac{\sum (\Delta y)^2}{(n-r)}$$

as a certainty equivalent for  $\sigma^2$  (where  $n$  is the number of observations and  $r$  is the rank of  $\underline{x}$ ).

---

\*  $\underline{x}$  denotes a matrix, and  $\underline{x}$  denotes a vector.

Having evaluated the posterior distribution of  $\underline{m}$  from the observed data, we focus our attention on using this distribution in combination with the formation half-width distribution to evaluate probability n-tuples. The probability that point  $(x_0, y_0)$  (Figure 28) is part of the trend body,  $p_t(x_0, y_0)$ , equals the probability that the half-width is greater than or equal to the distance between  $(x_0, y_0)$  and the actual centerline.\* Since the y-coordinate of the centerline for a given value of x is uncertain,  $p_t(x_0, y_0)$  equals the conditional probability that  $w \geq |y_0 - y_1|$  where  $y_1$  is the true centerline, times the probability that  $y_1$  is the

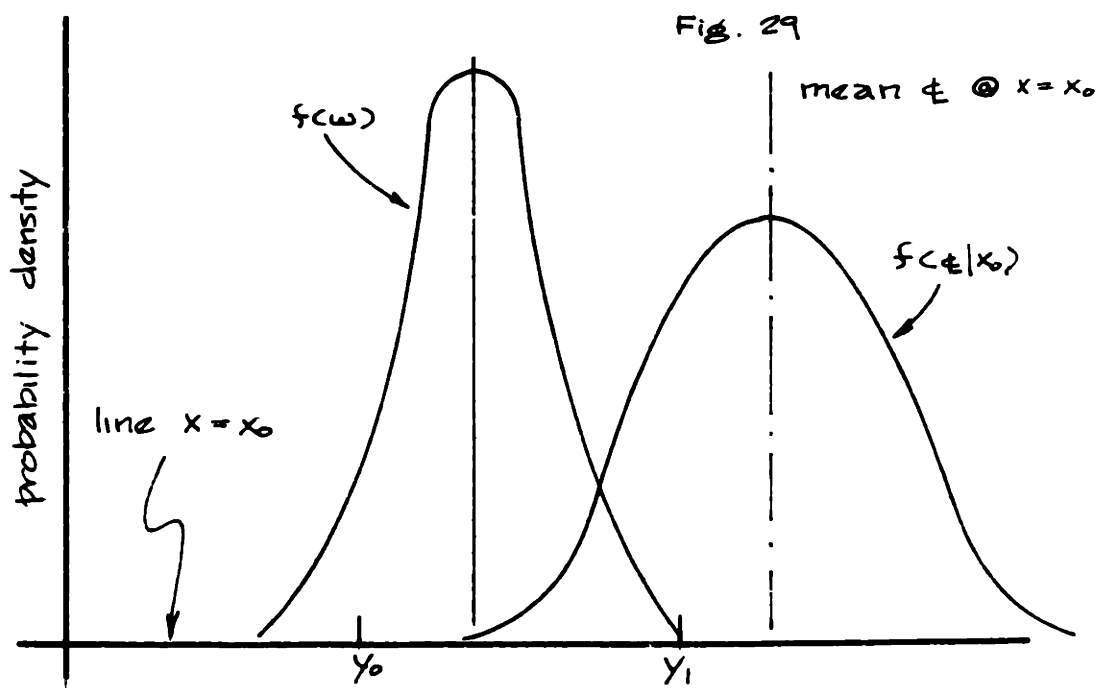


true centerline, integrated over all values of  $y_1$  (Figure 29).

\*Note that the 1/2-width is measured in the y direction and not perpendicular to the body. This may cause difficulty if the trend undergoes a large change in direction.



$$P_t(x_0, y_0) = \int_{y_1 = -\infty}^{\infty} f_{\xi|x}(y_1|x_0) (1 - F_w(y_0 - y_1)) dy_1 \quad 39$$



The cdf of  $w$ ,  $F_w(w_0)$ , equals the probability that  $w$  is less than  $w_0$ , or

$$F_w(w_0) = \int_{w=0}^{w_0} f_w(w_0) dw \quad 40$$

and  $f_{\xi|x}(y_1|x_0)$  is obtained from  $y = g(x)$  and  $f(\underline{m})$  by the common change of variable transformation

$$f_{y_1, y_2, y_3|x}(y_1, y_2, y_3|x_0) = f_{\underline{m}}(\underline{m}_0) / J\left(\frac{a, b, c}{y_1, y_2, y_3}\right) \quad 41$$

where  $J\left(\frac{a, b, c}{y_1, y_2, y_3}\right)$  is the Jacobian of  $y_i = g_i(a, b, c)$

$$J\left(\frac{a, b, c}{Y_1, Y_2, Y_3}\right) = \begin{vmatrix} \partial g_1 / \partial a & \partial g_1 / \partial b & \partial g_1 / \partial c \\ \partial g_2 / \partial a & \partial g_2 / \partial b & \partial g_2 / \partial c \\ \partial g_3 / \partial a & \partial g_3 / \partial b & \partial g_3 / \partial c \end{vmatrix} \quad 42$$

and  $y_2, y_3$  are dummy variables which are integrated out to give

$$f_{\epsilon|x}(y_1|x_0) = \int_{Y_2} \int_{Y_3} \frac{f_m(m_0)}{J\left(\frac{a, b, c}{Y_1, Y_2, Y_3}\right)} dy_2 dy_3 \quad 43$$

Our simplifying assumption that the likelihood of  $\Delta y$  is normally distributed resulted in a multivariate normal distribution for  $\underline{m}$ , and in turn results in a normally distributed  $f_{\epsilon|x}(y_1|x)$  with mean,

$$E(y | x_0) = \underline{b}x_0 \quad 44$$

where  $x_0 = (1, x_0, x_0^2)$  and  $\underline{b}$  is the mean of  $f(\underline{m})$ , and variance

$$\sigma_{\epsilon}^2 = \underline{x}^t \underline{V} \underline{x}.$$

#### Boundary Trends:

When data are available indicating a contact between two geological bodies (Figure 30), the trend of the boundary itself provides information about the classification of unexplored locations.

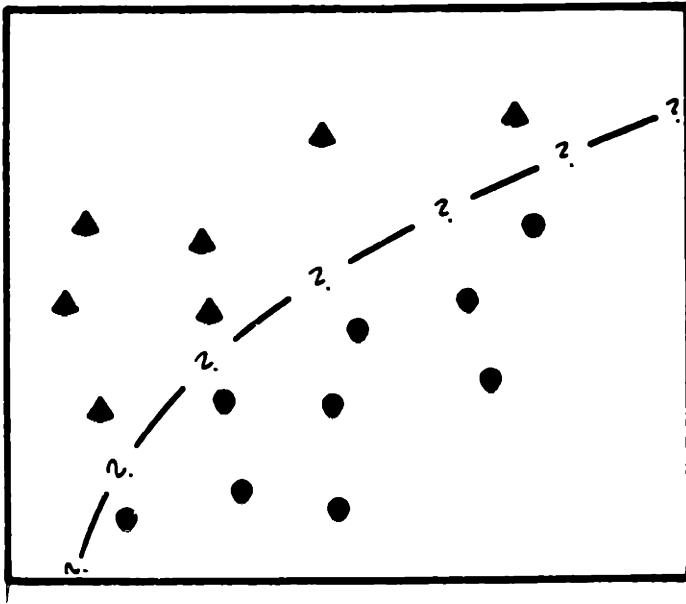


Fig. 30

Site map showing  
outcrops of two forma-  
tions & inferred contact

Given two adjacent geological bodies  $i$  and  $j$ , unexplored points are commonly considered as being in body  $i$  when  $p_i > p_j$  and in body  $j$  when  $p_i < p_j$ .<sup>\*</sup> At some point along a line joining a point in  $i$  with a point in  $j$  the most probable class changes and this point is estimated to be a contact between the two bodies. At this point  $p_i = p_j$ . So, by finding locations at which  $p_i = p_j$  we can make point estimates of the boundary dividing these two bodies, after which a curve can be fitted and the trend extrapolated.

One way to estimate locations where  $p_i = p_j$  is by applying the

---

<sup>\*</sup>This may not be appropriate for design, however, since the consequences of misclassification are not the same for all formations (e.g., more severe consequences are associated with assuming a weak, pervious formation to be a competent one than vice versa), but such considerations enter at the design stage. Here we are treating only an inference problem. The resulting information may be later used to make design decisions.

proximity rule. The likelihood ratio of class  $i$  to class  $j$  at any location is the ratio of the conditional probabilities of the observed data given the location is of class  $i$  to that given class  $j$ . When the prior probabilities are uniform, for  $p_i$  to equal  $p_j$  the likelihood ratio must equal 1.0. Consider the data shown in Figure 31. If we fix the value of  $x$  as  $x_0$  and the likelihood ratio as 1.0, we can solve for  $y_0$  and the point  $(x_0, y_0)$  will be an estimated point on the boundary between body  $i$  and body  $j$ .

$$\text{L.R.} = 1.0 = \frac{L(\text{data} | i)}{L(\text{data} | j)} \quad 45$$

From equations 19 and 22 (assuming  $p_{\xi|i}(d)$  isotropic)

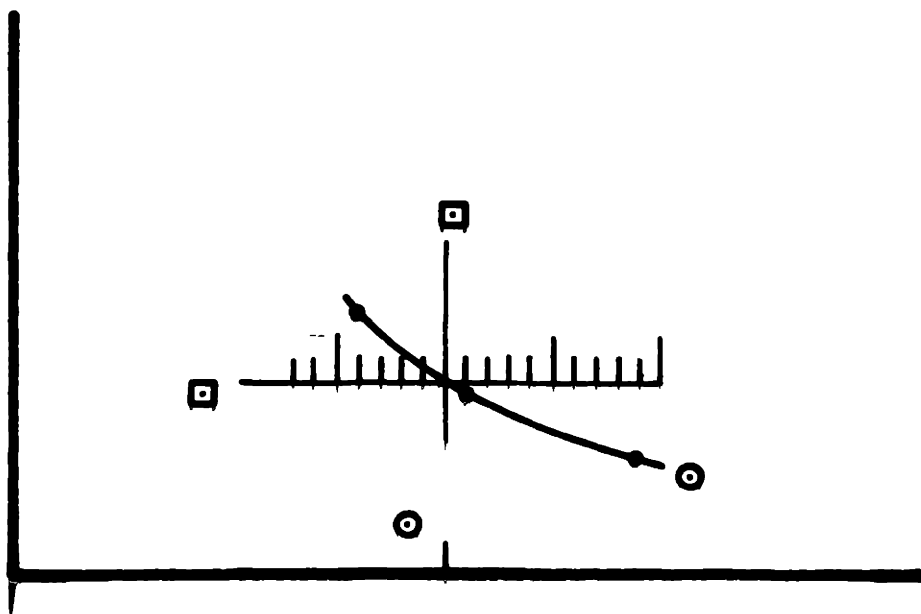
$$\begin{aligned} L(\text{data} | i) &= \prod_{z=1}^4 p_{\xi|i}(d) \quad 46 \\ &= (p_i + (1-p_i)e^{-2d_1})(p_i + (1-p_i)e^{-2d_2}) \\ &\quad \times (1 - (p_i + (1-p_i)e^{-2d_3}))(1 - (p_i + (1-p_i)e^{-2d_4})) \\ L(\text{data} | j) &= \prod_{z=1}^4 p_{\xi|j}(d) = \dots \quad 47 \end{aligned}$$

where  $d_1, d_2, d_3,$  and  $d_4$  are the distances from  $(x_0, y_0)$  to data points 1 through 4, respectively. Dividing equation 46 by 47, the L.R. (after some manipulation) is

$$\text{L.R.} = \frac{L(\text{data} | i)}{L(\text{data} | j)} = \frac{(1 + e^{-2d_1})(1 + e^{-2d_2})(1 - e^{-2d_3})(1 - e^{-2d_4})}{(1 - e^{-2d_1})(1 - e^{-2d_2})(1 + e^{-2d_3})(1 + e^{-2d_4})} \quad 48$$

which has been solved for three points along the line  $x = x_0$  and plotted. The point  $(x_0, y)$  at which  $\text{L.R.} = 1.0$  (i.e., the estimated contact) is found graphically.

Figure 31



Determination of an estimate for contact between body  $\square$  and body  $\odot$ . The L.R. at three points along  $x = x_0$  is determined and plotted. The estimated contact is that point on  $x = x_0$  at which L.R. = 1.0

Continuing so, we estimate contact points at several values of  $x_0$  (Figure 32). As a rule-of-thumb the lines of  $x_0$  should be spaced no closer than the sphere of influence of data points, because closer spacings produce residuals from the fitted curve which are autocorrelated (since they are based on the same data points) and difficult to treat analytically.

The probability that a point  $(x_0, y_0)$  is of class  $i$  equals the probability that the boundary crosses the line  $x = x_0$  at a  $y$  value less than  $y_0$ ; which, in turn, equals the probability that the trend curve plus a random component describing the variance of contact points about the trend curve has a  $y$  value less than  $y_0$ . As before, when  $f(\underline{m})$  is multivariate normal,  $f(y | x_0)$  is normal with mean  $\underline{b} \underline{x}$  and variance  $\underline{x}^t \underline{V} \underline{x}$ . If we assume the term accounting for scatter about the trend curve to be a zero mean, normally distributed r.v. with  $\sigma^2$ , then the sum of these two is also normally distributed, with mean  $\underline{m} \underline{x}$  and variance  $\underline{x}^t \underline{V} \underline{x} + \sigma^2$ . An interpolation of the boundary estimated in Figure 32 is shown in Figure 33.

#### Continuity:

In any mapping work involving trends a near-classic question is how far trends can be extrapolated. One limit, certainly, is set by dispersion of probability density of the trend line away from the observation set. At locations sufficiently removed, the probability vectors approach the prior probabilities. For lesser distances the

Figure 32

Hypothetical outcrops of adjacent formations  
 and determination of estimated contact points.  
 At each of seven arbitrarily selected  
 values of x, the value of  
 y making L.R. = 1.0 is  
 graphically determined

plan view of mapped area of site

● }  
 □ } represent different  
 formations

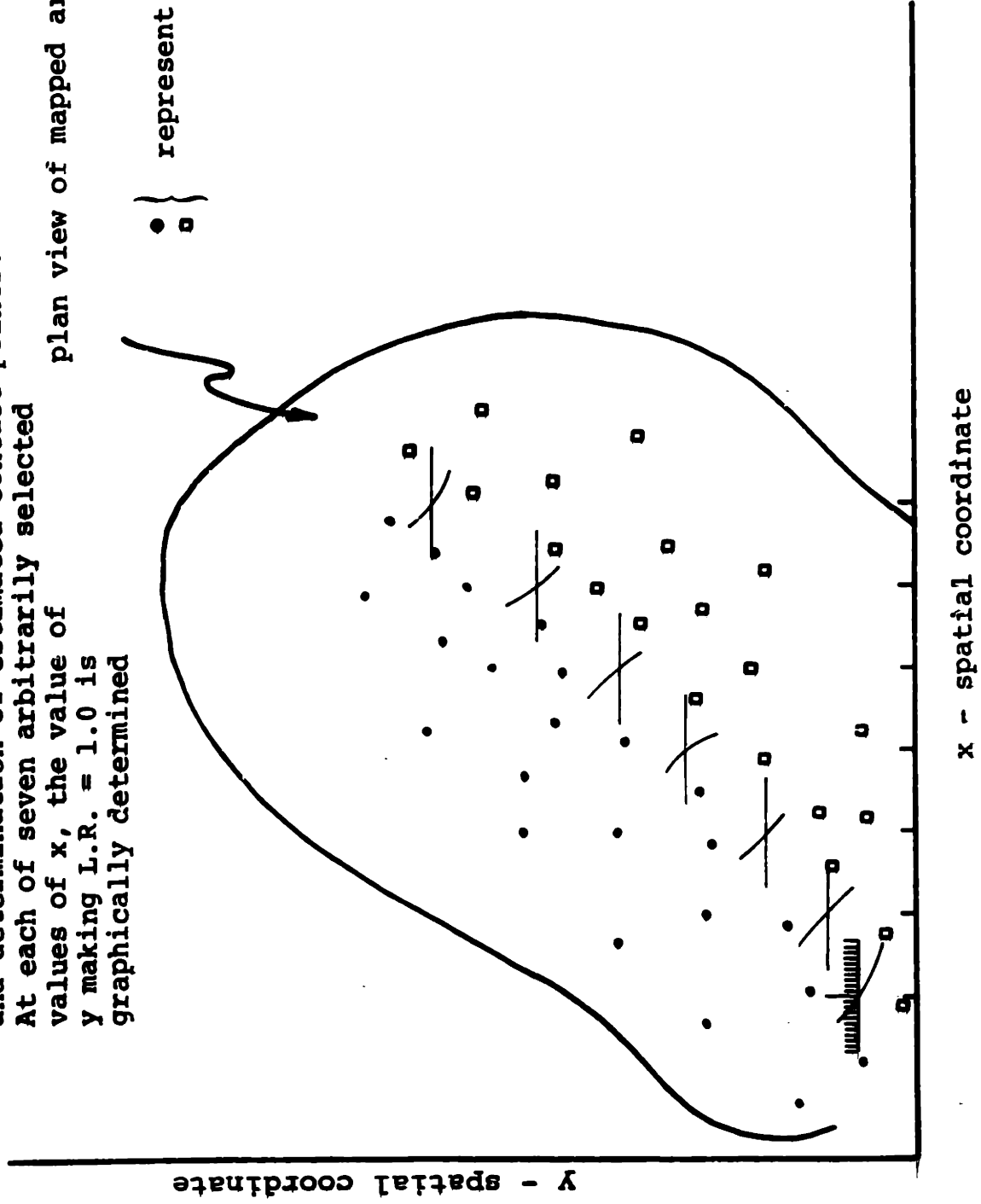
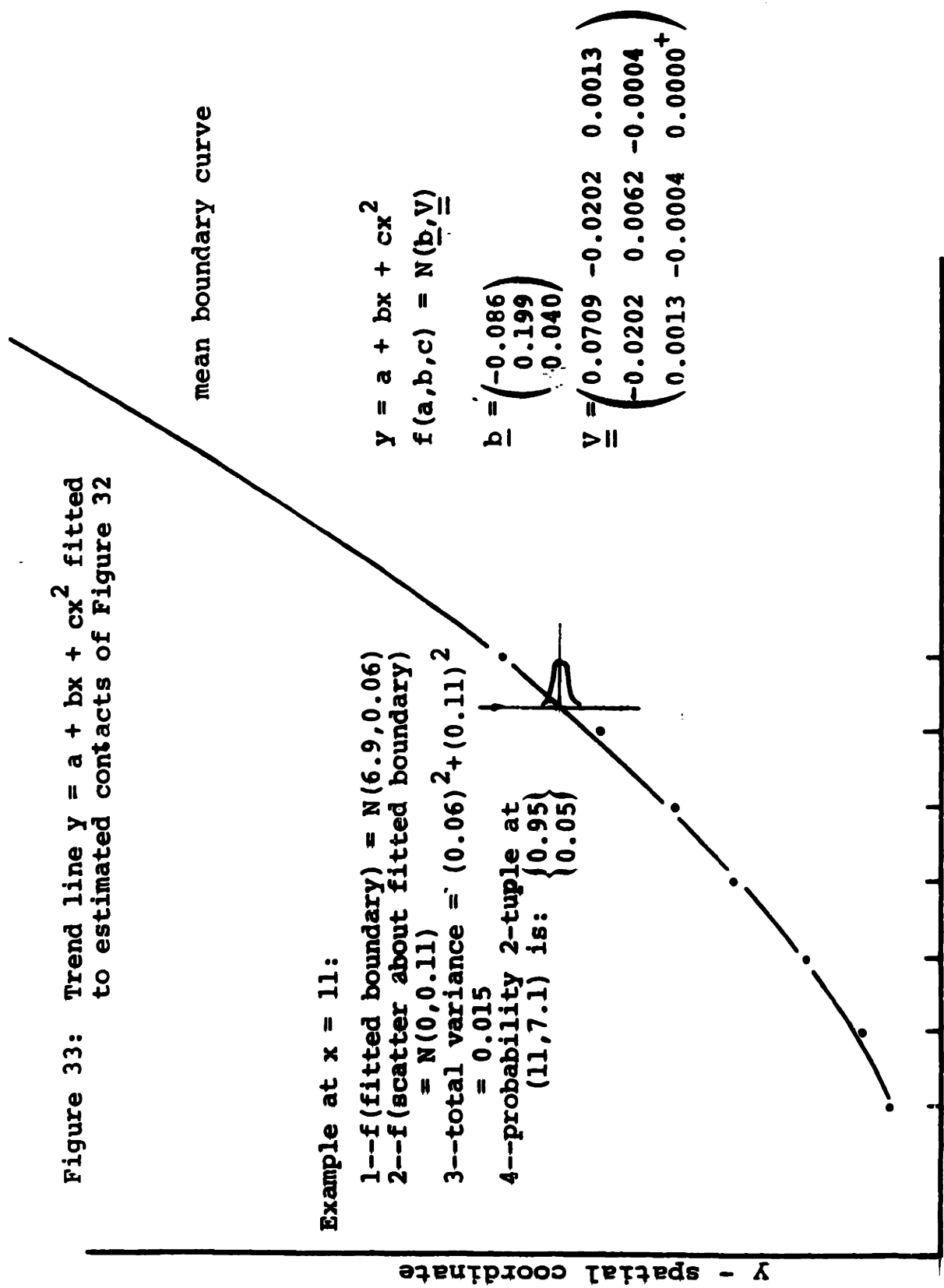


Figure 33: Trend line  $y = a + bx + cx^2$  fitted to estimated contacts of Figure 32



x - spatial coordinate



probability of the trending body being continuous must be evaluated by other means, usually based on insufficient information.

In regions where mapping has been undertaken nearby, or with bodies for which past experience is available, pdf's of body length can sometimes be constructed -- if only subjectively.

Let the pdf of body length be  $f_l(l_0)$ . The conditional probability density of the body having length  $l_0$  when we know that the length is at least  $l_1$  (Figure 34) is

$$f_{l|l_1}(l_0|l_1) = \begin{cases} \frac{f_l(l_0)}{\int_{l_1}^{\infty} f_l(l_0) dl_0} & l_0: l_0 \geq l_1 \\ 0 & \text{otherwise} \end{cases} \quad 49$$

Fig. 34

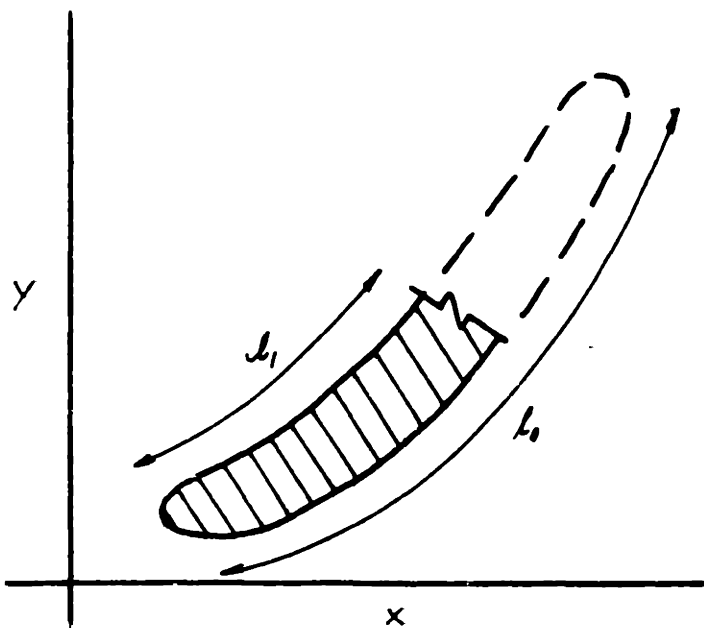
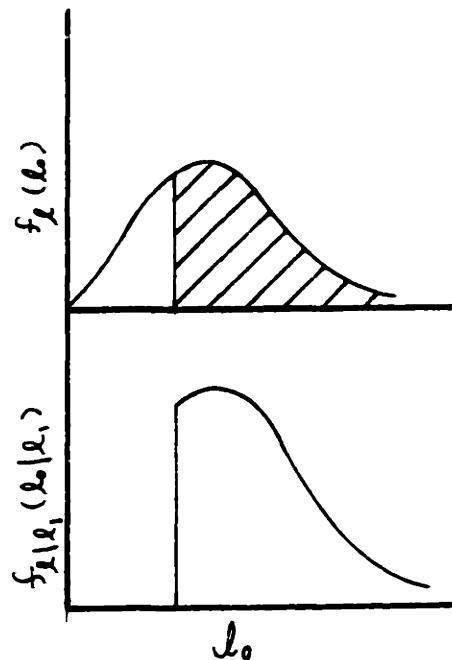


Fig. 35



which is just the original density normalized so that the area under the pdf from  $l_1$  to  $\infty$  equals 1.0 (Figure 35). Using equation 49 the probability of the body being continuous to any length  $l_0$  is

$$\begin{aligned} \Pr \{ \text{body con't to } l_0 \mid \text{con't to } l_1 \} \\ &= \Pr \{ \text{actual length} \geq l_0 \} \\ &= \int_{l_0}^{\infty} f_{l|l_1}(l_0|l_1) dl_0 \end{aligned} \quad 50$$

The probability n-tuple at any distal point must be a combination of the probability n-tuple inferred when the trend is continuous and that inferred when the trend is not continuous, weighted by the respective probabilities,

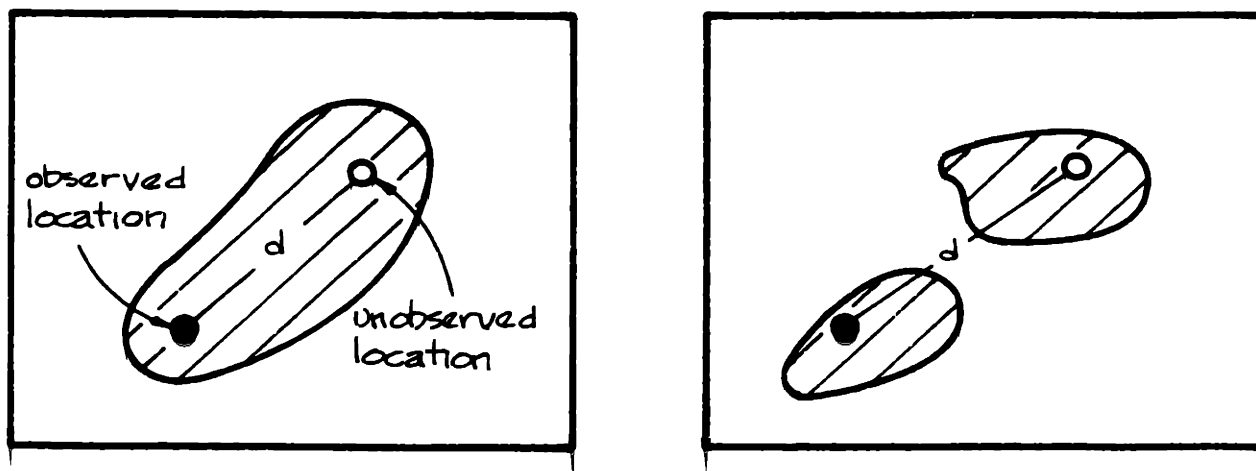
$$(p_i) = p(\text{con't}) (p_i | \text{con't.}) + p(\text{not con't.}) (p_i | \text{not con't.}) \quad 51$$

From the previous discussion we have a rule for inferring  $(p_i | \text{con't.})$  however, we lack a rule for inferring  $(p_i | \text{not con't.})$ . At first one might suggest using the proximity rule, but the proximity rule predicts that a string of trend body type elements can occur with probability greater than zero -- a logical contradiction since we've assumed the trend to be discontinuous. What we need to infer probability vectors conditioned on the trend not being continuous is a modified proximity rule.

In general terms, a point at distance  $d$  from an observed point of composition  $i$  can also be of composition  $i$  in two ways. Either the body in which the observation was made extends the distance  $d$ ,

or a second body of composition  $i$  exists (Figure 36). We can

Fig. 36

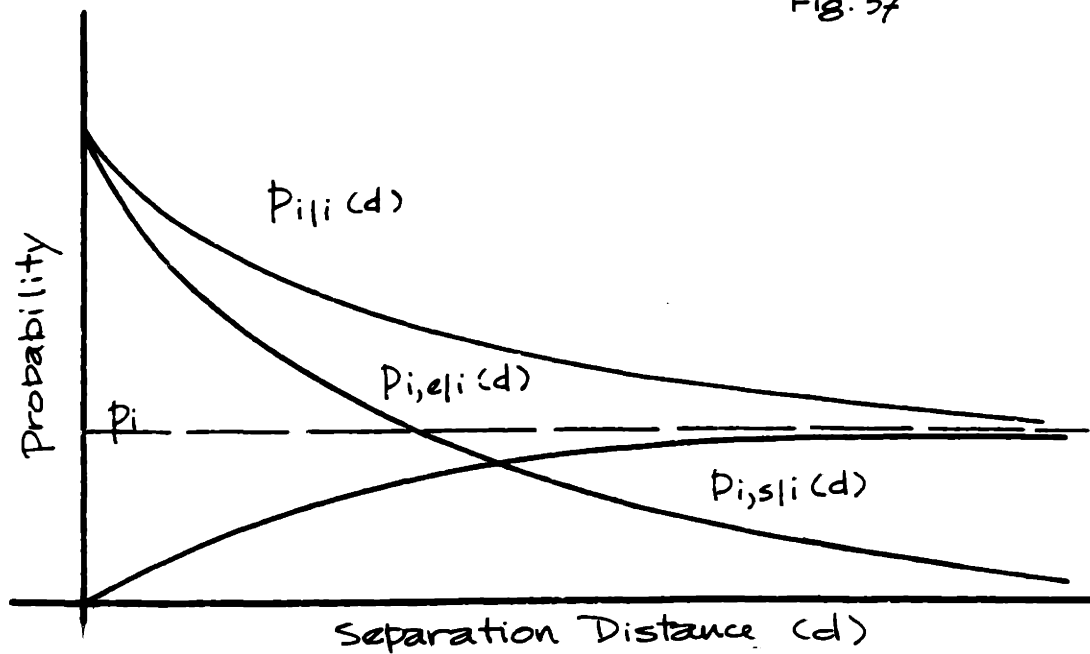


decompose the proximity rule into these two components, one accounting for bodies extending distance  $d$  and one for separate bodies.

$$P_{i|i}(d) = P_{i,e|i}(d) + P_{i,s|i}(d) \quad 52$$

where  $P_{i|i}(d)$  is the conditional probability of a point being of composition  $i$  when an observation of  $i$  has been made at a distance  $d$ , this is the probability from both sources;  $P_{i,e|i}(d)$  is the contribution due to bodies extending to  $d$ ; and  $P_{i,s|i}(d)$  is the contribution due to separate bodies. Graphed, these functions might look like Figure 37, but, of course, they require empirical assessment.

Fig. 37



The probability vector at some point distant from the trend would be

$$(P_i) = P_{tc} (P_i)_{tc} + (1 - P_{tc}) (P_i)_{mp} \quad 53$$

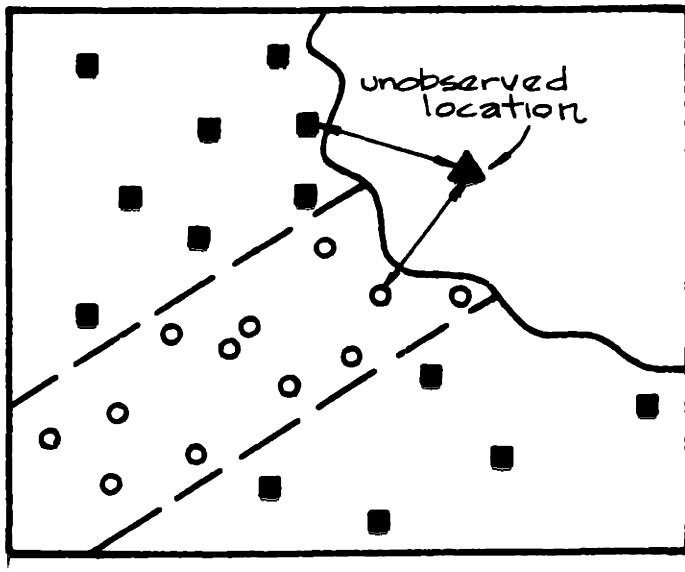
where  $p_{tc}$  is the probability that the trend is continuous,  $(P_i)_{tc}$  is the probability vector determined by the trend extrapolation rule, and  $(P_i)_{mp}$  is the probability vector determined by the modified proximity rule.

$$(P_i)_{mp} = \left\{ \frac{P_i L(\text{obs}'ns | i, \text{mp rule})}{\sum_i P_i L(\text{obs}'ns | i, \text{mp rule})} \right\} \quad 54$$

$$= \left\{ \frac{P_i \prod_{z_1} P_{e,s|i}(d) \prod_{z_2} P_{s|i}(d)}{\sum_i P_i \prod_{z_1} P_{e,s|i}(d) \prod_{z_2} P_{s|i}(d)} \right\}$$

where  $z_1$  is the set of observations belonging to the trending body and  $z_2$  is the set of observations not belonging to the trending body. The likelihood of the former is based on  $p_{i,s|i}(d)$  while the likelihood of the latter is based on the total proximity rule  $p_{i|i}(d)$ .

Fig. 38



when the trend is assumed not to continue the likelihood of  $\circ$  is based on  $p_{i,s|i}(d)$ , while that of  $\blacksquare$  is based on  $p_{i|i}(d)$ .

Estimating the various parameters and component functions seems no easy task, and requires substantial quantities of data. The probability of the trend continuing can be evaluated if a pdf of body length  $l_0$  is available. The functions  $p_{i,e|i}(d)$  and  $p_{i,s|i}(d)$  must be estimated from previous mappings of geologically similar regions using the sampling technique suggested on page 187. The function  $p_{i,e|i}(d)$  might be determined analytically if the size distribution of bodies of material  $i$  is known or estimated, as can  $p_{i,s|i}(d)$  if the distribution of body spacings

is known; but if a map of a nearby or similar region is available, empirical estimations are probably simpler and better approximations.

#### Trending Boundaries:

A second technique for estimating the length of trending bodies is to extrapolate the contacts or boundaries of the body and estimate the location of zero width. Before we have always implicitly assumed that the width of the body remained essentially constant along the body's length. Here we make a new assumption: we assume that width can be expressed as a trending function of length along the body (i.e., "wedge" shaped bodies).

The appropriate curve to fit to boundaries between formations depends, as always, on local geology and judgement -- which is perfectly admissible. Consider Figure 39 in which the trends of two boundaries are shown.

Fig. 39

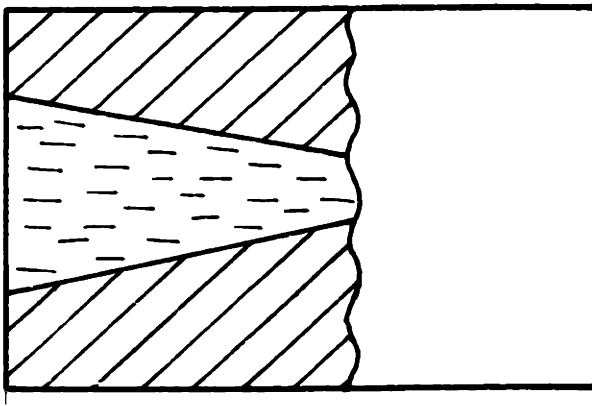
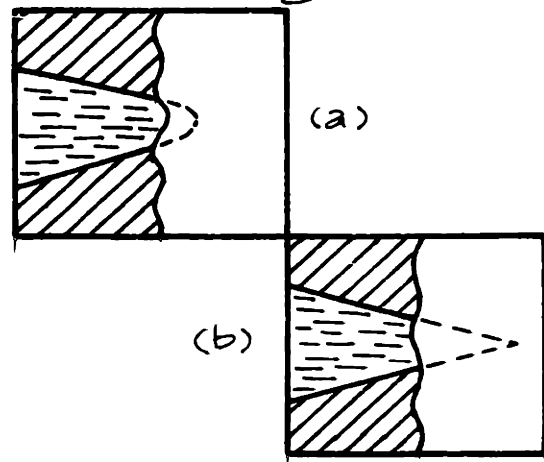


Fig. 40



Were the body a plunging anticline the boundaries might be extrapolated as in Figure 40a; were it fault emplaced they might be extrapolated as in 40b.

The projected length of a body can be thought of as the distance to the intersection of its boundaries, that is, the distance at which its width equals zero. Were we able to estimate the relationship between distance along the body,  $x$ , and width,  $w$ , we could then estimate the pdf of  $x$  conditioned on width equalling zero,  $f(x|w=0)$  from which we could infer the probability of the trend extending given distances.

When one attempts to extrapolate real bodies on geological maps by "covering" one end of a formation and fitting curves to the other end, he finds that low order curves seldom predict well, and high order curves become contorted once they leave the data set to which they have been fit. Statisticians never cease warning that one should not extrapolate empirically fitted curves unless a theoretical model leads to the particular family of curves fitted. This brings us to the distasteful conclusion that we may presume more sophistication than we are justified to by extrapolating geological trends. Of course, only extensive empirical comparisons will show for sure, and this would be a logical next step. The scope of the present investigation did not allow empirical comparisons with a sufficiently large number of previously mapped bodies to draw significant conclusions about the validity of extrapolations.

Much of our work in site mapping deals with interpolation --

perhaps even more than extrapolation. The approaches discussed in the present section are applicable to the interpolation problem as well as the extrapolation one (see, e.g., Figure 33), and the degree of confirmation of fitted curves for inferring probabilities of classification between observed points is considerably greater than for outside observed points (for a philosophical discussion of this point see Nagel's classic paper, 1938). In future investigations we may find that one can treat interpolation with an acceptable level of assurance, but not extrapolation.

### Section 5.2.3 Uncertainty Functions and Entropy

We would like to apply Bayesian decision theory to the problem of forming strategies for the allocation of exploration effort in mapping, but to do so we must possess the basic elements of a decision problem. These are a decision space,  $A$ , on which all possible terminal actions are defined, a state space (sometimes called a parameter space),  $\Theta$ , on which all possible states of nature are defined, and a real valued function  $u(a, \theta)$ , called a utility function, which is defined on the product space  $A \times \Theta$  (i.e., a real-value  $u(a, \theta)$  can be associated with every action-state pair). Thus, were we designing a foundation on a clay stratum of uncertain thickness, we could express our decision in terms of Bayesian theory: We possess a decision space ( $A = \{ \text{foundation type and dimensions} \}$ ), a state space ( $\Theta = \{ \text{thickness of stratum} \}$ ), and a utility function ( $u(a, \theta) = \text{some function of settlement with design specification}$



and thickness as parameters). We could include decisions on obtaining information about deposit thickness by considering their expected effects on the certainty with which we know the true state of nature prior to making design decisions (which in turn can be expressed as changes in the expected utility of action-state pairs). This requires two additional elements: an outcome space,  $Z$ , on which all possible outcomes of experiments to obtain information are defined; and a family of probability distributions,  $L(z|\theta)$ , called likelihood functions, defined on the product space  $Z \times \Theta$ , which express the probability of outcomes conditioned on the state of nature.

When we consider the application of Bayesian decision theory to mapping (i.e., recognition and reconstruction) a dilemma arises: in most mapping problems we know neither a decision space (for design) nor a utility function. Mapping is usually undertaken before proposed facilities are precisely located, and sometimes even before they are generally located. Commonly, the many uses to which recognition and reconstruction information will be applied cannot be foreseen at the time the information is collected; and even if they could, their number and variety would make a decision formulation intractable.\* So, given this state of affairs, how can we speak

---

\*These remarks are aimed at general mapping accompanying most exploration programs in rock masses. We can certainly conceive of situations in which mapping information is desired for specific design decisions (e.g., the thickness of a compressible stratum for settlement oriented design; the precise orientation and dimensions of a fractured zone for design against shear failure), and which may be handled using Bayesian decision theory. However, these situations represent special cases.

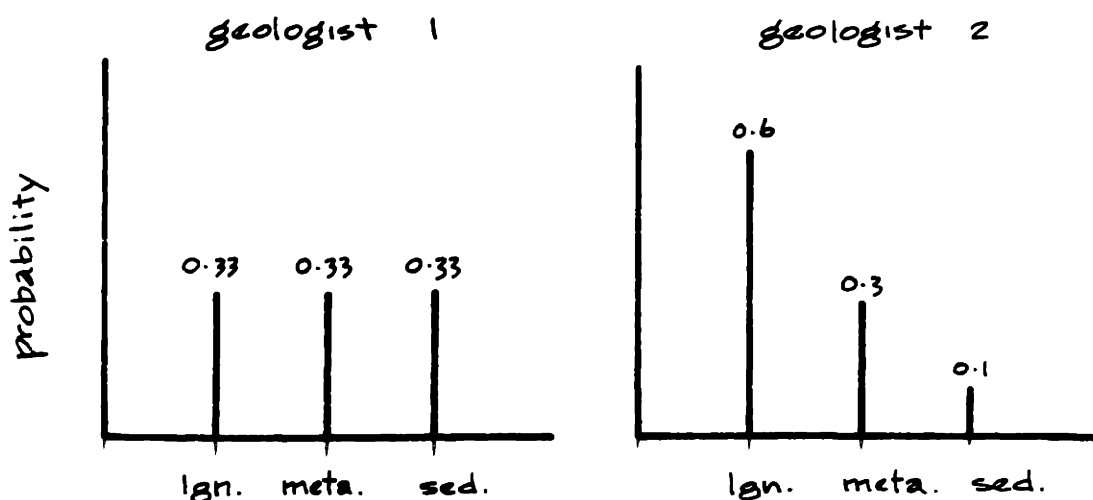
of optimizing allocations?

With no way of relating exploration information to design decisions and expected utilities, we have no way of analytically optimizing the amount of exploration effort expended commensurate with design problems; however, for a given amount of effort we do have intuitive concepts of what obtaining "the most information" from an expenditure means, even if we cannot relate this information beforehand to design decisions. Somehow we need a measure of that information, which can be used to evaluate effectiveness. Such an approach to optimization is not as desirable as a direct decision-theoretic one, but for the present must suffice (unless the precise decision space of design is known).

As we discussed in the last section, mapping-information takes the form of an association of an ordered  $n$ -tuple to each point in the space mapped. Considering one of these points, what is meant by directly measuring the information contributed by exploration? We can think of information, and will so define it, as being the change in uncertainty of correct classification. Uncertainty, in turn, we will directly measure.

What is meant by a direct measure of uncertainty? Consider two geologists faced with determining the origin of a particular body. After viewing the evidence they express their belief as in Figure 41.

Fig. 41



Intuitively we might say that the first geologist is less certain of the classification than the second, and we would like to define a function which in some way is a measure of this uncertainty.

For example, one such function is

$$\gamma = 1 - \max_i \{p_1, \dots, p_i, \dots, p_n\} \quad 55$$

where  $p_i$  is the probability of the  $i^{\text{th}}$  class or event. By this measure geologist 1's uncertainty is

$$\gamma_1 = 1 - 0.33 = 0.67$$

and 2's is

$$\gamma_2 = 1 - 0.60 = 0.40$$

An uncertainty function typically must be a minimum at those distributions  $\{p_i\}$  for which some  $p_i = 1.0$ , and a maximum at those distributions near  $\{1/n, \dots, 1/n\}$ ; it must be non-negative.

The "information",  $I$ , in an experiment is defined as the

expected change in uncertainty between the prior and the posterior distribution.

If we accept the common assumption (DeGroot, 1962) that an experiment can at worst contain no information,  $I$  must be non-negative. If  $\mathcal{V}$  is a real valued function defined on the set of all possible probability distributions  $\{p_i\}$ , DeGroot has shown that  $I$  is non-negative if and only if  $\mathcal{V}$  is concave. That is, if and only if

$$\mathcal{V}(a\{p_i\} + (1-a)\{q_i\}) \geq a\mathcal{V}\{p_i\} + (1-a)\mathcal{V}\{q_i\} \quad 56$$

where  $\{p_i\}$  and  $\{q_i\}$  are any probability distributions and  $a$  is any constant between zero and 1.0. Therefore, any non-negative concave function  $\mathcal{V}$  defined on the set of all probability distributions will be called an uncertainty function. Examples of uncertainty functions, other than (1), are  $[1 - \sum_{i=1}^n p_i^2]$  57 and Shannon's entropy function

$$H = -c \sum_{i=1}^n p_i \log p_i \quad 58$$

in which  $c$  is a constant.

If we further restrict the set of uncertainty functions that we will consider to those possessing the properties

1.  $\mathcal{V}$  continuous in  $p_i$
2. For  $\{p_i\} = \{1/n, \dots, 1/n\}$ ,  $\mathcal{V}\{p_i\}$  increases with increasing  $n$
3. The uncertainty in a compound event equals the weighted sum of the uncertainty in successive events.

$$\mathcal{V}(AB) = \mathcal{V}(A) + \mathcal{V}(B|A)$$

Shannon (1948) has shown that the only one satisfying these constraints is 58.

When entropy is extended to continuous distributions it becomes,

$$H = \lim_{\Delta x_i \rightarrow 0} \left[ - \sum_i p(x_i) \Delta x_i \log p(x_i) \Delta x_i \right] \quad 59$$

$$= - \int f_x(x_0) \log f_x(x_0) dx_0 - \lim_{\Delta x_i \rightarrow 0} \sum p(x_i) [\log \Delta x_i] \Delta x_i \quad 60$$

rather than the analogue of 58,

$$- \int f_x(x_0) \log f_x(x_0) dx. \quad 61$$

Note that the second term in 60 goes to infinity as  $\Delta x_i \rightarrow 0$ .

For simplicity, entropy in continuous distributions is commonly defined at 61, or, in  $n$  dimensions,

$$H = - \int_{x_1} \cdots \int_{x_n} f(x_1, \dots, x_n) \log f(x_1, \dots, x_n) dx_1 \cdots dx_n \quad 62$$

despite difference from the limiting form.\*

The entropy of discrete distributions deals with probabilities of events, which, having definite values, are invariant with respect to coordinate changes. Equation 61, however, deals with density distributions, whose form changes as the coordinate system changes according to

$$f(y_1, \dots, y_n) = f(x_1, \dots, x_n) / J \left( \frac{x_1 \cdots x_n}{y_1 \cdots y_n} \right) \quad 63$$

---

\* Since only changes in entropy are usually considered, the second term in equation 60 drops out and some workers have used 60 as the continuous definition of entropy.

where  $y_1 \dots y_n$  is the new coordinate space,  $x_1 \dots x_n$  is the old coordinate space and  $J \left( \frac{x_1 \dots x_n}{y_1 \dots y_n} \right)$  denotes the Jacobian

of the transformation. So, the entropy of continuous distributions is not invariant with respect to coordinate changes, but equals the former value minus the expected logarithm of the Jacobian of the transformation. In the continuous case entropy must be considered a measure of uncertainty relative to a certain standard (i.e., the coordinate system  $x_1, \dots, x_n$ ). Differences in entropies (e.g., information), nevertheless, are invariant with respect to coordinate transformations.

The entropy function has properties which make it desirable as an uncertainty measure for mapping. An important characteristic, relative to other common measures, is that entropy is dependent on the probabilities of every classification not simply the most or least probable. Design decisions, in general, depend on the probabilities of other than the most probable classification, and changes in these should affect our measure. Secondly, entropy is a continuous function of the probabilities of every classification and thus is easier to treat analytically than discontinuous functions. Lastly, entropy is additive over compound events, which means that data of the state-of-nature at two locations leads to the same change in uncertainty whether it is collected for both locations simultaneously or in different stages.

We can use entropy as a measure of uncertainty in a total map by taking spatial averages, which for a discretized map of  $n$  elements is,

$$\bar{H} = \frac{1}{n} \sum_{i=1}^n H_i = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^k p_{ij} \log p_{ij} \quad 64$$

where  $p_{ij}$  is the probability of class  $j$  at element  $i$ , and  $k$  is the number of possible classes. If  $\bar{H}$  is a measure of uncertainty in a geological map, expected changes in  $\bar{H}$  can be used as a measure of the information gained by effort allocations and can be used as a criterion of optimality. That is, other things equal, we will favor effort allocations which offer the largest expected change in  $\bar{H}$ .

In general, the continuous function  $H(x,y)$  which assigns an entropy value to each point on the map will be complicated and an approximation by homogeneous discrete elements will be necessary. As always, optimization depends both upon the expectations of outcomes of effort allocations and upon the rules of inference in which those outcomes are used.

The use of entropy in describing pattern reconstruction has not been treated in the literature, although two somewhat similar applications have been suggested. Danskin (1964) proposed using entropy as a criterion in searching for enemy military installations, and Dowds (1961) mentioned that an entropy measure might be used to describe success ratios in oil exploration.

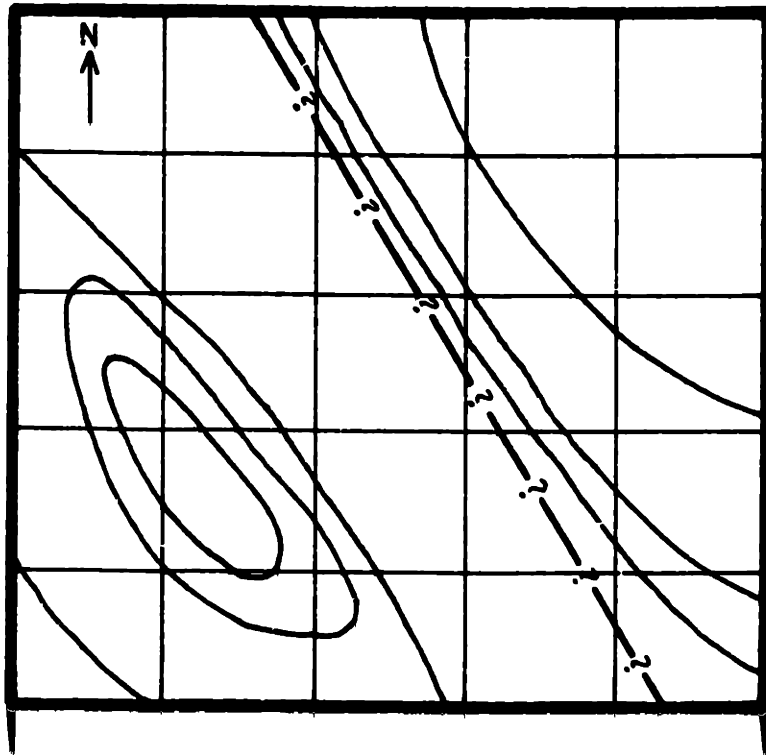


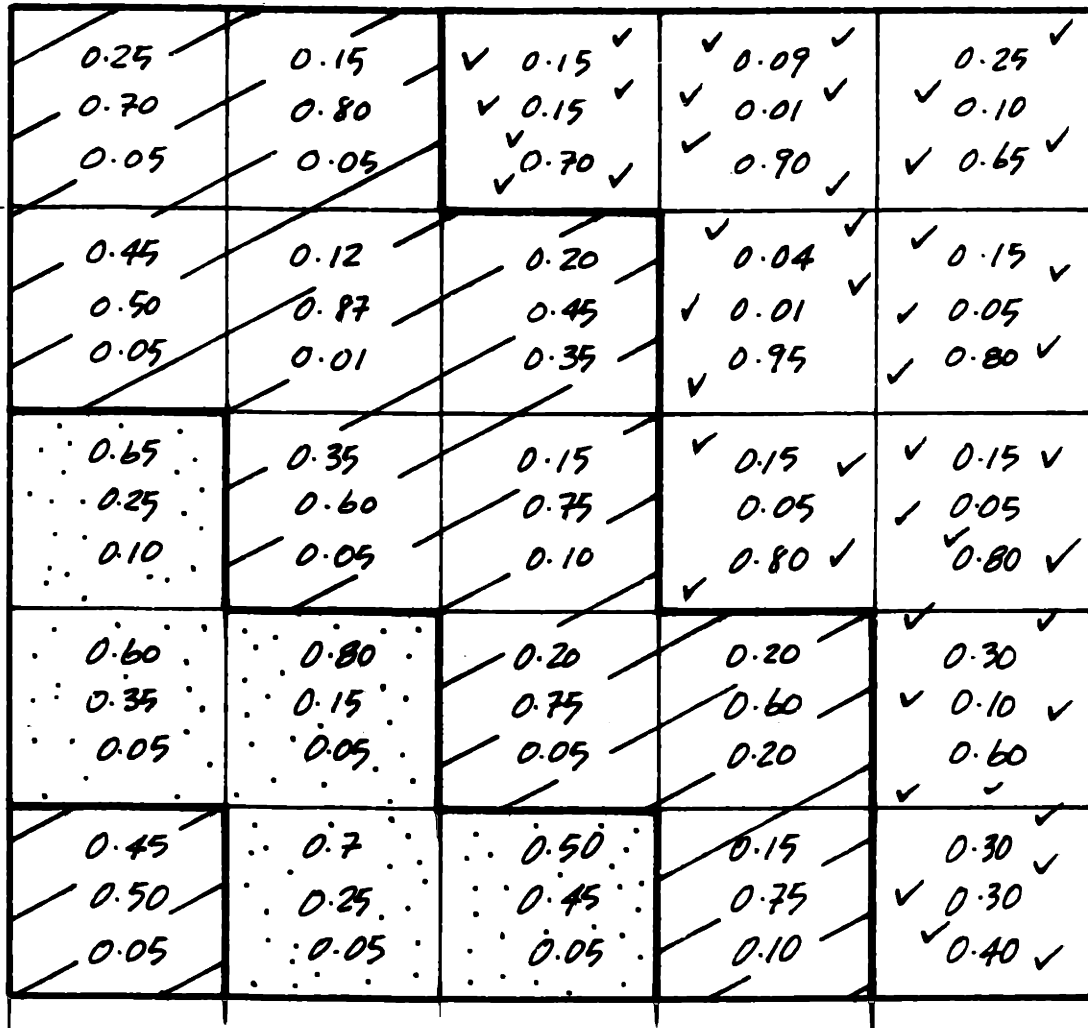
Fig. E-1

#### Example: Entropy of a Discretized Probabilistic Map

Shown above is a topographic map of a hypothetical dam site. After initially viewing the site the engineer has concluded that the mountain in the SW of the site is a "hog-back" composed of a sandstone-shale sequence, while the NE mountain is granitic. Because of its greater resistance to weathering he has decided that the summit of the SW mountain probably is sandstone, while the valley is probably of less-resistant shale (discluding faults for the moment). The engineer subjectively constructs the array of probability vectors shown on the next page and now wishes to determine the gross "uncertainty" in his evaluation before proceeding on to allocating effort.



Figure E-2



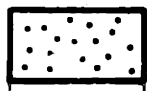
Discretized Map showing subjective probability 3-tuples

$p(\text{sandstone})$

$p(\text{shale})$

$p(\text{granitic})$

for each element. Partitioning according to the common rule: maximum probability is shown,



sandstone



shale



granite

Entropy for probability 3-tuple

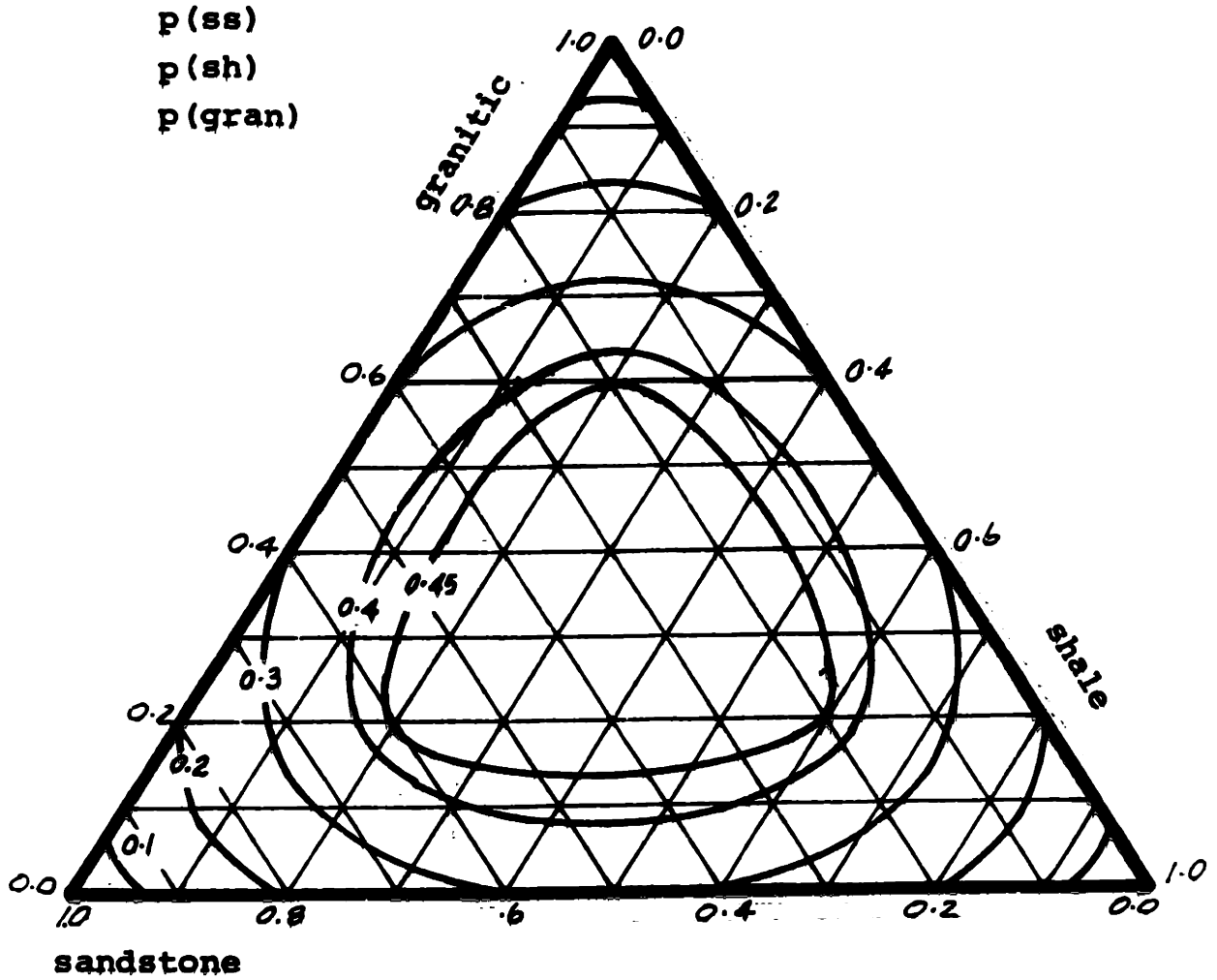


Figure E-3

0.32	0.27	0.36	0.16	0.37
0.37	0.18	0.48	0.10	0.27
0.37	0.36	0.32	0.27	0.27
0.36	0.27	0.32	0.45	0.39
0.37	0.32	0.37	0.32	0.47

$$\bar{H} = 0.32$$

FIG. E-4  
Entropy of Elemental Areas

The engineer's gross uncertainty, manifested in the probabilistic map of Figure E-2, equals the ave. entropy, or  $\bar{H} = 0.32$ .

We know from our discussion of proximity considerations in Section 4.2.2 that the compositions of proximal elements are not mutually independent. The probability of element B being sandstone is increased if we are given that A is sandstone; or more generally, the probability of B being sandstone is increased if the probability of A being sandstone is increased. One might attempt to balance prior probability assignments to conform with proximity rules; but we submit that this is not justified in light of the approximated nature of subjective probability assessments, intuitive consideration of dependence in assessing probabilities, and large changes in the probabilities caused by exploration observations.

#### Section 5.2.4 Optimal Allocation

A major benefit of a quantitative approach to exploration is the capability of maximizing the efficiency of effort allocations. This optimization depends both on the criterion of optimality and on the rules used to infer states-of-nature from observations. In preceding sections we have discussed both. We now face the problem of combining them mathematically to obtain optimal strategies of allocation.

As will become apparent, the computational difficulty of optimizing reconstruction allocation is considerable, unless unrealistically simplifying assumptions are made. Although the entropy criterion itself does not add appreciable complexity, the number of random variables is great and the rules for inference are involved. Deriving mathematical solutions to these formulations is beyond the scope of this work, and in the particular case of sequential allocation, solutions to the whole class of problems can only be obtained by dynamic programming or direct enumeration. Computational problems must be overcome before real allocations can be optimized.

We divide our discussion into two parts. The first treats single-stage allocations in which all available effort is allocated at one time, and the second treats sequential allocations in which after each unit of effort is allocated the probability distributions on the state-of-nature are modified and new optimal strategies determined.

### Single Stage Allocations

Let the site be composed of  $m$  elemental areas within each of which geology is homogeneous. To each element we may allocate a quantity of effort  $\varphi(k)$ , where  $\varphi(\cdot)$  is a continuous non-negative function subject to the constraint,

$$\sum_{k=1}^m \varphi(k) = \Phi \quad 65$$

in which  $\Phi$ , the total effort, is a constant.

Associated with element  $k$  is another non-negative function  $I(\varphi(k))$  which describes the expected change in entropy (i.e., information) in the total map by allocating  $\varphi(k)$  amount of effort to element  $k$ .

The optimization problem is to maximize the sum of  $I(\varphi(k))$  over  $k$ , subject to the constraint 65. Computational difficulty arises because the  $k$  functions  $\varphi(k)$  are not generally independent.

To begin we shall consider the simple case of independent elements, that is, effort allocated to element  $k$  affects only the entropy of  $k$  (ensuring that the  $I(\varphi(k))$  are independent).

Because there are no interaction effects, the functions  $I(\varphi(k))$  may be easily formulated and the optimal allocation depends only on their shape, which Danskin (1964) has demonstrated to be concave-convex when certain assumptions are made. A set of convex functions is illustrated in Figure 42a, and a set of concave functions in 42b. Maximizing the sum of convex or concave functions is a familiar problem in operations research (Churchman, et al, 1957).

Fig. 42a

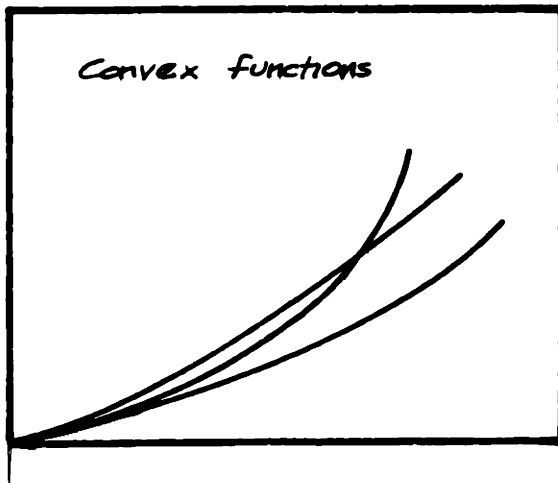
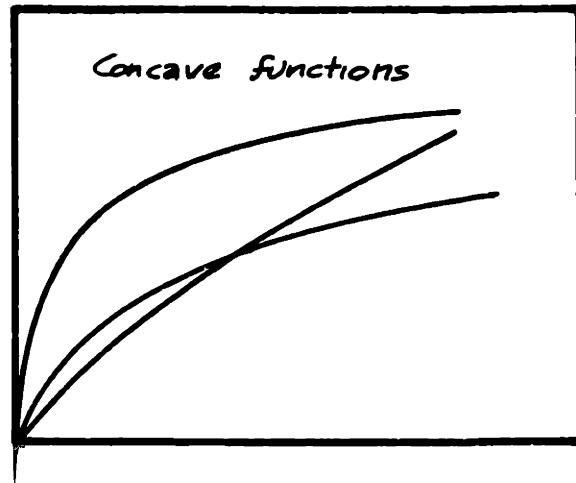
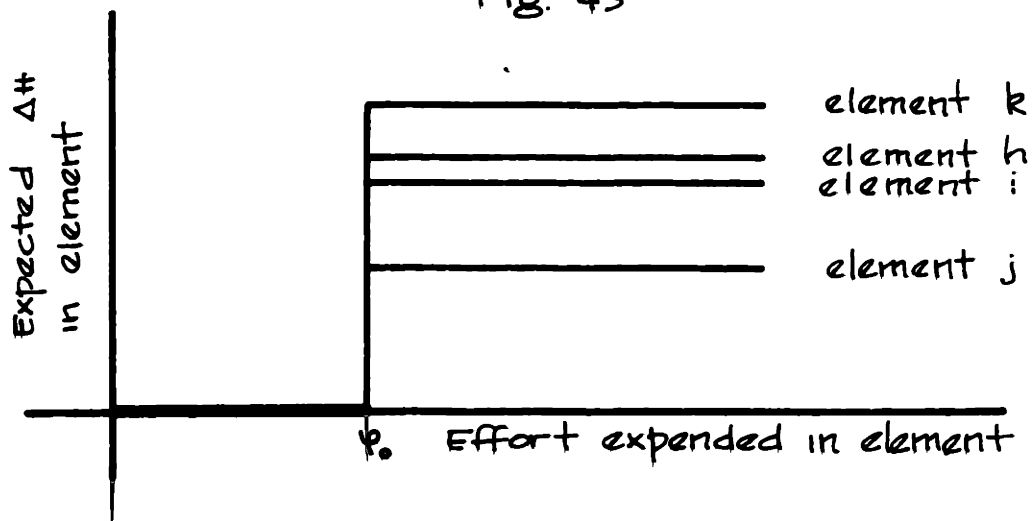


Fig. 42b



If  $I(\varphi(k))$  are step functions (Figure 43), as might be the case in boring observations, the problem becomes trivial: Allocate one unit,  $\varphi_0$ , of effort to each of the elements about which we are least certain.

Fig. 43



An illustration will show precisely how optimization might be accomplished in a real problem. Consider the problem of field mapping a large site extensively covered with overburden. In inferring the class to which an element belongs, one looks for certain evidence offering characteristics which are more commonly associated with some classes than with others (e.g., "float," vegetation, residual soils, drainage patterns, etc.) (Lahee, 1961).\*

Let the probability that a class  $i$  element exhibits characteristic  $j$  be  $q(j|i)$ . The probability of observing characteristic  $j$  in an element of class  $i$  is

$$p(j|i) = q(j|i) p\{j \text{ obs'd} | j \text{ exists}\}$$

Now, the more effort allocated to an element the greater the probability of observing a characteristic, if it exists. We will devote attention to the form of this relationship, called the "detection function," in Chapter 5, but for now we will simply assume that it has an exponential saturation form (Figure 44)

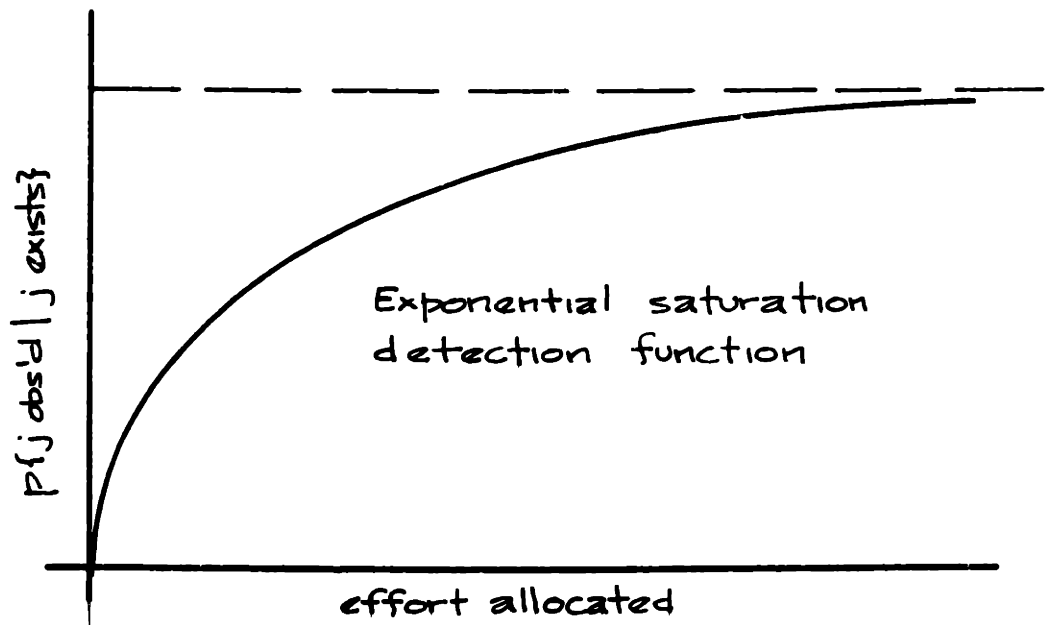
$$p\{j \text{ obs'd.} | j \text{ exists}\} = 1 - e^{-c_j \varphi}$$

where  $\varphi$  is the effort and  $c_j$  is a constant depending on the characteristic (i.e., some characteristics are hard to find, some easy).

---

\*We have before ignored the problem of observations containing information which is not sufficient to classify a location with perfect certainty because proximity rules become considerably complicated when we do not. Here we are not concerned with proximity rules, since we assume independence, and assuming perfect observations would make the problem trivial.

Fig. 44



The probability that element  $k$  is of class  $i$  when characteristic  $j$  has been observed follows from Bayes' Theorem, where  $p_i$  is the prior probability of class  $i$ ,

$$P(i|j) = \frac{p_i f(j|i)(1 - e^{-c_j \varphi})}{\sum_i p_i f(j|i)(1 - e^{-c_j \varphi})} \quad 66$$

The entropy of an element prior to exploration is,

$$H_0 = - \sum_i p_i \log p_i \quad 67$$

and the posterior entropy, given observation of  $j$  is,

$$H' = - \sum_i P(i|j) \log P(i|j) \quad 68$$

The conditional change in entropy, therefore, is

$$(\Delta H)_j = \sum_i (P(i|j) \log (P(i|j)) - p_i \log p_i) \quad 69$$



The expected change in entropy equals equals the weighted sum of the conditional changes over the possible outcomes of the exploration, or,

$$E(\Delta H) = \sum_j \sum_i p_j (p(i|j) \log p(i|j) - p_i \log p_i) \quad 70$$

Where,

$$p_j = \sum_i p_i p(j|i) \quad 71$$

Combining equations 66, 70, and 71,

$$E(\Delta H) = \sum_j \sum_i p_i q(j|i) (1 - e^{-c_j \varphi}) \times \log \frac{p_i q(j|i) (1 - e^{-c_j \varphi})}{p_i \sum_i p_i q(j|i) (1 - e^{-c_j \varphi})} \quad 72$$

$$= \sum_j \sum_i p_i q(j|i) (1 - e^{-c_j \varphi}) \log \frac{q(j|i)}{\sum_i p_i q(j|i)} \quad 73$$

The second derivative of equation 73 is everywhere negative, which means that 73 is concave. So, the optimal allocation problem is one of maximizing the sum of several concave functions subject to the linear constraint that the total effort be constant (equation 65). This is simply a problem in nonlinear programming.

#### Dependent Elements:

Computational difficulty enters when non-independent elements are considered, because proximity and trend rules become important.

A one-dimensional condition is shown in Figure 45 which may be expanded to higher dimensions. Let  $x_k$  be the location of the center

of element  $k$ . If element  $k$  is observed (assuming perfect classification) to be class  $j$ , the probability vector at any element  $h$  (by the proximity rule and Bayes' Theorem)\* would be (equation 24).

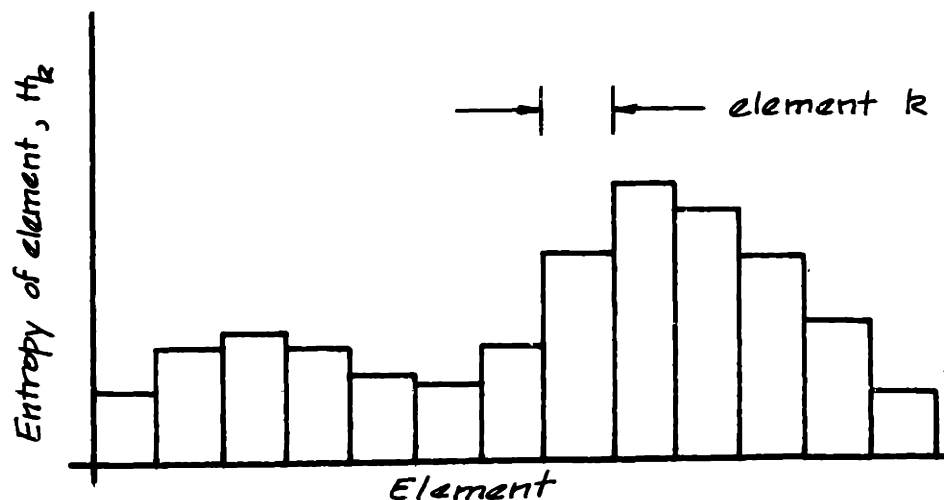


Fig. 45

$$\{P'_i\}_{x_k} = \frac{P_i P_{j|i}(1x_k - x_h)}{\sum_i P_i P_{j|i}(1x_k - x_h)} \quad 74$$

Similarly, were the elements  $x_k$ ,  $k = 1, \dots, m$ , observed to be, respectively,  $J_k$ , then (equation 25)

$$\{P'_i\}_{x_h} = \frac{P_i \prod_{k=1}^n P_{J_k|i}(1x_k - x_h)}{\sum_i P_i \prod_{k=1}^n P_{J_k|i}(1x_k - x_h)} \quad 75$$

\* In one dimension trend considerations do not enter the analysis.

An optimal allocation is one that maximizes expected change in entropy, or correspondingly, minimizes the expected posterior entropy. The expected posterior entropy of an element equals the sum of the conditional posterior entropies, given a possible outcome of observations, weighted by the probability of those outcomes. The sum is taken over all possible outcomes, or,

$$E[H_h] = \sum_{J_1} \cdots \sum_{J_m} \left[ \sum_i p_i' \log p_i' \right] p(J_1, \dots, J_m) \quad 76$$

where in this case  $p_i'$  refers to the probability of class  $i$  in the element being considered,  $h$ . The random variables  $J_1, \dots, J_m$  are the possible classifications (i.e., possible outcomes of elements  $k = 1, \dots, m$ ) upon which the  $p_i'$  depend, and  $p(J_1, \dots, J_m)$  is the probability mass function (pmf). Substituting equation 75 into 76 and averaging over all elements

$$E[H'] = \frac{1}{m} \sum_{h=1}^m \left\{ \sum_{J_1} \cdots \sum_{J_m} \left[ \sum_i \frac{p_i \prod_{J_k} p_{J_k|i}(x_k - x_h)}{\sum_i p_i \prod_{J_k} p_{J_k|i}(x_k - x_h)} \log \frac{p_i \prod_{J_k} p_{J_k|i}(x_k - x_h)}{\sum_i p_i \prod_{J_k} p_{J_k|i}(x_k - x_h)} \right] p(J_1, \dots, J_m) \right\} \quad 77$$

The optimal allocation is that set of  $n$  locations  $x_k$  (where  $n =$  total number of observations =  $\Phi$ ) which minimizes equation 77.

We shall not attempt a solution to equation 77, but if we do assume that sophisticated optimization techniques can be used to minimize it, we are still left with the problem of evaluating the pmf  $p(J_1, \dots, J_m)$ . This is no easy task in itself because by the

very fact that we use a proximity rule we assume the  $J_k$  to be autocorrelated, and therefore the marginal pmf's (i.e., the prior probability vector of each element) cannot simply be multiplied together to get the joint pmf. The conditional pmf's, however, are available from the trend rule itself (i.e.,  $p_{j|i}(d)$ ) so direct evaluation of the joint pmf is a possibility if the number of classes and elements is small (there are  $(\# \text{ classes})^m$  possible outcomes).

The two-dimensional extension of equation 77 considers only the proximity rule in assigning values to  $\{p_i\}$ . Inclusion of a trend rule complicates the optimization even more, because, although the form of the formulation is not altered, the expression for  $p_i$  must be represented by equation 53. All of the previous comments apply here as well, and again, the difficulty is solely computational.

### Sequential Strategies

When the engineer allocates sequentially, his choice at any stage can depend on all observations made in previous stages, so intuitively, at least, we might think that sequential allocations would return more information for a given amount of effort than single-stage allocation (sequential procedures are not always better, however, as we shall demonstrate in Chapter 5). The sequential performance of experiments aimed at lowering some uncertainty or risk function has been dealt with extensively in the statistics literature, but with the exception of a few special cases analytical solutions do not exist. (Quite fortunately for us,

sequential search decisions, treated in Chapter 5, is one of these special cases.)

Since no analytical solution exists for this class of problem, it is usually solved by enumerating many or all of the possible series of outcomes, and selecting the sequence of experiments which is optimum. In Bayesian decision analysis complete enumeration is called "preposterior analysis."

The sequential allocation problem is a so-called Markovian decision process in that the expected change in the state of the process (i.e., expected change in entropy) depends only on the current state and the observation chosen. Markovian decision processes are usually optimized by dynamic programming which is a backward induction technique that reduces the number of branches of the decision process requiring enumeration (see, Bellman, 1957). Although in theory the optimal sequential procedure can be constructed this way, practical application is difficult. As the number of states increases, as in site exploration, the number of branches requiring enumeration increases exponentially. Furthermore, optimal sequential procedures are not at all straightforward, or intuitive. Even in seemingly simple problems the optimal procedure may be quite complex (e.g., see Bradt and Karlen, 1956).

Only when elements are independent can we find the optimal sequential procedure. However, due to the very condition of independence the problem becomes trivial. Since information on one element does not change the uncertainty at other elements the

optimal sequential strategy is essentially the same as the optimal single-stage allocation.

#### Comment

The use of Bayesian decision theory in formulating strategies of effort allocation for reconstruction problems is hindered by general ignorance of the decision space (and consequently of a utility function) of design decisions. Therefore, a criterion of optimality other than maximizing expected utility is needed. We have proposed the direct measure of uncertainty by Shannon's entropy function. This approach is less preferable than a decision theoretic one, but allows optimization in ignorance of the specific uses to which reconstruction data will be put. If, in any specific case, the decision space and utility function are known, Bayesian decision theory should be used.

Further work may overcome the computational difficulties of optimizing single-stage allocations, but for the foreseeable future we will be able to optimize sequential strategies only through dynamic programming (the cost of which increases exponentially as the number of stages and process states increase). A fruitful topic of future inquiry might be the development of sub-optimal approximations for single-stage allocations.

Approximations to the optimizing solution (for other than non-perfect classifying observations) which do not account for interaction of entropy changes (i.e., the expected change in entropy

from allocating 2 observations is not equal to the sum of the expected changes of each one taken independently) beg the real question because such approximations would only lead to optimal strategies which prescribe observations at the centers of the most uncertain elements of the site -- a conclusion we could have made without analysis.

Finally, while we cannot at present determine allocation optima, we can compare specific alternate plans. These comparisons require only an estimate of the joint pmf  $p(J_1, \dots, J_m)$  and can be readily made for systematic cases.

### Section 5.3 Continuous Attributes (Regionalized Variables)

In addition to the discrete attributes one normally thinks of when considering geotechnical maps (i.e., those discussed in the first part of this chapter), there exist attributes which are defined on a continuous space -- ones which are not constrained to belong to discrete classes. Two such attributes are index properties and ground water elevation. The mapping of such attributes is often a required part of exploration (see, Knill and Jones, 1965; or Fookes, 1969).

As in the discrete case, treatment of continuous attributes involves two related considerations, rules of inference for predicting values at unobserved locations and criteria of optimality for allocating exploration effort. We may treat both of these by considering a mathematical model of the continuous attribute which

relates a probability density function to each point in space. Fundamentally, such models are based on statistical regression. They fit a surface to observed data and associate an uncertainty and a residual error with that surface. The surface is an estimate of the expected value of the attribute at any point, and has an uncertainty associated with it because it is only an estimate; the residual error is a measure of the dispersion of the actual value of the attribute about its expectation (Figure 46). So, we can infer values of the attribute by making predictions with the model, and we can optimize allocations of exploration effort by determining expected effects on the parameters of the model.

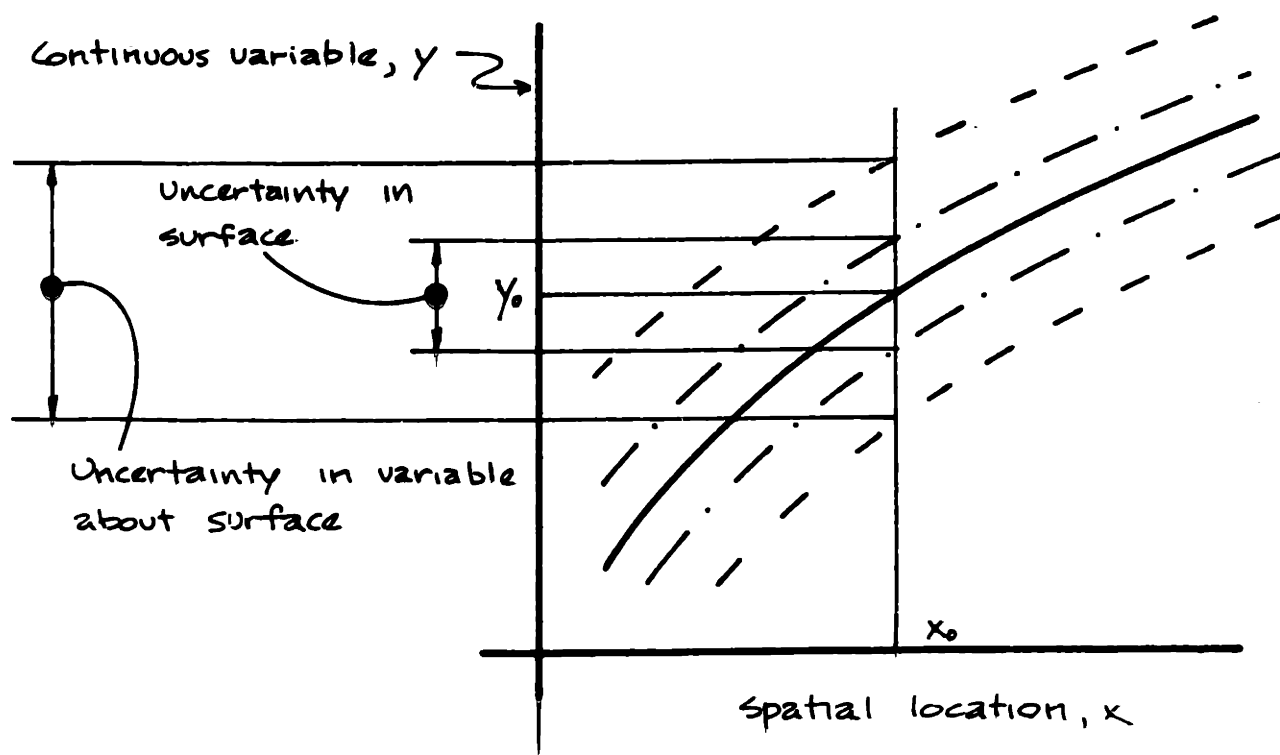


Fig. 46



The predominant use for the mapping of continuous attributes (or "regional variables" as they are called by Matheron) is in measuring and reporting mechanical properties. As such, its consideration would more properly be included in work on allocations for mechanical testing. However, since topics of mechanical testing are not considered in this thesis, and since mapping regionalized variables is related to exploring geometries of site geology, we present in this section a brief discussion of the techniques and difficulties involved.

Trend surface analyses, as these techniques are known in the English literature, have received considerable attention over the past decade. In fact, one observes that the annual number of articles published on trend surface analysis seems to be increasing exponentially, to the point that these methods now command major portions of textbooks on statistical methods in geology (e.g., Krumbain and Graybill, 1965; Harbaugh and Merriam, 1968). Harbaugh and Merriam have summarized the English literature, and Ramonava (1970) has presented an extensive bibliography which also includes work in the Russian literature.

To the author's knowledge, all of this work on trend analysis adopts a relative frequentist approach to probability, and in general fits low order polynomials (quadratic through quintic) or Fourier series to observed data. Considerable attention is usually given to the statistical significance of successive terms in the models fitted, and the most common goals are descriptions of

paleogeology and developing functional relationships between parameters of geological processes.

Most of the work in the English literature assumes local variations in regionalized variables to be independent. That is, the residual error from the trend surface at any one location is assumed not to be correlated with the residual error at adjoining locations. In reality, of course, these residuals are autocorrelated.

Matheron, and his French school of "geostatistics," have developed a method of analysis, called "universal Kriging," which accounts for autocorrelation between residuals, and is based on the analysis of stochastic time series initiated by Kolmogorov (1941) and Wiener (1949). Unfortunately, Matheron's work has appeared almost uniquely in French, although attempts are now underway to introduce it into the English literature (Huijbregts and Matheron, 1970; Watson, 1971).

Recently, attempts have been made to introduce statistical methods for treating regionalized variables into the literature of soil mechanics (e.g., Cornell, 1971), but this work has been mostly oriented toward probabilistic design rather than exploration.

### Section 5.3.1 Inference (Prediction)

Trend surface analysis is based on the concept that the observed points,  $z(x_i)$ ,  $i = 1, \dots, n$ , are points on a surface  $z = z(x)$ , where  $x$  is spatial location, and  $z(x)$  is a random function. The

random function  $z(x)$  is usually considered to be the sum of two other random functions,

$$z(x) = t(x) + e(x) \quad 78$$

where  $t(x)$  is slowly varying with distance and  $e(x)$  is rapidly varying:  $t(x)$  might be thought of as regional variation of  $x(x)$ , and  $e(x)$  might be thought of as local variation. The expected value of a regional variable, when  $e(s)$  is assumed to have zero mean (i.e.,  $E(e(x)) = 0$ ), is,

$$E(z(x)) = E(t(x)) + E(e(x)) = E(t(x))$$

If the  $e(x)$ 's are assumed mutually independent, trend analysis (in terms of predicting  $z(x_0)$  at unobserved  $x_0$  and predicting functions of  $z(x)$ ) simply reduces to finding the function  $t(x)$  which "best" fits observations. This is the model generally treated in the English literature.

Matheron's model is somewhat different. He assumes that  $z(x)$  is composed of two components: a deterministic function  $d(x)$  (called "drift") which describes the expected value of  $z(x)$  as a function of  $x$ , and a zero-mean autocorrelated random function  $u(x)$  which describes local variations from  $d(x)$ ,\*

$$z(x) = d(x) + u(x) \quad 79$$

---

\*The difference between the drift and the trend,  $t(x)$ , (of trend analysis) is that the former is a deterministic function while the latter is a random function.

Most work in the area of stochastic processes, such as Matheron's adopts the simplifying assumption of stationarity, which implies that a probability distribution is not dependent upon location. Since  $E(z(x))$  does depend on location, however, the function  $z(x)$  is non-stationary. If we could transform  $z(x)$  into a stationary function we could simplify our analysis. If we assume that  $u(x)$  is isoskedastic (i.e., the variance of  $u(x)$  is the same everywhere) then we can treat it as stationary. We can make predictions about  $z(x)$  by subtracting the drift, treating  $u(x)$  as a zero mean autocorrelated stationary process, and adding  $d(x)$  back in after the analysis is completed.

The value of  $u(x)$  at any location can be predicted by a linear combination of the observed values at nearby locations,

$$u(x_0) = \sum_{i=1}^n a_i u(x_i) \quad 80$$

where the coefficients  $a_i$  depend on the autocorrelation of  $u(x)$ , and are chosen such that the variance of the prediction about the true value is a minimum,

$$\text{variance} = E \left[ \left( u(x_0) - \sum_{i=1}^n a_i u(x_i) \right)^2 \right] = \text{minimum.} \quad 81$$

Let the observed values of  $u(x_i)$  be,

$$u^t = \{ u(x_1), \dots, u(x_n) \} \quad 82$$

and let,

$$\underline{a}^t = \{ a_1, \dots, a_n \} \quad 83$$

Then equation 81 becomes,

$$E[(u(x_0) - \sum_i^n a_i u(x_i))^2] \equiv E[(u_0 - \underline{a}^t \underline{u})^2] \quad 84$$

By the linear nature of the expectation operator (see, e.g., Davenport, 1970)

$$E[(u_0 - \underline{a}^t \underline{u})^2] = E[u_0^2] - 2E[u_0 \underline{a}^t \underline{u}] + E[\underline{a}^t \underline{u} \underline{u}^t \underline{a}] \quad 85$$

The covariance of two variables is defined as

$$\text{cov}[u(x_i), u(x_j)] \equiv c(u_i, u_j) = E[u(x_i)u(x_j)] \quad 86$$

and is related to the correlation coefficient,  $\rho_{ij}$ , of  $u(x_i)$  and  $u(x_j)$  by the relationship

$$c(u_i, u_j) = \frac{\rho_{ij}}{(\text{var } u_i \text{ var } u_j)^{1/2}} \quad 87$$

Let

$$\underline{V} = \begin{bmatrix} c(u_1, u_1) & c(u_1, u_2) & \dots & c(u_1, u_n) \\ c(u_2, u_1) & & & \vdots \\ \vdots & & & \\ c(u_n, u_1) & \dots & & c(u_n, u_n) \end{bmatrix}$$

and

$$\underline{c} = \begin{bmatrix} c(u_0, u_1) \\ c(u_0, u_2) \\ \vdots \\ c(u_0, u_n) \end{bmatrix} \quad 88$$

Substituting into equation 85,

$$E[(u_0 - \underline{a}^t \underline{u})^2] = E[u_0^2] - 2 \underline{a}^t \underline{c} + \underline{a}^t \underline{V} \underline{a}$$

which can be minimized by setting the derivative with respect to  $\underline{a}$  equal to zero. This results in the solution,

$$\underline{a} = \underline{V}^{-1} \underline{c}$$

which is equivalent to the familiar solution (Cox and Miller, 1965; Agterberg, 1970; Watson, 1971),

$$\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{21} & & & \\ \vdots & & & \vdots \\ \rho_{n1} & \cdots & & \rho_{nn} \end{bmatrix} \begin{bmatrix} \rho_{o1} \\ \rho_{o2} \\ \vdots \\ \rho_{on} \end{bmatrix}$$

where  $\rho_{oi}$  is the correlation coefficient of  $u(x_o)$  and  $u(x_i)$ .

The variance of the estimate  $\underline{a}^t \underline{u}$  about the true value  $u_o$  is just equation 85, or, since  $E(u_o^2) = 0$

$$\text{Var}(u_o - \underline{a}^t \underline{u}) = \underline{a}^t \underline{V} \underline{a} - 2 \underline{a}^t \underline{c}$$

89

To summarize, the procedure for estimating the value of a regional variable at an unobserved location is the following:

1. Remove the deterministic drift of the regional variable,  $d(x)$ , by fitting a low-order mathematical surface to the data. Then subtract the so-determined surface from the gross data to result in a set of residuals  $u(x)$  whose expected value is zero.
2. From this set of residuals, empirically estimate the autocorrelation function  $R(i,j)$  which describes the correlation between two terms  $u(x_i)$  and  $u(x_j)$  as a function of the distance separating them,  $(|x_i - x_j|)$ .
3. Use  $R(i,j)$  to determine  $\rho_{ij}$  for each pair of observations.
4. Solve for the coefficient vector  $\underline{a}$ .

5. Estimate  $u(x_0)$  and the variance in the estimate.
6. Add the drift component at  $x$ , determined in step 1, to  $u(x_0)$  to obtain an estimate of  $z(x_0)$ .

The drift function in certain instances (we will return to this point) can be assumed to have the form (Whittle, 1963)

$$d(x) = \sum_j b_j g_j(x) \quad 90$$

in which  $g_j(x)$  are functions of spatial location and  $b_j$  are scalar constants which usually must be estimated from the observed data. As in trend-surface analysis, the functions  $g_j(x)$  are often taken to be of the form  $g_j(x) = x^j$ , or of terms in a Fourier expansion.

If we assume the  $u(x)$  to be jointly normally distributed, the estimate of the coefficients  $b_j$  with minimum variance is also the maximum likelihood estimate (i.e., the most probable in a Bayesian sense) (Watson, 1971). Let there be  $n$  observations,  $z_i$ ,  $i=1, \dots, n$ , each taken respectively at location  $x_i$ , and let these sets of data be represented by the vectors,

$$\underline{z} = (z_1, \dots, z_n) \quad 91$$

$$\underline{x} = (x_1, \dots, x_n) \quad 92$$

From the multivariate normal distribution the likelihood of  $\underline{z}$ , given a coefficient vector  $\underline{b}$ ,

$$\underline{b} = (b_1, \dots, b_r) \quad 93$$

is

$$L(\underline{z}|\underline{b}) \propto \exp\left(-\frac{1}{2}(\underline{z} - \underline{b}^t \underline{x}) \underline{V}^{-1} (\underline{z} - \underline{b}^t \underline{x})\right) \quad 94$$

which is maximized by minimizing

$$(\underline{z} - \underline{b}^t \underline{x}) \underline{V}^{-1} (\underline{z} - \underline{b}^t \underline{x}) \quad 95$$

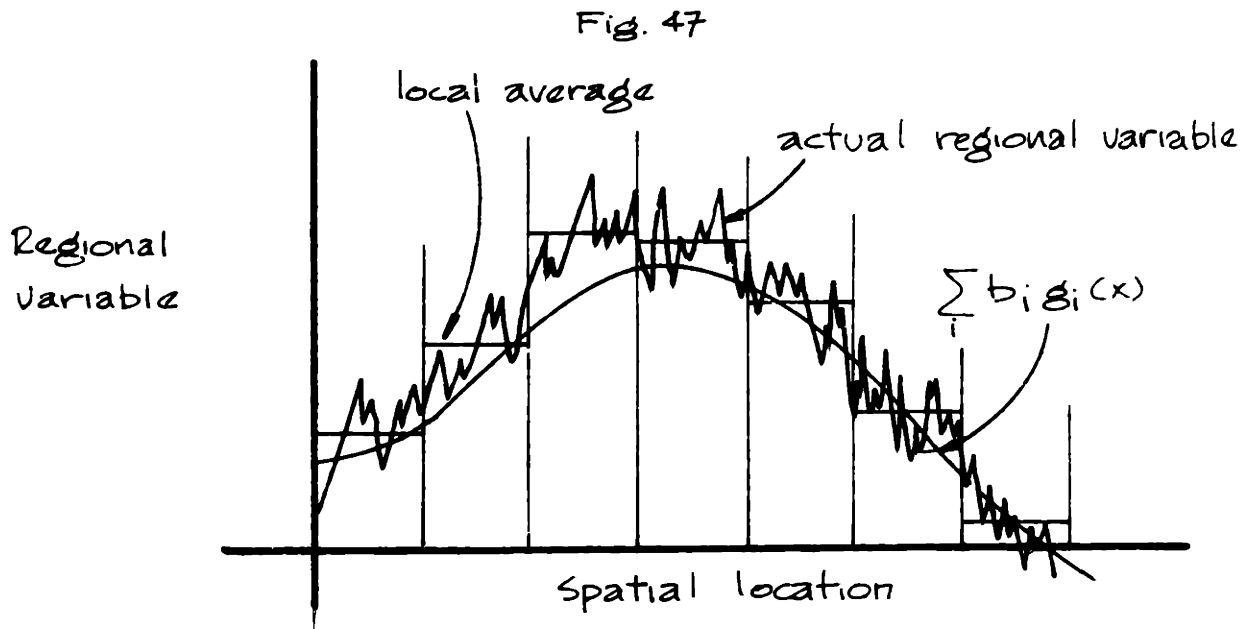
where  $\underline{V}$  is the covariance matrix of equation 87. When the autocorrelation of residuals is zero (i.e., when the residuals are independently normally distributed),  $\underline{V} = \sigma_u^2 \underline{I}$  and the minimization of 95 becomes a least squares criterion. When the autocorrelation is not zero, and  $\underline{V}$  is known a priori, the vector  $\underline{b}$  which minimizes 95 is (Whittle, 1963),

$$\underline{b} = (\underline{x} \underline{V}^{-1} \underline{x})^{-1} \underline{x} \underline{V}^{-1} \underline{z} \quad 96$$

Usually, however,  $\underline{V}$  must be estimated from the observations themselves and can be done so only after the drift is fitted. So, an approximation to the drift function must be used. One approximation is made by using the method of least squares (i.e., assuming the residuals to be independent); this does not have the property of minimum variance, but it is an unbiased fit. Agterberg (1966) reports that, empirically, the least squares fit is "very close" to the minimum variance fit (i.e., from equation 100) when the number of data points is greater than 100. For data sets of less than 100 one might attempt an iterative solution, although the iterative series is not necessarily convergent (C.A. Cornell, personal communication, 1972).. Zellner (1971) discusses other methods for handling autocorrelated residuals.



As Whittle (1963) and vanMarke (1972, personal communication) point out, a model for  $d(x)$  of the form 90 is only a useful tool in prediction if the drift of  $z(x)$  is relatively smooth, or if some professional reason exists for preferring such a model. In other cases the variance of a prediction can be decreased by approximating the drift with a local average or trend (Figure 47), because this decreases the variance of  $u(x)$ .



### Section 4.3.2 Allocating Sampling Effort

The optimal allocation of exploration effort to mapping regionalized variables depends on the use which is to be made of the resulting information. For example, in a related problem of current interest (S. Davis, personal communication, 1972), an attempt is being made to estimate the heavy metal content of top-soil in an urban park as a base level with which to compare subsequent changes

caused by air pollution. This is done by measuring concentrations at several points, fitting a drift function, and integrating over the area of interest. In this case an optimal allocation would be one which minimized the expected error of the estimate of total metal content. In design for structures in rock, the average of a regionalized variable within limited sectors of the site is often of primary interest (e.g., the average rock quality within a proposed abutment). Pragmatically, sampling for this value can be adequately handled by classical theory (summarized in Section 6.A), the application of which, in regional variable problems, has been extensively investigated by Matern (1960). Briefly Matern's conclusions are that for regional variables with isotropic autocorrelation, systematic sampling with approximately equal sampling intervals in all directions (e.g., square grids in two dimensions) gives the least variance in the estimate  $\bar{z} = \frac{1}{n} \sum_{i=1}^n z(x_i)$  (i.e., greatest precision) of all common sampling plans. When one includes some measure of sampling cost in proportion to travel distance between points, systematic plans with slight anisotropy (e.g., rectangular grids) give the highest ratio of precision to cost (the optimum aspect ratio being a function of unit travelling cost and the distance over which properties are correlated) -- these findings are essentially the same as Cochran's (1946) and Yates' (1948). Cochran (1963) discusses the precision of an estimate of  $\bar{z}$  from one systematic sample by making assumptions about the form of the drift function; this work is summarized in Section

A Bayesian approach would be an alternate, but not necessarily better, way to treat optimization. In this approach one would first fit the drift by obtaining a pdf of the model coefficients,  $f_{\underline{m}}(\underline{m}_0)$ . Treating the coefficients as random variables, a pdf of  $\bar{z}$  can be derived by integrating the drift and dividing by spatial area or volume. Although Bayesian analysis does result in a pdf of  $\bar{z}$  which can be interfaced with design decisions, classical sampling is easier to use in this case and its theory is well developed.

Although sampling for estimates of the autocorrelation function is a nontrivial task, it is precisely akin to sampling for any other statistic. When large numbers of samples are taken the central limit theorem implies that fluctuations of the sample correlation coefficient about the real value are normally distributed with mean zero and variance proportional to  $1/n$  (where  $n$  is the number of samples). Because there is fluctuation in the sample estimates we usually hypothesize a model for the autocorrelation and estimate its parameters. For example, in geological processes an exponential autocorrelation function is frequently hypothesized,

$$\rho_d = k e^{-ad} \quad 97$$

where  $\rho_d$  is the correlation coefficient for values of  $u(x_i)$  separated by distance  $d$ , and  $k$  and  $a$  are constants.

In order to estimate the autocorrelation function at points closer than the spacing of systematically placed observations additional sampling is necessary. Due to the quasi-exponential nature

of the autocorrelation function (i.e., a plot of correlation vs. log distance is usually approximately linear) data from close observations markedly increases the precision of estimates of the autocorrelation model coefficients. In sampling, Baker (1972, personal communication) has suggested using a systematic plan in conjunction with several observations randomly placed within one sampling interval (i.e., within one grid spacing) of any one observation.

The precision of estimates of model coefficients can be determined from conventional or Bayesian regression by plotting the data on a transformed coordinate system, in which the fitted model plots as a straight line (e.g.,  $ke^{-ad}$  is linear with relation to log distance).

There is debate in the literature, however, concerning the appropriate form of the autocorrelation function. The exponential function 97, or its anisotropic analogue\*

$$\rho_{d_x d_y} \propto \exp(-\sqrt{a_1^2 d_x^2 + a_2^2 d_y^2}) \quad 98$$

is in agreement with empirical data for many geological processes, yet in continuous cases a satisfactory general random process model

\*Van Marke (1970) has suggested using the model

$$= e^{-a_1(d_x)^2 - a_2(d_y)^2}$$

due to its mathematical tractability.

(particularly for the two-dimensional case) has not been constructed which results in exponential autocorrelation. For additional treatment of this point one should refer to Whittle (1963) or Bartlett (1966) who summarize some of the issues involved, or Agterberg (1970) who discusses the current literature on autocorrelation functions in geology. For our purposes an empirically determined function is certainly appropriate as long as one appreciates that the model is only a convenient summary of past observations and has no divine validity in itself.

If the drift is fitted assuming no autocorrelation between residuals, the empirically determined autocorrelation function will be biased (Agterberg, 1970). This means that the expected value of the estimate does not equal the true autocorrelation. Biasedness does not necessarily mean that the estimate is not a "good" one, since unbiasedness is only one of several criteria which can be used to evaluate estimators; but we will not venture into this point since it is treated in many statistics textbooks (e.g., Benjamin and Cornell, 1970).

### Section 5.3.3 Summary

Continuous attributes of site geology are primarily associated with mechanical properties. Problems of inferring their value at unobserved locations and of optimizing exploration effort allocations can be treated by assuming a mathematical model which considers the variable to lie on a surface,  $z(x)$ . A commonly used model divides

the function  $z(x)$  into two components: a deterministic drift  $d(x)$  which approximates  $E(z(x))$  and a zero-mean autocorrelated random function  $e(x)$  which describes local variations about the mean (i.e., about  $d(x)$ ).

At present, problems of inference and effort allocation can be treated by relative frequentist methods already developed. These methods are easy to use and result in confidence limits on predictions, and minimum variance (of estimations) sampling plans. Work is needed, however, on simultaneous treatment of multiple regionalized variables (e.g., sampling for estimates of strength and permeability).

The development of Bayesian methods which could be interfaced with a comprehensive decision-theoretic approach for regionalized variables would be desirable, and should be part of investigations into the exploration of mechanical properties.

## CHAPTER 6

SEARCH THEORY6.0 Introduction

An engineer, faced with designing a structure founded on limestone, is concerned about the possible existence of solution cavities. He has a general idea of their geometry, and also of the size cavity which would cause trouble at certain depths below the surface. How much effort should he expend trying to locate cavities (which, after all, might not even exist), and how should he distribute that effort over the site?

This is an example of a class of exploration problems which, in this work, will be called problems of search. The problem of search is basically to locate or detect a geological anomaly of certain description in an efficient way, subject to the initial probability distribution of its location, and in the presence of what is most aptly described as "noise." In site exploration, the problem is to maximize the probability of a find at a total exploration cost commensurate with the consequences of not finding significant geological details.

The suspicion that there exist adverse geological anomalies rests upon the engineer's knowledge of geology, as does the prior likelihood of the anomaly's location. These judgemental feelings

can be expressed as subjective probabilities (as discussed in chapter 2), and may be evaluated by the methods discussed there. In this chapter we assume that these evaluations have been already made. Throughout the chapter we will refer to the anomalies being searched for as "targets," the distribution of effort in space as "allocation," and the distribution of effort in time as "strategy." We will consider both single-stage search, in which all of the available effort is expended in one allocation, and sequential search, in which the effort is expended in several allocations after each of which the data are evaluated. We will consider throughout the discussion that the geometries and dimensions of targets are either known or assumed.

### 6.0.1 Organization of Chapter

As with pattern-reconstruction allocations, we can consider search allocations to be made either in single-stage or sequentially. The basic division of the present chapter is along these lines. In discussing each of the two classes we will first consider the optimal plan of searching with a fixed resource, and then the decision of how much total effort is justified for a particular problem.

### 6.0.2 Brief History of Search Theory

The initial major theoretical contribution to rationalizing search procedures came as a result of antisubmarine warfare efforts by the United States military during World War II. The product of



many people's work, it was first published by B. O. Koopman (1946) as a classified report to the Department of the Navy. The report was declassified in 1958, but selected portions of it had been revised and published in the professional literature earlier (Koopman, 1956a, 1956b, 1957).

Search theory was seen to have application to a host of problems, both military and civilian. Gluss (1959) applied the theory to searching for defective components in electronic circuitry; Posner (1963) applied the theory to searching for satellites lost in space; Morse (1970) applied the theory to searching for books in libraries (and to establishing policies for new-book shelving practices). DeGuenin (1961) and Dobbie (1963) extended search theory to exploration techniques more general than Koopman's while Charnes and Cooper (1958) modified the basic search problem of Koopman for solution by linear-programming techniques.

The most common optimizing criterion is maximum probability of a find, although a minor amount of work (Danskin, 1964; Mela, 1961) has been based on changes in uncertainty (in both investigations, the entropy function).

Enslow (1966) has presented an extensive annotated bibliography of search-theory articles published prior to 1965.

Interest in sequential procedures of search--procedures in which each unit of effort is allocated on the basis of all preceding allocations--began to be evidenced in the literature during the middle 1960's (Black, 1962; Chew, 1967), and were treated from an

entirely different approach than were single-stage allocations. As shall be discussed in Section 5.5, sequential search is fortunately one of the few sequential decision problems for which an analytically optimized strategy can be found. Most sequential decision problems require partial or total enumeration of possible alternatives and consequences.

There are few published contributions to using rational search procedures in engineering geology (including soil mechanics), and essentially all of them are in the mining literature (e.g., Slichter, 1955; Drew, 1966; Brown, 1960; Engle, 1957). At this time (1972) the use of simple analytical search techniques (e.g., grid search-- Section 5.2.3) is only starting to creep into civil engineering practice.

### 6.1 Single Stage Search

Single-stage search refers to the commitment of the total available search effort in one stage, the outcome of any one quantum of effort not being reviewed until the entire effort has been expended. This is a common case in site exploration. A program of exploration is constructed, a contract let to perform the exploration, and the results submitted to the engineer in the form of a report.

We shall consider several allocation schemes for the search effort, starting with the simplest (analytically) and proceeding through allocations which are in some sense optimal. We commence with random search, discuss the use of grid patterns, and end with

allocations which maximize the probability of finding the target when the prior pdf of the target's location is not uniform. While random search is not a technique frequently used in practice, the discussion of random search will serve as a foundation for the remaining sections.

### 6.1.1 Random Search

Random allocation of search effort is not an efficient technique for search, as we shall see, but it is easy to treat analytically. Random search refers to the procedure which allocates each quantum of effort independent of the allocation of all other quanta. Concisely, the necessary and sufficient conditions for random search are

1. Each small element of area within the site has an equal probability of being search by a quantum of effort;
2. Each quantum of effort has an equal probability of being located at any elemental area within the site.
3. The selection of the location of any one quantum of effort in no way influences the selection of the location of any other quantum.

The probability of finding a target using random search follows the binomial law since each quantum of effort is independent and has the same probability of success. Let the probability of success of any one quantum of effort (e.g., a boring) be  $p$ . The probability of failure (i.e., not detecting the target) on any one quantum is, therefore,  $(1-p)$ . The probability of not detecting the target with

n quanta is, by the binomial theorem

$$\text{Pr} \left\{ \begin{array}{l} \text{no detection} \\ \text{w/ } n \text{ quanta} \end{array} \right\} = \binom{n}{n} (1-p)^n p^0 \quad 1$$

$$= (1-p)^n \quad 2$$

The probability of detection is the complement of (2), or

$$\text{Pr} \left\{ \begin{array}{l} \text{detection} \\ \text{w/ } n \text{ quanta} \end{array} \right\} = 1 - (1-p)^n \quad 3$$

For  $p \leq 0.1$ , this is approximately

$$\text{Pr} \left\{ \text{detection w/ } n \text{ quanta} \right\} \cong 1 - e^{-np} \quad 4$$

#### EXAMPLE:

A target whose planar area is 100 sq. m. is being searched for in a site whose area is 10,000 sq. m. What is the probability of detection as a function of the number of borings randomly located in searching for the target?

SOLUTION:  $p = 10^2/10^4 = 0.01$

$$\text{Pr} \left\{ \text{detection with } n \text{ borings} \right\} = 1 - e^{-n(0.01)}$$

(see Figure 1)

For low values of n (Figure 1) the probability of detection is approximately linear with n. The increasing deviation from linearity as n increases reflects the growing probability that more than one boring will strike the target as n increases.

For small quantities of exploration effort or for low probabilities of success on one quantum, random search plans offer approximately

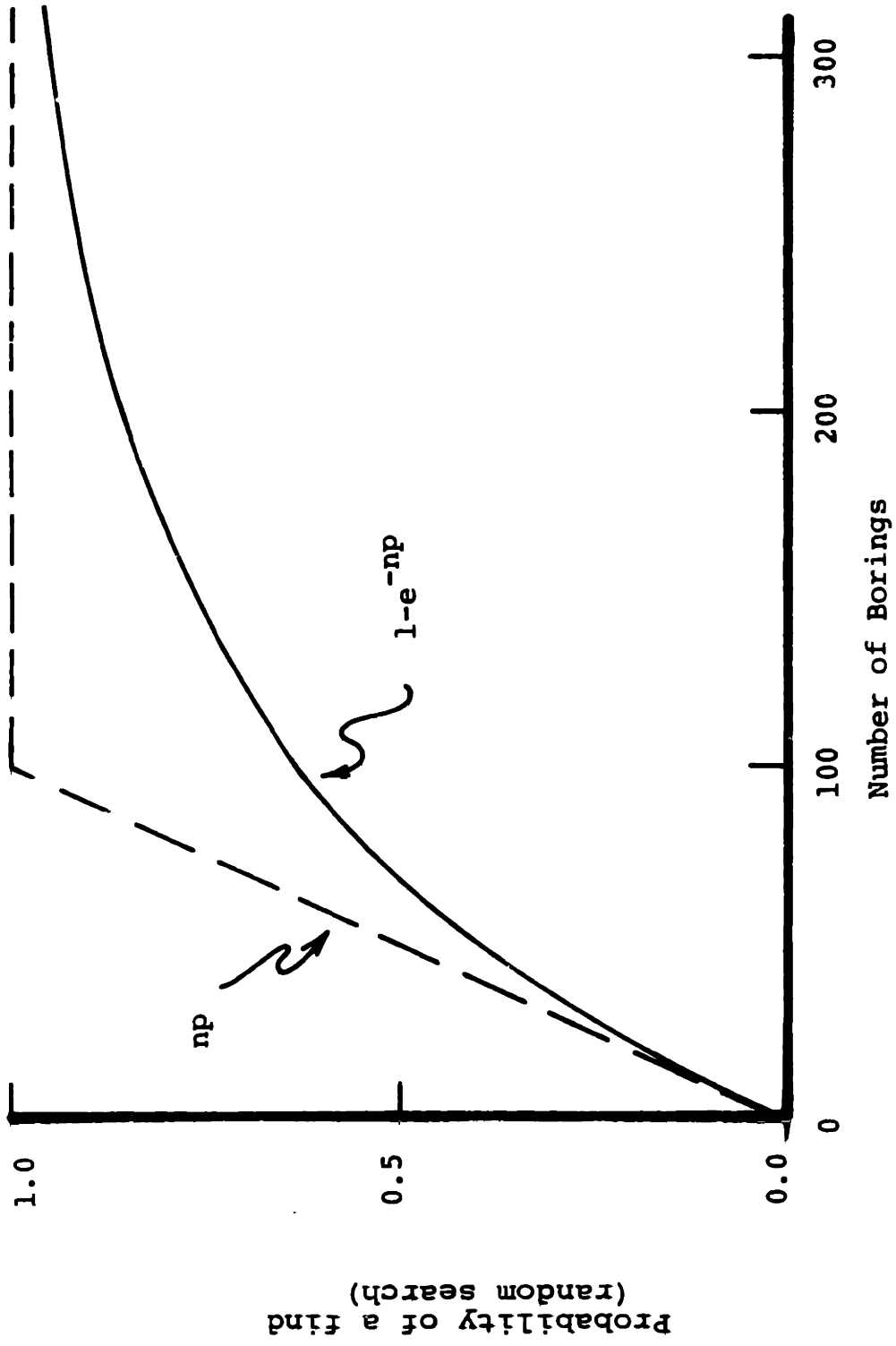


Figure 1

the same probability of a find as do systematic plans; however, as the overall probability of a find rises, the probability of a find with random plans becomes smaller than that with systematic plans.

### 6.1.2 Grid Search

Locating exploration effort in a grid pattern is the most common allocation in civil engineering practice, the primary reasons being that grid allocations are easy to specify, and coverage of the site is insured. Nearly every textbook in applied soil mechanics and engineering geology discusses the appropriate grid spacings for different types of soil or rock conditions (e.g., Table 1).

TABLE 1  
(after Sowers and Sowers)

Structure or Project	Spacing	
	ft	m
Highway	1000 - 2000	300 - 600
Earth dam	100 - 200	30 - 60
Borrow pits	100 - 400	30 - 120
Multistory Buildings	50 - 100	15 - 30
Manufacturing plants (one story)	100 - 300	30 - 90

In search problems grid-pattern allocations are appropriate for targets without preferred location at the site. That is, targets whose prior pdf's of location are uniform. Grid allocations are appropriate for these targets because they distribute effort uniformly

over the site, which, as we will prove in Section 2.3, maximizes the probability of a find.

There is an infinite variety of grid and target geometries that could be considered. Instead of treating each separately and drawing up a large series of probability tables (i.e., one for each combination of grid and target), we will consider only a few simple geometries and try to extract a general procedure for evaluating the probability of a find for any grid allocation.

### Point Grids

Analyzing grid allocations for the probability of finding an object of given geometry and size is generally straightforward since grids are repeating geometries, and only one cell needs to be considered in the analysis. Consider first the one dimensional problem of finding an object whose location can be described by the single

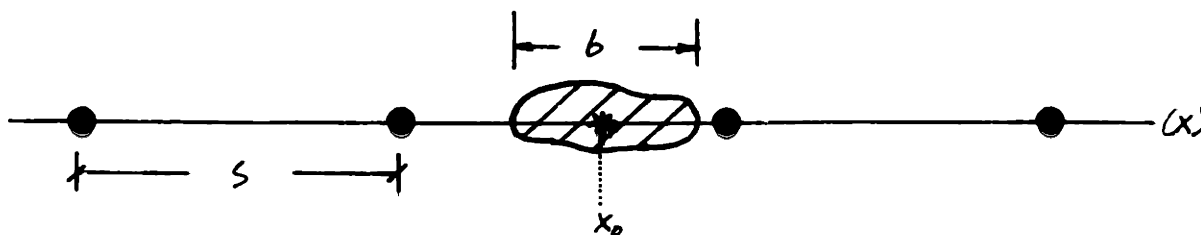


Fig. 2

variable  $x$ . If the size of the object in the  $x$ -direction is  $b$  (Fig. 2) and if the center of the object is located at  $x_i$ , then a find will be reported if and only if one point of the exploration grid falls within the interval  $(x_i - b/2, x_i + b/2)$ . Letting the grid spacing be  $s$ ,

the probability of a find is, neglecting boundary problems,\*

$$p(\text{find} | b, s) = b/s \quad \text{for } b \leq s \quad 5$$

When we expand the problem to two dimensions the probability of a find becomes dependent upon geometry as well as the size of the target (Figure 4). Consider a circular target of diameter  $b$ , whose center is randomly located in the  $(x,y)$  plane (i.e., each small element of area  $dx dy$  has an equal probability of containing the center of the circle). The exploration grid can be broken down into repeating elements of the form shown in Fig. 5, each containing a boring at its center. If the center of the target falls within a circle of diameter  $b$  about the boring, a find will be recorded. If it falls outside of this circle, no find will be recorded. Since

---

\* We have implicitly assumed the target to be continuous in setting up the problem statement. However,  $p(\text{find} | b, s)$  depends only on the length of the intersection of the target with the  $x$  axis and not its continuity. The probability of finding the target shown in Figure 3 is also  $b/s$ .

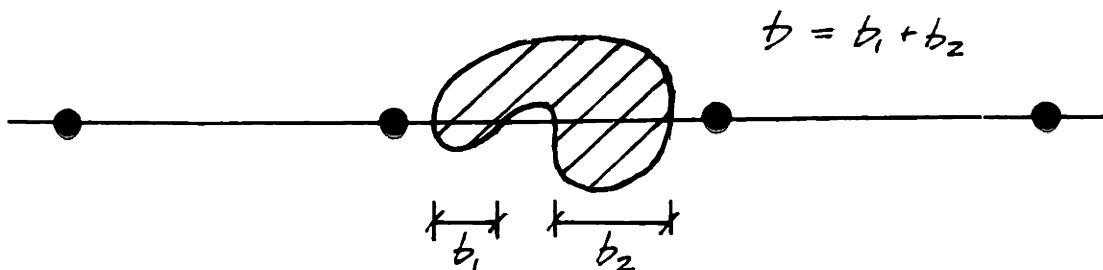


Fig. 3



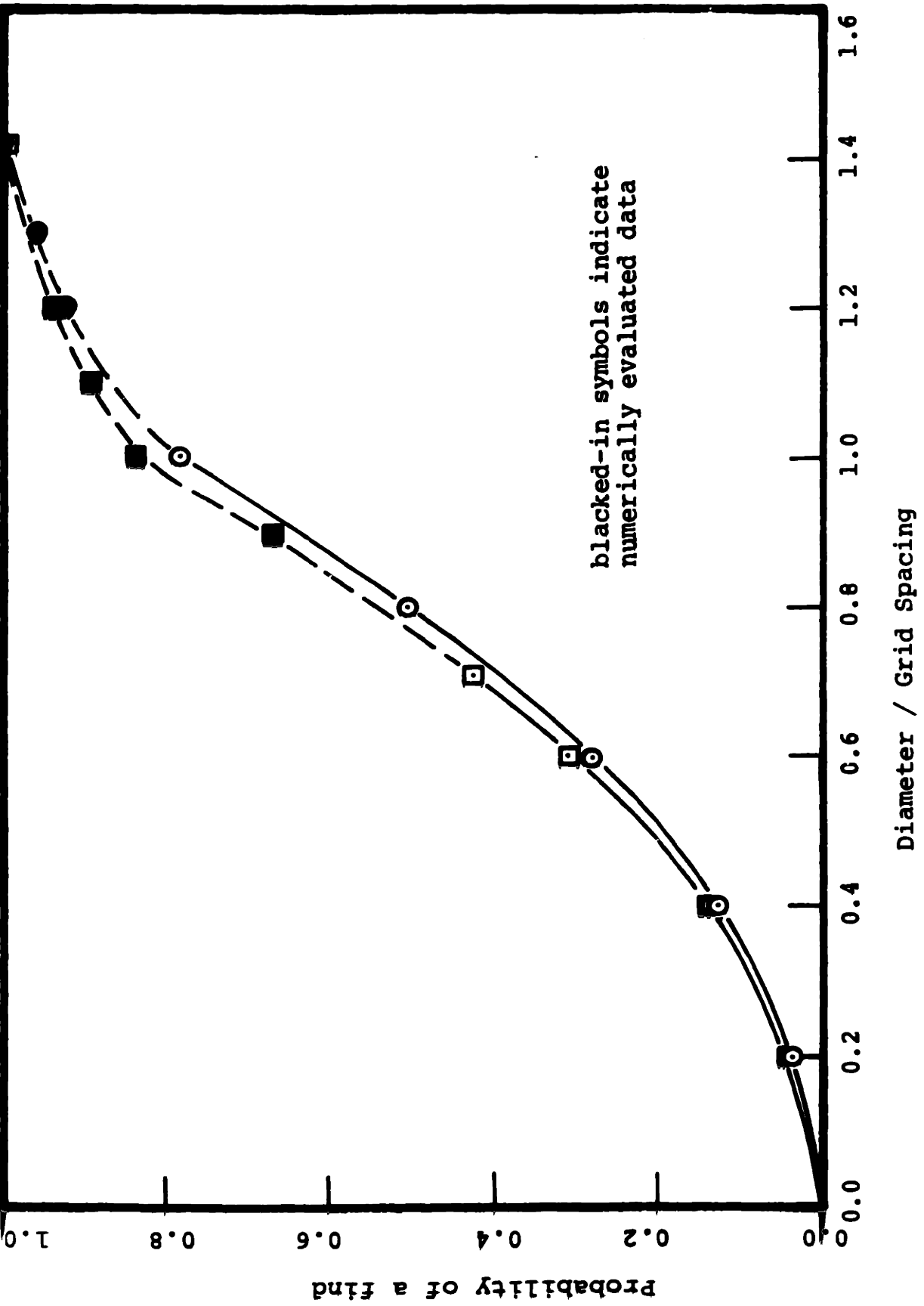


Figure 4: Probability of a find for circular and square targets

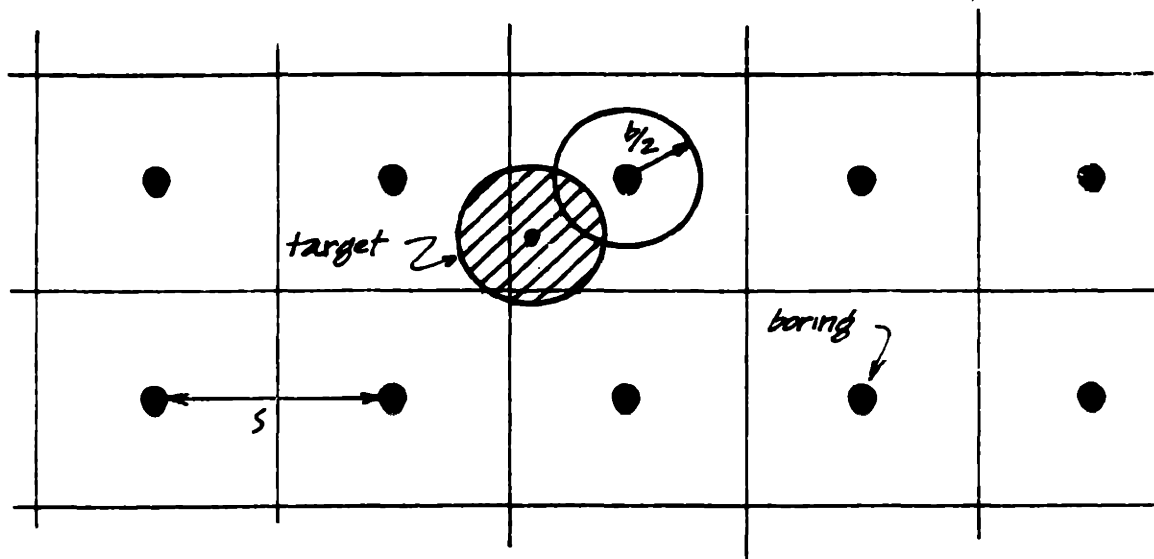


Fig. 5

Since the element of Figure 5 is repeating over the zone of search, the probability of a find is

$$p(\text{find} \mid b, s) = \frac{\text{Area } \textcircled{\bullet}}{\text{Area } \square} \quad \text{for } b \leq s$$

$$= \frac{\pi}{4} \left( \frac{b}{s} \right)^2 \quad 6$$

When  $b > s$  the circles within which a find is recorded overlap (Figure 6). This complicates the problem only in that the area within which a find is recorded in each element is difficult to specify analytically (Figure 7). The dashed portions of the probability curves in Figure 4 were obtained graphically.

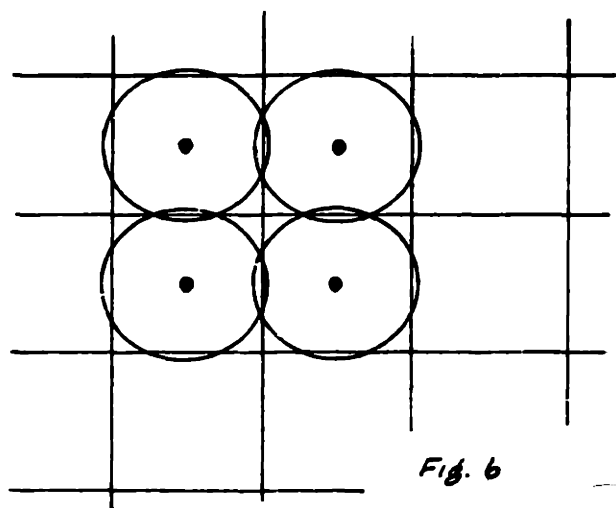


Fig. 6

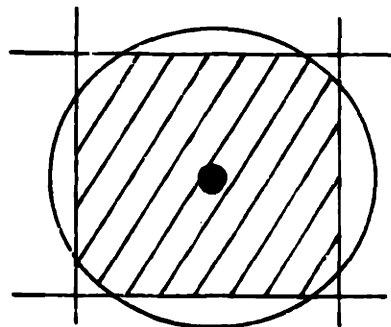


Fig. 7

We'll compare equation 6 with the probability of finding a square target with the same grid. The new determination is more involved than the circular target was because the distance between the center of the target and its perimeter varies from  $b/2$  to  $b/(2)^{1/2}$  depending on the angle  $\theta$  shown in Figures 8 and 9. The diameter of the detection circle shown in Figure 9 then also varies from  $b$  to  $(2)^{1/2}b$ , and must be treated as a random variable.

Fig. 8

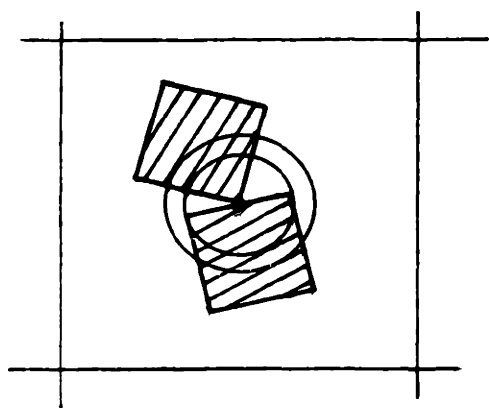
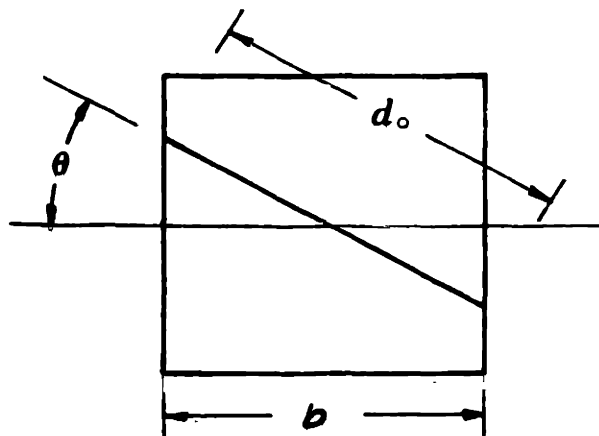


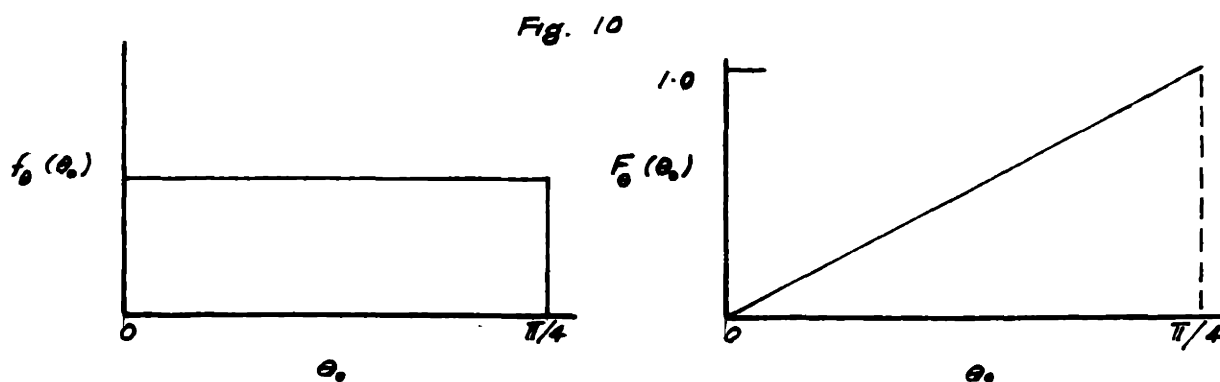
Fig. 9



To determine the probability of a find we must integrate the product of the conditional probability given  $d$  and the pdf of  $k$ , over all possible values of  $d$ .

$$\begin{aligned}
 p(\text{find}) &= \int p(\text{find}|d) f_d(d) dd \\
 &= \int \frac{\pi}{4} \frac{d^2}{s^2} f_d(d) dd \\
 &= \frac{\pi}{4s^2} \int d^2 f_d(d) dd \\
 &= \frac{\pi}{4s^2} E(d^2) \quad \text{for } b \leq s/\sqrt{2} \quad ?
 \end{aligned}$$

Since the diameter  $d$  is a function of  $\theta$ , the pdf of  $d$  can be determined from the pdf of  $\theta$ . With no prior information on  $\theta$  its pdf will be assumed uniform (Figure 10), and its cumulative



distribution function (cdf) linear, since

$$F_{\theta}(\theta_0) = \int f_{\theta}(\theta) d\theta \quad 8$$

$$= k\theta \quad 9$$

where  $k$  is a constant. For the area under the pdf to equal 1.0,  $k$  must equal  $4/\pi$ . We know that (Figure 9)

$$\theta_0 = \cos^{-1} b/d_0 \quad \text{for } b \leq d_0 \leq \sqrt{2}b \quad 10$$

So, if  $\theta \leq \theta_0$ , then  $d \leq d_0$ , or

$$\begin{aligned} F_d(d_0) &= F_{\theta}(\theta_0) \\ &= F_{\theta}(\cos^{-1} b/d_0) \\ &= 4/\pi (\cos^{-1} b/d_0); \quad b \leq d_0 \leq \sqrt{2}b \end{aligned} \quad 11$$

The pdf of  $d_0$  is determined by differentiating the cdf, or

$$\begin{aligned} f_d(d_0) &= \frac{d}{d(d_0)} F_d(d_0) \\ &= \frac{d}{d(d_0)} (4/\pi \cos^{-1} b/d_0) \\ &= \frac{4/\pi (b/d_0^2)}{\sqrt{1 - b^2/d_0^2}}; \quad b \leq d_0 \leq \sqrt{2}b \end{aligned} \quad 12$$

Figure 11 is a plot of  $f_d(d_0)$ . The expected value of  $d_0^2$  equals the second moment of  $f_d(d_0)$ , or

$$\begin{aligned}
 E[d^2] &= \int d_0^2 f_d(d_0) dd_0 \\
 &= \frac{4}{\pi} \int_b^{\sqrt{2}b} \frac{b}{1 - b^2/d_0^2} dd_0 \\
 &\approx 1.1 b^2 \quad *
 \end{aligned}$$

13

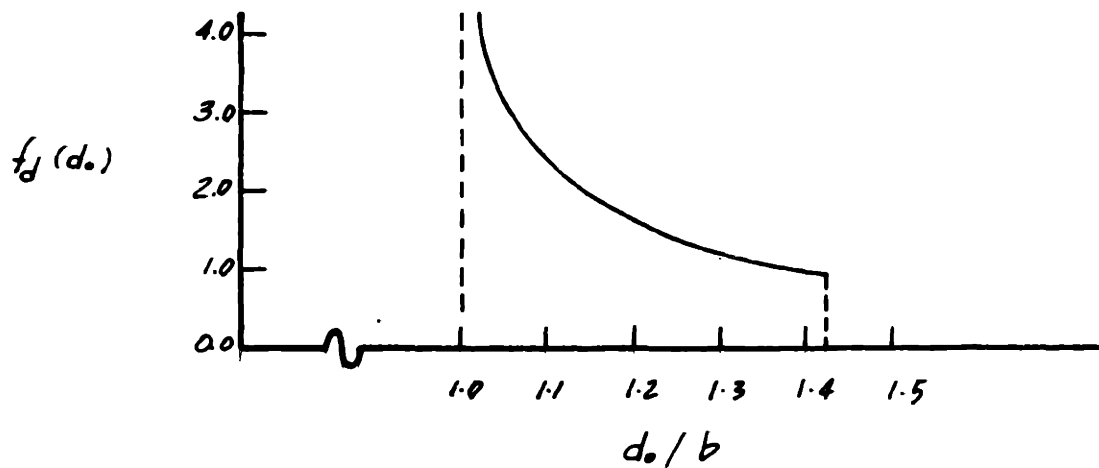


Fig. 11

Therefore, the probability of finding a square target of minimum dimension  $b$ , whose orientation is unknown, with a square grid of spacing  $s$  is

$$\begin{aligned}
 p(\text{find} \mid \square \text{ target}, b, s) &= \pi/4 (1.1 b^2)/s^2 \\
 &= 0.86 b^2/s^2 \quad b \leq s/\sqrt{2} \quad 14a
 \end{aligned}$$

Compare this to the probability of finding a circular target of diameter  $b$  with the same grid

$$p(\text{find} \mid \circ \text{ target}, b, s) = 0.785 b^2/s^2 \quad b \leq s \quad 14b$$

\*This solution has been evaluated numerically using Simpson's Rule.

The probability of finding the square target is 1.1 times the probability of finding the circular target of the same width. However, since the area of the square target is 1.275 times the area of the circular target, the probability of finding a square target is less than the probability of finding a circular target of the same area.

Other target shapes may be treated in a similar fashion once the functional relationship between  $d$  and  $\theta$  is determined (e.g., Figure 12). Slichter (1955) has presented some work on rectangular targets (Figure 21) and Savinskii (1965) has presented a comprehensive set of tables and graphs for elliptical targets and rectangular grids.

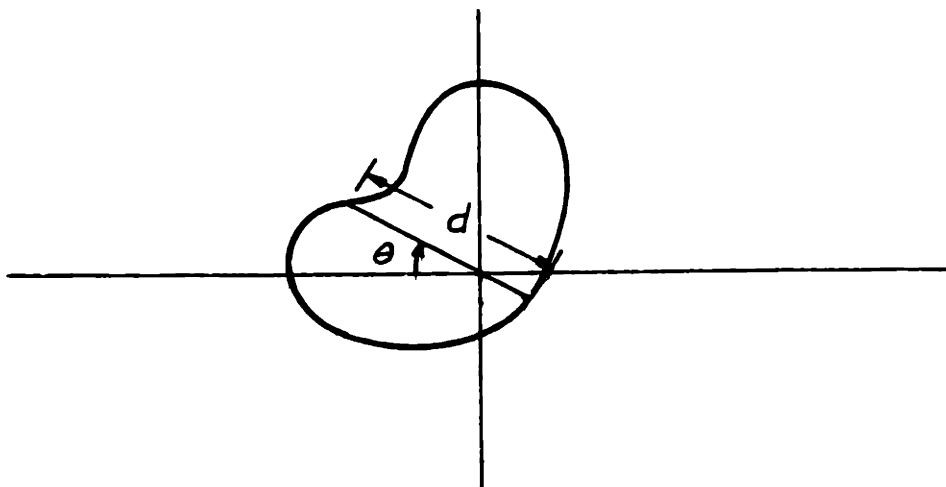


Fig. 12

Equation 14 is only valid for  $b \leq s/\sqrt{2}$  since for larger values the circle within which detection may occur overlaps the unit grid element. We do know that for  $b \geq \sqrt{2} s$ , the probability of a find, is 1.0; and furthermore, we know that the true curve must lie

below equation 14 for  $b > s/\sqrt{2}$ , since overlapping lowers  $p(\text{find})$ . Consider Figure 13. As before, if a target center lies within  $b/2$  of a boring a find is recorded with probability 1.0. Now, if a target center lies in a ring of width  $\Delta d$  at distance  $d_i$  from a boring (cross-hatched area in Figure 13) a find is recorded if and only if  $d_o \geq d_i$  in the direction of the boring. Letting the area of this ring within the elemental grid square be  $A_i$  and the area of the circle of radius  $b/2$  within the elemental grid square be  $A_o$

$$p(\text{find}) = \frac{A_o + \sum_i A_i Pr(d_o \geq d_i)}{s^2} \quad 15a$$

For  $d_i \geq s$ , as  $\Delta d \rightarrow 0$ ,

$$A_i \rightarrow 4 \left( \frac{\pi}{2} - 2 \cos^{-1} \frac{s/2}{d_i/2} \right) d_i/2 \, dd_i \quad 15b$$

and for  $d_i \leq s$ , as  $\Delta d \rightarrow 0$ ,

$$A_i \rightarrow \pi d_i \, dd_i \quad 15c$$

Since  $Pr(d_o \geq d_i) = (1 - F_d(d_i))$ ,

$$p(\text{find}) = 1/s^2 \left\{ A_o + \int_{d_i=b}^s \pi d_i (1 - F_d(d_i)) \, dd_i + \int_{d_i=s}^{\sqrt{2}b} 4 \left( \frac{\pi}{2} - 2 \cos^{-1} \frac{s}{d_i} \right) \frac{d_i}{2} (1 - F_d(d_i)) \, dd_i \right\} \quad 15d$$

For  $b \leq s/\sqrt{2}$ , equation 15d yields

$$p(\text{find}) = \pi/4 (1.11 b^2/s^2) \quad 15e$$

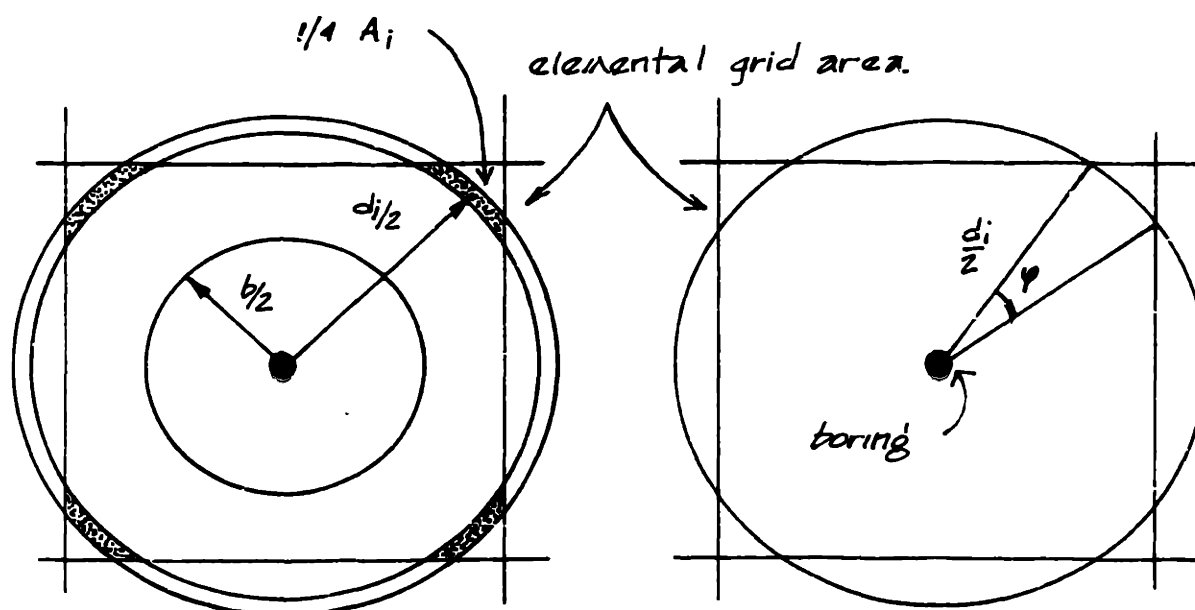


corresponding well with the numerically evaluated equation 14.

For  $b \geq \sqrt{2}$  equation 15a becomes difficult to solve analytically, and numerical methods were used to obtain the values plotted in Figure 4. Obviously, these values are approximations.

Fig. 13

Fig. 14



The general procedure for determining the probability of a find with grid allocations becomes clear. Given that targets exist at a site, they must each lie within one of the repeating units of the grid pattern. Given that the targets are randomly located, the center of each target is randomly positioned within the repeating elemental grid unit in which it lies. Within an elemental grid unit we can define two mutually exclusive and cumulatively exhaustive regions A and B with the following properties: If the center of the target falls within region A a find will be recorded; if it falls

within region B, no find will be recorded. The probability of finding any one target given that it exists somewhere at the site, equals the ratio of area A to the total elemental grid unit area  $A + B$ .

When a target is non-circular in cross-section the size of area A changes with changes in the relative orientation of target and grid. In this case the expected area  $E(A)$  is determined and used to compute the probability of a find  $P = E(A)/(A + B)$ .

When a target dimension exceeds the bounds within which simple equations describing probability of a find are valid, geometry becomes a complicating factor, and probabilities may need to be obtained either graphically or by simulation.

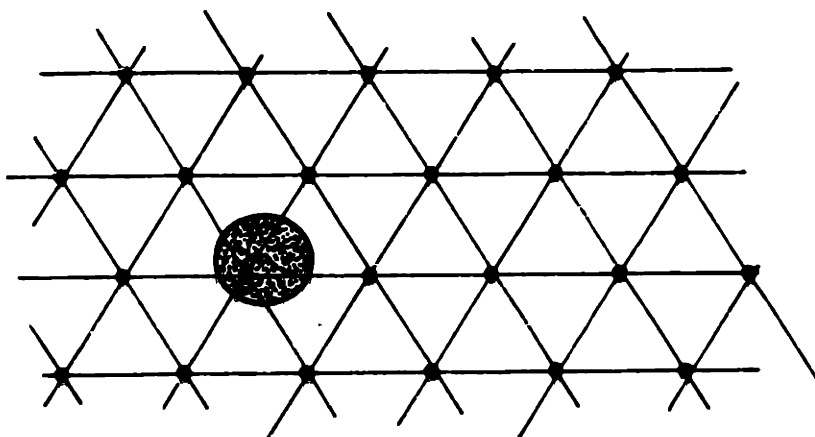
We will define a "triangular" grid as one in which any three mutually adjacent points form the vertices of an equilateral triangle (Figure 15). With a triangular grid, no point in the plane is further than 0.577s from the closest grid point, as compared with 0.707s with a square grid. Using the procedure just developed, we will compare the probability of finding a circular target using a triangular grid to that using a square grid, when the number of grid points (e.g., borings) is fixed; we shall show these probabilities to be identical. (This is not to say that in some other application triangular grids might not be superior -- only, in search problems when the target is circular and square and triangular grids give identical probabilities of finding a target.)

Consider a very large rectangular site of dimensions  $a_1$  by  $a_2$  (large enough with respect to grid spacing that we will neglect

boundary effects). If  $n$  borings were to be allocated in a triangular grid over the site the grid spacing would be

$$s_t = (\sin 60^\circ a_1 a_2 / n)^{1/2} \quad 16$$

FIG. 15



Similarly, for a square grid the grid spacing would be

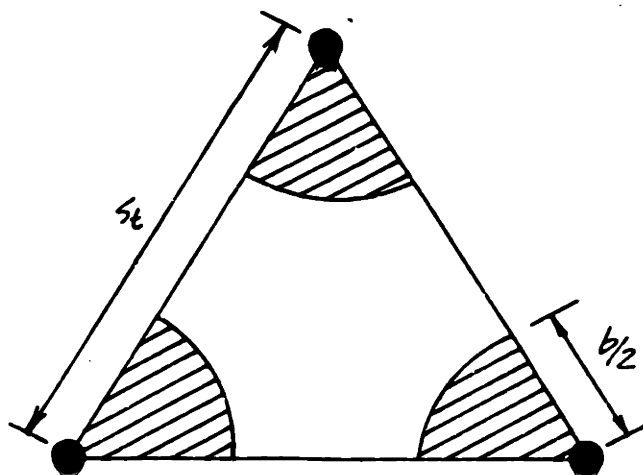
$$s_s = \sqrt{a_1 a_2 / n} \quad 17$$

$$s_s = 0.94 s_t \quad 18$$

The probability of finding a circular target with a triangular grid equals the ratio of the crosshatched area in Figure 16 to the total elemental area (Figure 17).

$$p_A (\text{find } s_t, b) = \frac{\pi/4 (b^2/s_t^2)}{\cos 30^\circ} \quad b \leq s_t \quad 19$$

Fig. 16



From p. 266 the probability of a find within the square grid is

$$p_{\square}(\text{find} | s_s, b) = (\pi/4)(b^2/s_s^2) \quad b \leq s_s \quad 20$$

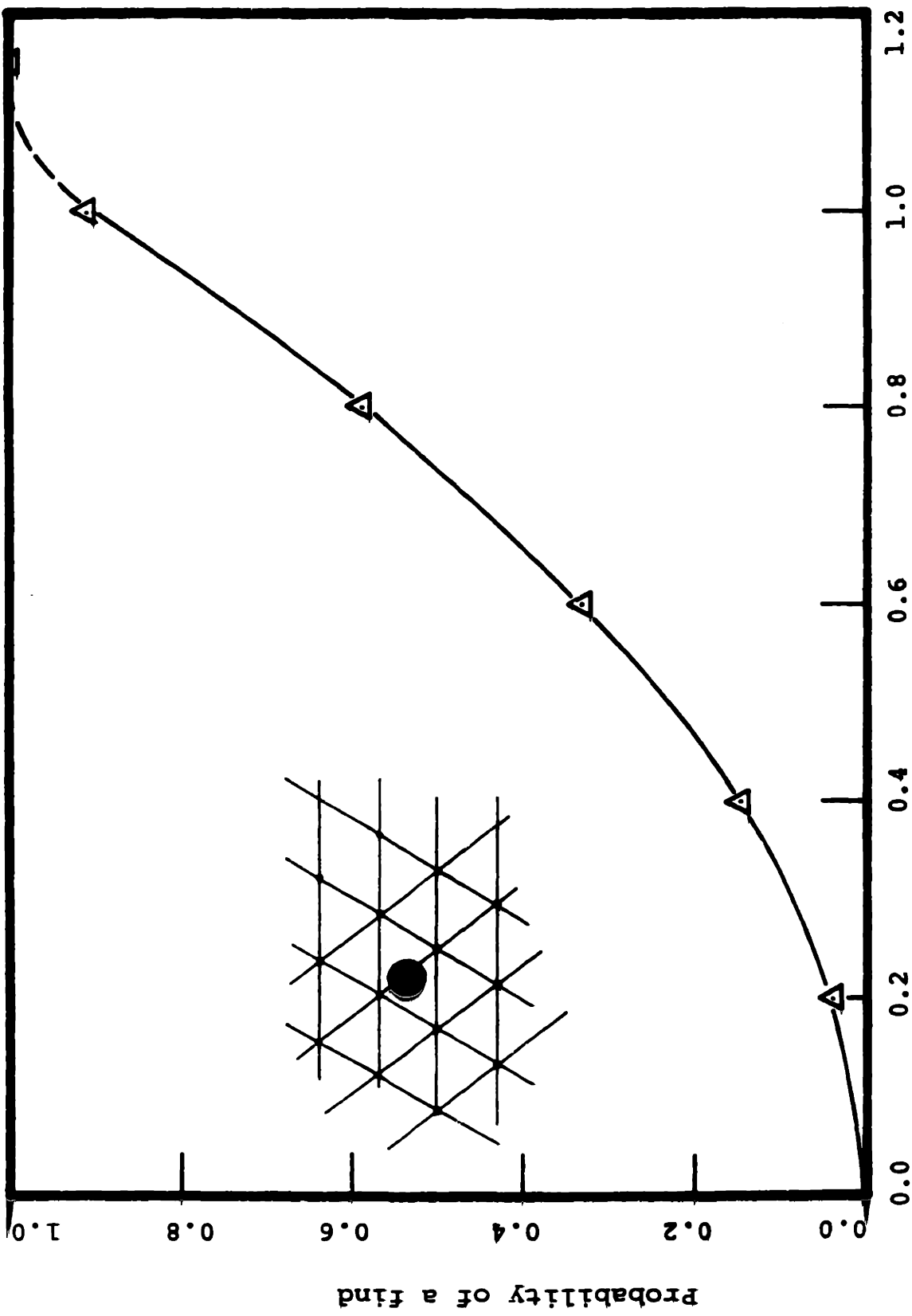
Combining equations 18 and 20

$$p_{\square}(\text{find} | s_s, b) = \left(\frac{\pi}{4}\right) \left(\frac{b^2}{\cos 30^\circ s_t^2}\right) \quad b \leq s_s \quad 21$$

Therefore, the probability of a find is the same for both a triangular and a square grid when the total number of grid points is fixed (i.e., fixed total search effort).

### Line Grids:

We have thus far considered only point grids: grids in which the effort is allocated at the intersections of grid lines. In geophysical work, field reconnaissance, and certain other search techniques the



Target Diameter/Grid Spacing

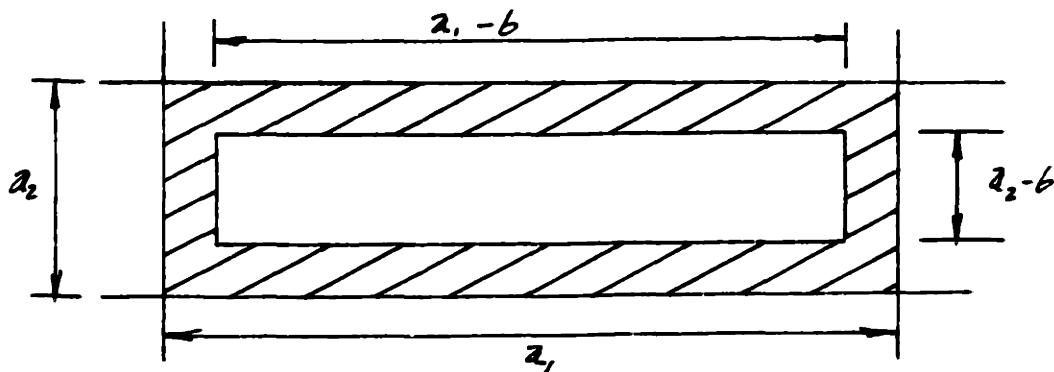
Figure 17: Probability of a find for circular targets and triangular grids

effort is allocated along the grid lines themselves. Thus, the allocation is continuous along lines rather than localized at points. The probability of a find using line grids equals the probability that a line will intersect the target. We will consider only rectangular grids and circular targets to illustrate how the probability of a find is determined.

The probability of a find with a rectangular line grid equals the probability that the center of the target falls into the cross-hatched area of the elemental unit shown in Figure 18.

$$p(\text{find} / a_1, a_2, b) = \begin{cases} 1 - \left[ \frac{(a_1 - b)(a_2 - b)}{a_1 a_2} \right] & b \leq a_2 \leq a_1 \\ 1.0 & b \geq a_2 \end{cases}$$

FIG. 18

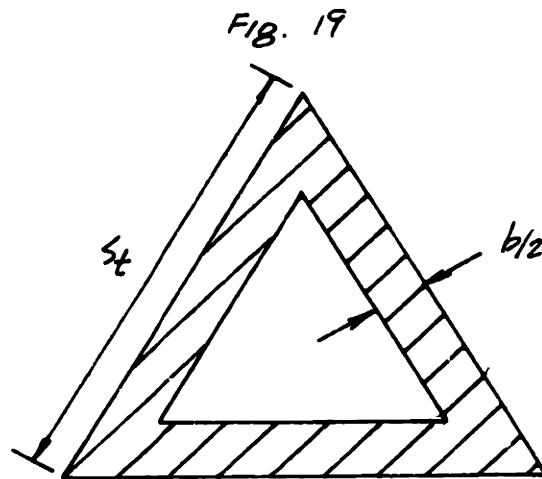


The probability of a find with a triangular grid equals the probability that the target center falls into the cross-hatched area of the elemental unit shown in Figure 19.

$$p_{\Delta}(\text{find} | b, s_t) = 1 - \frac{(s_t - b(\cos 30^\circ + \sin 30^\circ))^2}{s_t^2}$$

$$b \leq 0.5775 s_t$$

23



As before, we may compare the probability of a find using square and triangular grids. The area searched per unit of traverse (neglecting boundary effects), respectively, is .

$$A_s / L_s = 1/2 s_s \quad : \text{ square grid}$$

$$A_t / L_t = 0.29 s_t \quad : \text{ triangular grid}$$

For a given-size large site and given total length of traverse,  $A/L$  must be a constant, so  $s_t = 1.72s_s$ . Substituting into equations 22 and 23 we find

$$\frac{p(\text{find} | \square \text{ grid})}{p(\text{find} | \Delta \text{ grid})} = \frac{1 - \left(\frac{s_s - b}{s_s}\right)^2}{1 - \left(\frac{1.72s_s - b(1.65)}{1.72s_s}\right)^2} \quad 24$$

Therefore, a square line grid gives a higher probability of finding

a circular target ( $b \leq 0.575s_t \cong s_s$ ) than does a triangular line grid.

### Target Obliquity:

Up to this point we have concerned ourselves solely with targets which were approximately equidimensional (i.e., circular and square). We now turn our attention to planar targets which have one dimension larger than the other. Two shapes we will consider are the ellipse and the rectangle. Once the target ceases to be equidimensional the orientation of the target acquires importance (see the discussion of square targets in Section 2.2.1), and prior probabilities of the target's orientation (other than the uniform prior implicitly assumed in equally likely  $\theta$  on page 268) must be considered. We will let the ratio of the maximum dimension of the target,  $a$ , to the minimum dimension,  $b$ , be called the target aspect ratio  $\lambda = a/b$ . We shall first concern ourselves with uniform prior probabilities of target orientation, and evaluate the sensitivity of the probability of a find to changes in the obliquity of the target; following this we consider the use of rectangular grids in searching for oblique targets for which some orientation information exists prior to searching; and finally, we will consider the problem of laying out grids which maximize the probability of a find for a given target and prior pdf of orientation.

The graphs in Figures 20 and 21 show the relationship between probability of a find and target obliquity for elliptical and



(after Savinski, 1965)

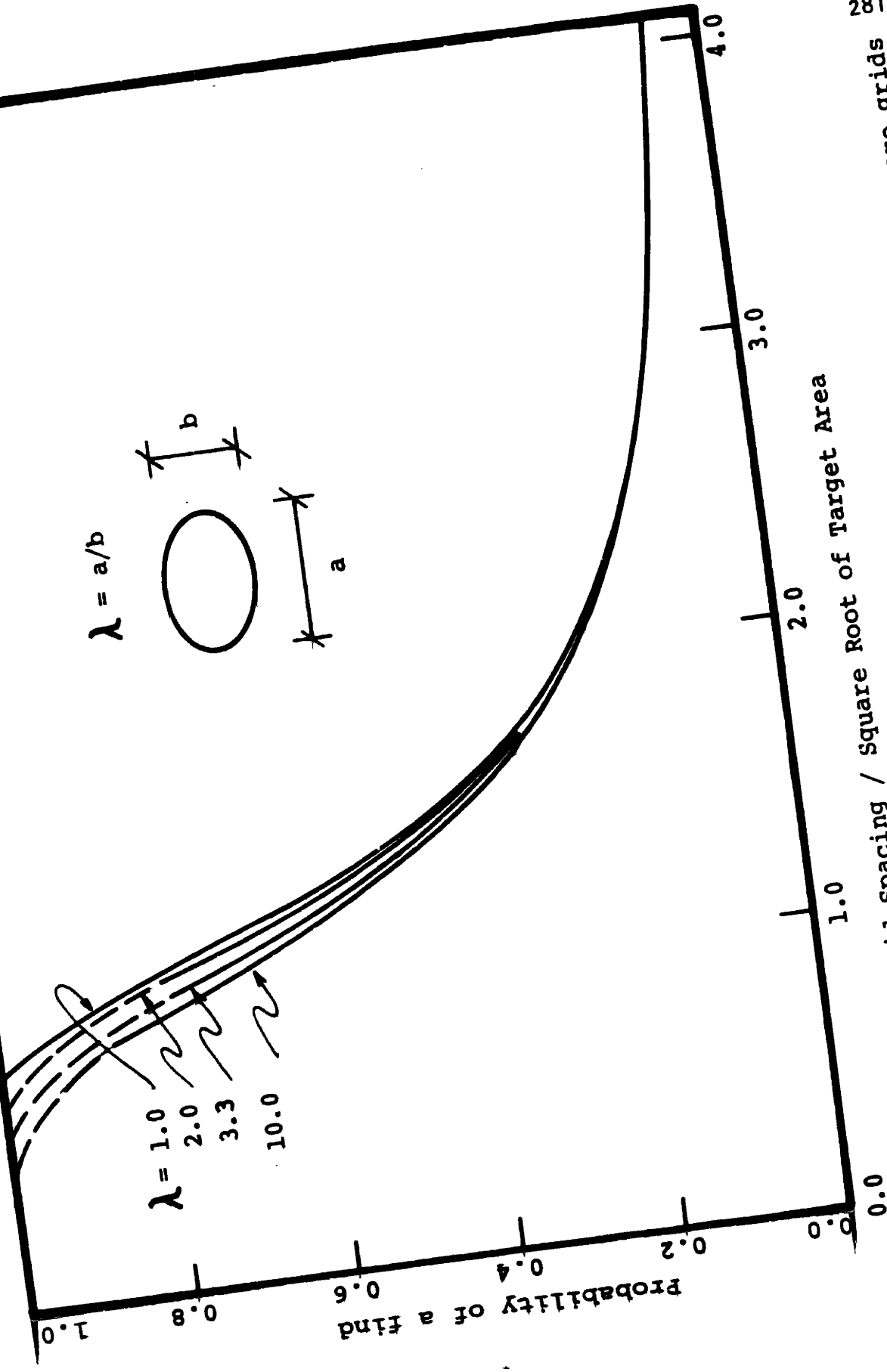


Figure 20: Probability of a find for elliptical targets, square grids

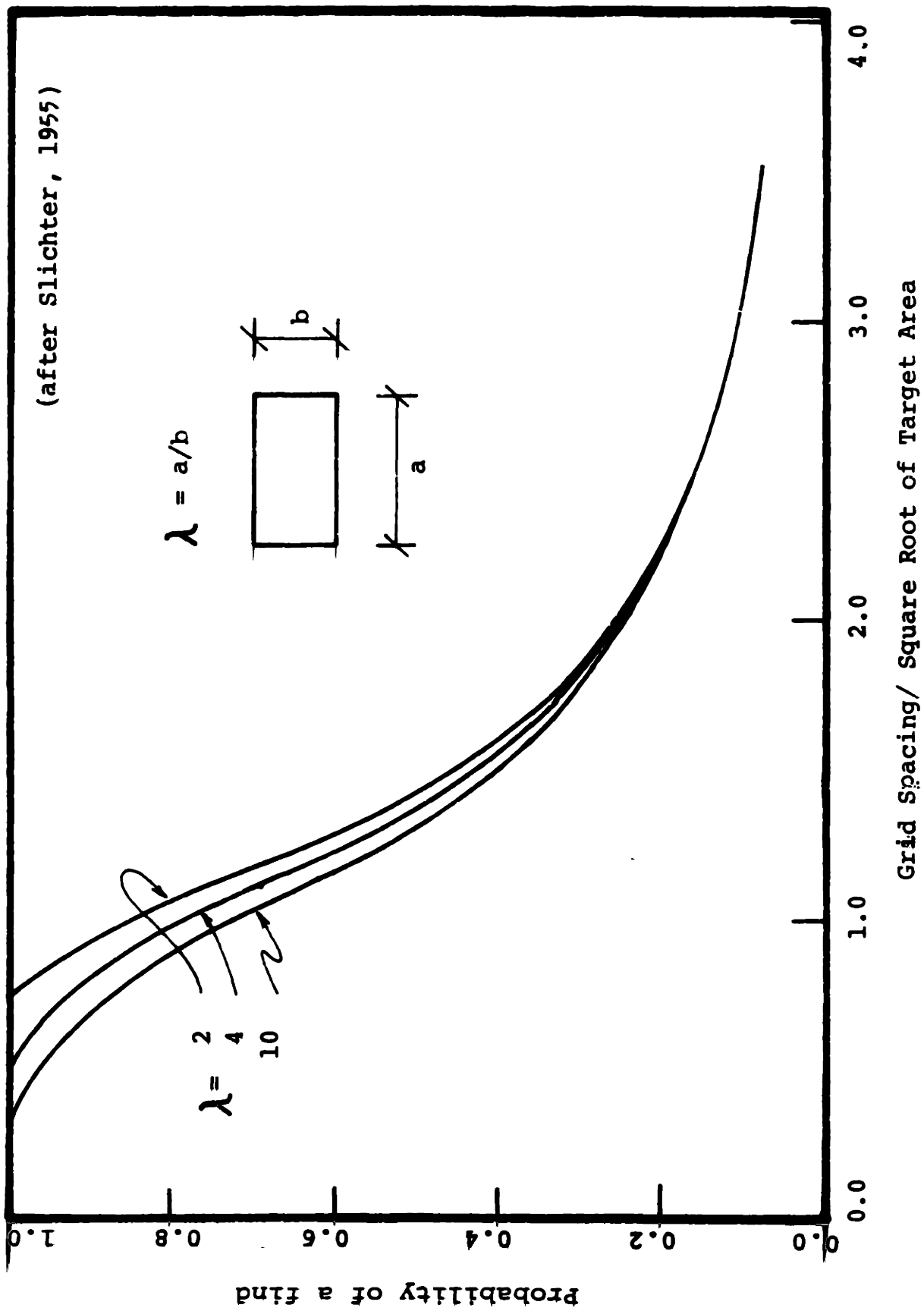


Figure 21: Probability of a find for rectangular targets, square grids

rectangular targets and uniform probability on orientation. The important thing to note here is that once the grid spacing exceeds about 1.5 times the square root of the target area the probability of a find is insensitive to changes in target obliquity. The error for  $s$  greater than 1.5 times the square root of the target area is less than 5%. This means that in most practical applications the assumption of a circular target causes little error. As the grid spacing becomes smaller, however, the actual obliquity of the target is important and must be considered. Errors of 20 to 25% can occur by neglecting target obliquity in the region of smaller spacings.

#### A Priori Preferred Orientation of Target

We turn our attention for a moment to the case in which we have some information a priori about the orientation of an anisotropic target (i.e., a target with  $\lambda > 1.0$ ). The probability of finding an elliptical target and the probability of finding a rectangular target are essentially the same when the obliquities of the two targets are the same (Figures 20 and 21), so we may restrict our attention to the elliptical case with no appreciable loss of generality.

Possessing information both about the target aspect ratio and orientation enables one to design a rectangular grid having a higher find probability than when orientation is assumed random. Consider an elliptical target of dimensions  $a$  by  $b$  and a rectangular grid,  $s_1$  by  $s_2$  (Figure 22).

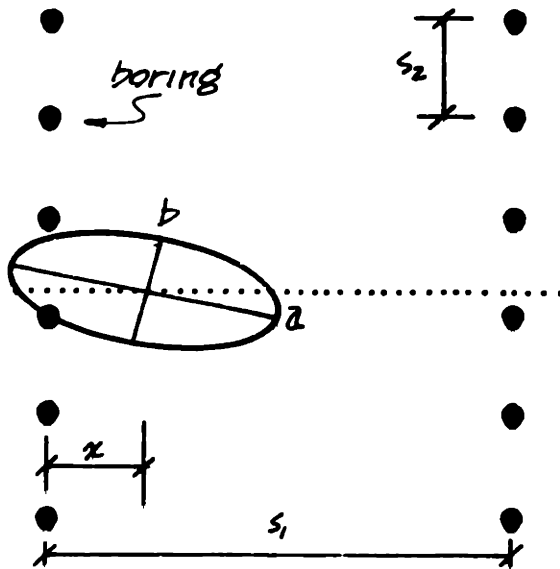


Fig. 22

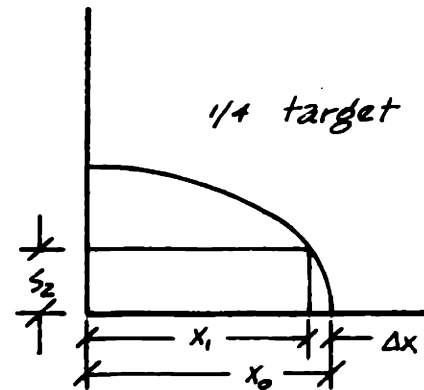


Fig. 23

Savinskii (1965) has shown that the probability of none of the points hitting an elliptical target oriented at  $\theta$  to the major axis of the grid is

$$\Pr \{ \text{no hits} | \theta \} = \frac{2}{s_1 s_2} \left\{ s_2 \Delta x - \int_{x_1}^{x_0} \mathcal{L}(x, \theta) dx \right\} + \frac{s_1 - bc}{s_1} \quad 25$$

where

$$\mathcal{L}(x, \theta) = \frac{2a}{bcz} \sqrt{\frac{b^2 c^2}{4} - x^2} \quad 26a$$

and

$$c^2 = \sin^2 \theta + (a^2/b^2) \cos^2 \theta \quad 26b$$

Starting from Savinskii's formula and accounting for the probabilistic nature of  $\theta$ , the probability of no hit occurring is

$$\Pr (\text{no hit}) = \int_{\theta} \Pr (\text{no hit} | \theta) f(\theta) d\theta \quad 27$$

where

$f(\theta)$  = the pdf of target orientation

The optimum rectangular grid obliquity  $s_1/s_2$  can be determined by finding the parametric values  $s_1^*$  and  $s_2^*$  which minimize  $\text{Pr}(\text{no hit})$ . Savinskii has done this for the cases

$$f(\theta) = \begin{cases} \text{uniform} & -30^\circ \leq \theta \leq 30^\circ \\ 0 & \text{otherwise} \end{cases}$$

and

$$f(\theta) = \begin{cases} \text{uniform} & -90^\circ \leq \theta \leq 90^\circ \end{cases}$$

Tables still need to be constructed for other distributions, particularly for uniform distributions with other limits (e.g.,  $\pm 10^\circ$ ), and circular normal distributions (e.g., see Gumbel, et al., 1953), although for the present, Savinskii's tables for uniform priors within the limits  $\pm 30^\circ$  can be used to obtain approximate solutions for the optimal grid spacings  $s_1^*$  and  $s_2^*$  when approximate alignment is known. Constructing additional tables will require considerable computation time.

Reviewing Savinskii's tables and discussions, two conclusions can be drawn. First, when the preferred orientation is known, the optimum grid obliquity is approximately proportional to the target obliquity,

$$(s_1/s_2)^* \approx \lambda$$

28

Second, the optimal orientation of the long axis of the grid is parallel to the long axis of the target.

These ideas and Savinskii's tables are not unknown in civil engineering practice, having been used by the Mexican firm DIRAC in searching for voids beneath building foundations (Cornell, 1972, personal communication).

#### Posterior Probability of Targets Existing Given No Find

The numbers and sizes of targets found with a given search effort,  $\Phi_0$ , offers evidence which may be used to refine our assessment of the total number of targets existing at a site, and the distribution of their sizes. We will first consider total number.

Let the actual number of targets at a site be  $n$  (without perfect information,  $n$  is a random variable) and the number found in a given search be  $m$ . If the prior pmf of  $n$  is  $p_n(n_0)$ , the posterior pmf may be found by Bayes' Theorem,

$$p'_n(n_0) = \frac{p_n(n_0) L(m/n_0, \Phi)}{\sum_{n_0} p_n(n_0) L(m/n_0, \Phi)} \quad 29$$

The probability that one or more targets exist somewhere at the site when no find is recorded is then

$$\Pr \left\{ \begin{array}{l} \text{one or more} \\ \text{targets exist} \end{array} \middle| \text{no find, } \Phi \right\} = 1 - \left[ \frac{p_n(0) L(0/0, \Phi)}{\sum_{n_0} p_n(n_0) L(0/n_0, \Phi)} \right] \quad 30$$

#### Example

An engineer is designing a structure on karst topography and is concerned about the existence of solution cavities. He has available

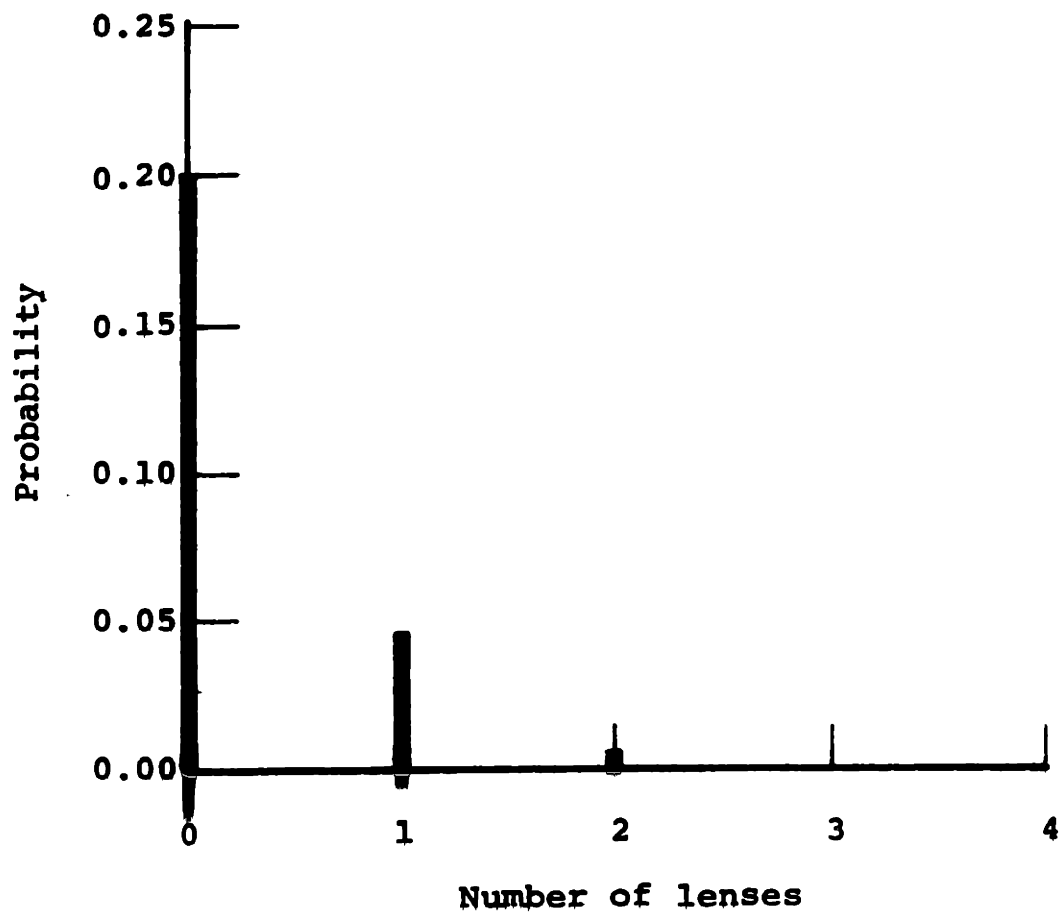
the data from a previous boring program in which borings were allocated in a square grid with 50 ft. spacings. In experience with adjacent areas the average number of voids for the area he is interested in has been found to be 0.225, and the voids occur randomly in the horizontal plane. Assuming the number of voids to be Poisson distributed, he computes the prior pmf of the number of voids to be that shown below. Given that no voids were found in the previous exploration, what is the probability that one or more exist at the site?

For simplicity we will assume that the size of voids is known to be  $b = 20'$  (a prior pdf of size could be introduced, which would lead to more involved calculations but not change the general form of the solution). Let  $L(0 | n, s)$  equal the likelihood that no voids are found using a grid spacing  $s$ , given that  $n$  voids exist. From Bayes' Theorem,

$$\begin{aligned}
 & P(\text{one or more voids exist} \mid \text{no finds}, s=50, d=20) \\
 &= 1 - \frac{p(0)L(0|0,50)}{p(0)L(0|0,50) + p(1)L(0|1,50) + p(2)L(0|2,50)} \\
 &= 1 - \frac{0.8(1.0)}{0.8(1.0) + 0.18(1 - \frac{20^2}{450^2}) + 0.02(1 - \frac{20^2}{450^2})^2} \\
 &= 0.175
 \end{aligned}$$

Therefore, the exploration data decreased the probability that a void exists undetected.

Figure 24



Prior probability of number of lenses

Poisson distributed



Example

Prof. W. G. Baker (1971, personal communication) has suggested a problem of considerable interest. In the investigation of large clay deposits, zones or lenses of high moisture content are sometimes encountered. Assuming that all these zones can be considered geometrically and dimensionally similar, what inferences can we draw about the total number of zones if our search has encountered  $m$  of them (i.e., we seek the posterior pmf of the number of zones,  $p'_n(n_0)$ ).

To avoid geometric complexities we will assume that the vertical dimension of the lenses is small compared to the deposit depth, and hence the probability of two lenses being at the same elevation and intersecting one another (i.e., having a center to center spacing less than  $b$ ) is negligible. We will further assume that systematic borings of spacing  $s$ , completely penetrating the deposit are used as a search tool (Figure 25).

If the centers of the lenses are considered randomly distributed in the horizontal plane with density  $\lambda$  (i.e.,  $\lambda$  = average number of lens centers/unit area), we may represent the prior pmf of the number of lenses in the total area by the Poisson distribution,

$$p_n(n_0) = \frac{(\lambda A)^{n_0} e^{-\lambda A}}{n_0!}$$

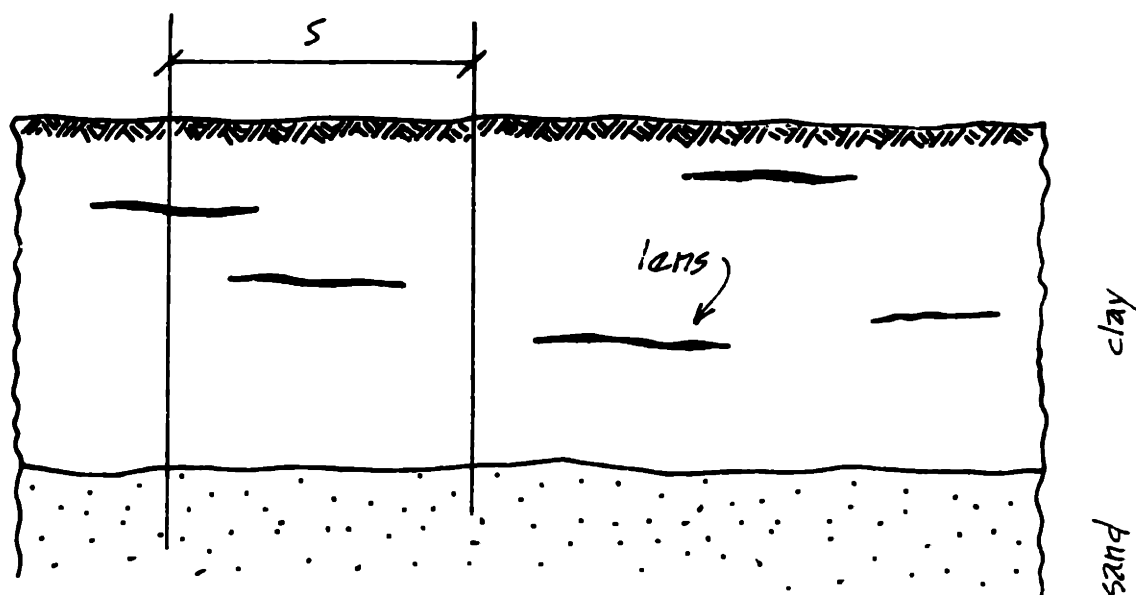


Fig. 25

The likelihood of  $m$  finds, from the binomial theorem, is

$$L(m/n, \Phi) = C_m^n (p(\text{find}))^m (1-p(\text{find}|\Phi))^{n-m} \quad 31$$

where  $C_m^n$  are the combinations of  $n$  items taken  $m$  at a time.

Consider a site area 5000' by 2000' searched by a square grid of borings 100 ft. on center. If the lenses are approximately circular with 50' diameters, and the prior pmf of the number of lenses is Poisson with  $\lambda = 10^{-5}$  (i.e.,  $10^{-5}$  lens centers are expected per square foot), then the posterior pmf when, say, 3 lenses have been found is

$$P'_n(n_0) \propto \left( \frac{10^{n_0} e^{-10}}{n_0!} \right) \frac{n_0!}{3!(n_0-3)!} (0.166)^3 (0.834)^{n_0-3}$$

Figure 26 shows the prior and posterior pmf's. In this case the exploration program shows that there are probably more voids existing at the site than originally thought.

Were the target dimension not deterministic but represented by some prior pdf, we could compute a posterior pdf of dimension in much the same way as we have just treated total number of targets. We do this by first assuming the distribution of  $b$  to belong to a family of distributions (e.g., Normal, Gamma, etc.). If  $\alpha$  and  $\beta$  are coefficients of the distribution model, we can record our a priori uncertainty about  $b$  in the form of a prior joint pdf of  $\alpha$  and  $\beta$ ,  $f_{\alpha, \beta}(\alpha_0, \beta_0)$ . The sample information can be used to modify and hence refine our estimate of the pdf of  $b$ .

$$f'_{\alpha, \beta}(\alpha_0, \beta_0) \propto f_{\alpha, \beta}(\alpha_0, \beta_0) L(b_1, \dots, b_m | \alpha_0, \beta_0, \Phi) \quad 33$$

where  $b_1, \dots, b_m$  are the sizes of the targets found and  $\Phi$  is the effort. If, as in the previous example, grid borings are used as a search tool,

$$L(b_1, \dots, b_m | \alpha_0, \beta_0, \Phi) = \prod_{i=1}^m f_b(b_i | \alpha_0, \beta_0) \left(\frac{\pi}{4s^2}\right) b_i^2 \quad 34$$

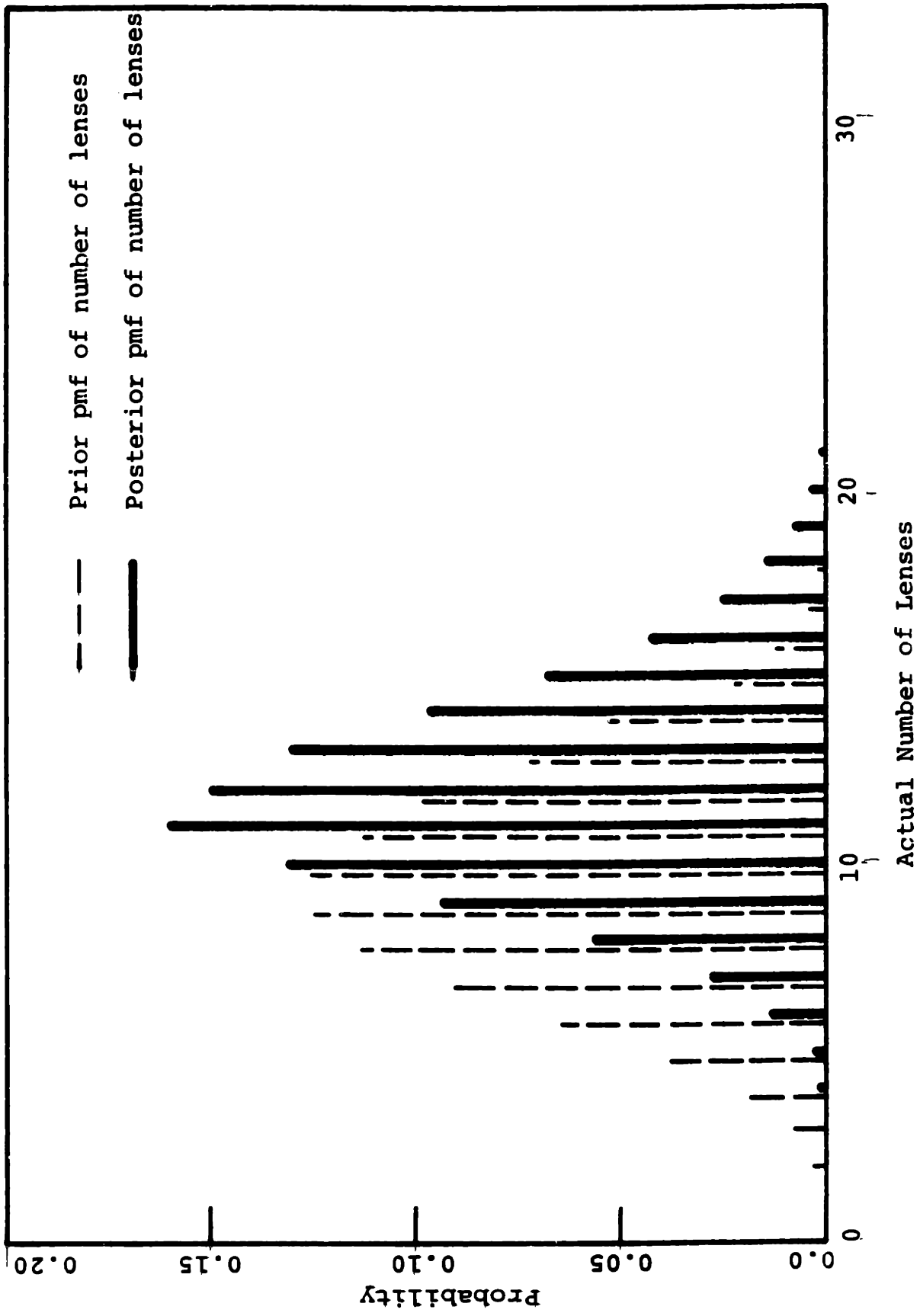


Figure 26: Prior and Posterior pmf's of Total Number of Lenses

and,

$$f_{\alpha, \beta}(\alpha_0, \beta_0 | b_1, \dots, b_m) = \frac{f_{\alpha, \beta}(\alpha_0, \beta_0) \prod_{i=1}^m f_b(b_i | \alpha_0, \beta_0) (\pi/4s^2) b^2}{\iint f_{\alpha, \beta}(\alpha_0, \beta_0) \prod_{i=1}^m f_b(b_i | \alpha_0, \beta_0) (\pi/4s^2) b^2 d\alpha_0 d\beta_0}$$

The prior distribution of  $b$  irrespective of the uncertain parameters  $\alpha$  and  $\beta$  (i.e., the marginal prior distribution) is,

$$f_b(b_0) = \iint f_b(b_0 | \alpha_0, \beta_0) f_{\alpha, \beta}(\alpha_0, \beta_0) d\alpha_0 d\beta_0$$

and similarly, the posterior distribution of  $b$  irrespective of  $\alpha$  and  $\beta$  is

$$f'_b(b_0) = \iint f_b(b_0 | \alpha_0, \beta_0) f'_{\alpha, \beta}(\alpha_0, \beta_0) d\alpha_0 d\beta_0$$

The application of Bayes' theorem to continuous variables frequently results in complicated equations, which can only be solved numerically. For this reason, one usually makes assumptions which simplify calculations. One such assumption is that the prior,  $f_{\alpha, \beta}(\alpha, \beta)$ , is a member of the family of conjugate distributions to the likelihood function,  $L(b_1, \dots, b_m | \alpha, \beta)$ .\* Such an assumption restricts our prior assessments, yet greatly reduces computation time. The development of conjugate distributions for common priors and search tools would be a valuable contribution to rational search.

---

\* A conjugate distribution is one such that if the prior distribution of  $\alpha$  and  $\beta$  belongs to this family, then for any set of observations  $(b_1, \dots, b_m)$  the posterior distribution of  $\alpha$  and  $\beta$  also belongs to this family.

### Section 6.1.3 Optimum Allocation

Grid and random search patterns allocate effort uniformly over the region of search. When there exists no prior information to indicate that some elemental areas are more likely than others to contain a target (i.e., uniform prior), there is no reason to allocate in any manner other than uniformly. Uniform allocation produces the highest probability of finding the target in this case.

When the prior probability distribution is not uniform -- that is, when some elemental areas are more likely to contain a target than others -- non-uniform allocations may increase the probability of a find. These allocations reduce, simply, to exerting more effort in those areas where a target is more likely to be.

The probability of finding a target in a certain elemental area equals the probability of finding it, given that it exists there, multiplied by the probability that it exists there.

$$P(\text{find in } \Delta A) = P(\text{find} \mid \text{exists in } \Delta A) P(\text{exists in } \Delta A)$$

35

By allocating more effort to those areas that have relatively a high probability of containing the target, and less effort to those areas that have relatively a low probability of containing the target we increase  $p(\text{find} \mid \text{target exists in } \Delta A)$  where the multiplying factor (i.e.,  $p(\text{exists in } \Delta A)$ ) is the greatest, thus increasing the total probability of a find.

The "detection function" of an exploration technique relates conditional probability of a find to amount of exploration effort

allocated in an elemental area. In our discussion of grid allocations we assumed this function to be constantly 1.0 (i.e., if a target existed in a search element it was found). When the prior pdf of location is uniform, imperfect observation does not alter the analysis significantly -- changing only the posterior pmf of total number of targets. When the location prior ceases to be uniform imperfect sampling assumes great interest because most (imperfect) detection functions obey diminishing returns. Allocations of increasingly large amounts of exploration effort do not produce proportionally large increases in the conditional probability of a find, and the optimum amount of effort to be allocated in a particular area is a function both of the probability of the target existing there and the rate of diminishing returns.

### Linear Detection Functions

Certain search techniques have detection functions which do not obey the law of diminishing returns. One such technique is systematic spacing of borings. Borings are discrete quantities of effort, and their density of allocation as a function of continuous area or length obeys a relationship similar to that shown in Figure 27. Were we to consider discrete elements of location (i.e., elemental areas or lengths) we could represent density of effort as a function of the number of borings per element, which, when large, can be

approximated by the continuous relationship\*

$$\varphi(s) \propto 1/s^n \quad 36$$

where  $n$  is the dimension of the search (e.g.,  $n = 2$  for search over a plane). Since the conditional probability of a find with systematic borings (equations 6,7) is proportional to  $1/s^n$

$$p(\varphi(x)) = k \varphi(x) \quad 37$$

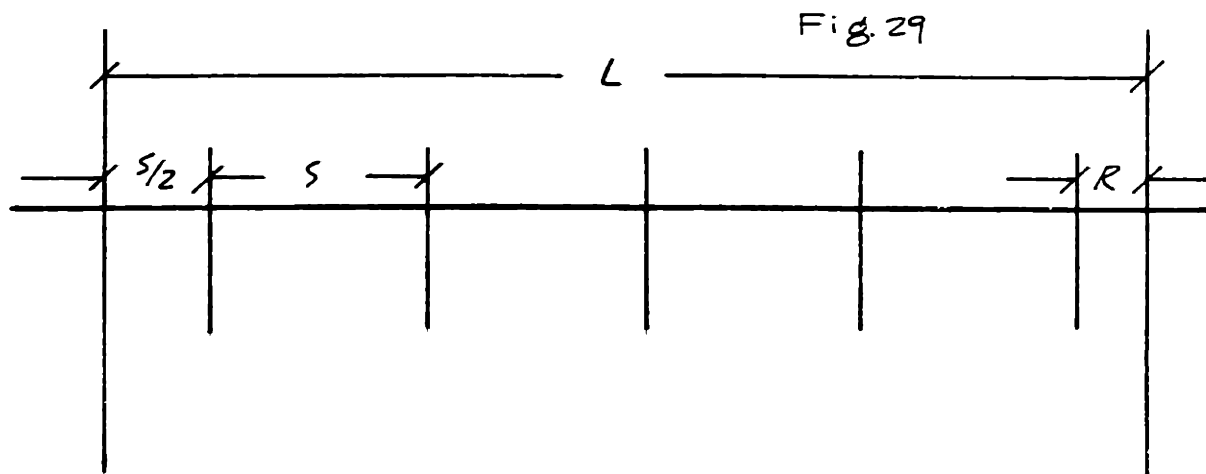
and systematic borings obey a linear detection function.

---

\*The requirement of  $s$  small relative to elemental area arises due to boundary effects, in that, a partial grid spacing at the boundary is not equivalent to a partial boring. On the average, assuming  $s$  small is neither conservative nor non-conservative, since the true probability of a find will be symmetrically distributed about  $k \phi(s)$ . Considering the one dimensional case, if  $n_s$  is the integral number of times that the element length less  $s/2$  is divisible by  $s$ , and  $R$  is the remainder,

$$R = (L - s/2) - n_s s$$

assuming  $s$  small is conservative (i.e.,  $p(\text{find}) \geq k \phi(s)$ ) if  $R \leq s/2$  and non-conservative is  $R > s/2$ .





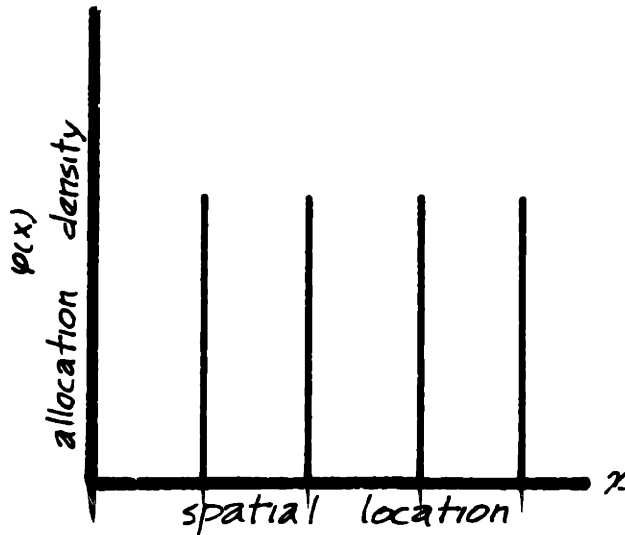


Fig. 27

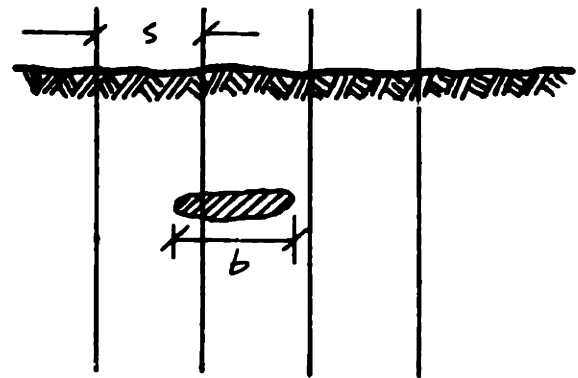


Fig. 28

Referring to equation 35, the probability of finding a target is

$$p(\text{find}) = \int_{-\infty}^{\infty} f_x(x_0) k \varphi(x_0) dx_0 \quad 38$$

$$= k \int_{-\infty}^{\infty} f_x(x_0) \varphi(x_0) dx_0 \quad 39$$

which is maximized by

$$\varphi(x_0) = \begin{cases} \bar{\Phi} & x_0: f_x(x_0) = \max. \\ 0 & \text{elsewhere} \end{cases}$$

In other words, by placing all of the effort in the elemental area having the highest probability of containing a target, the overall probability of a find is maximized.

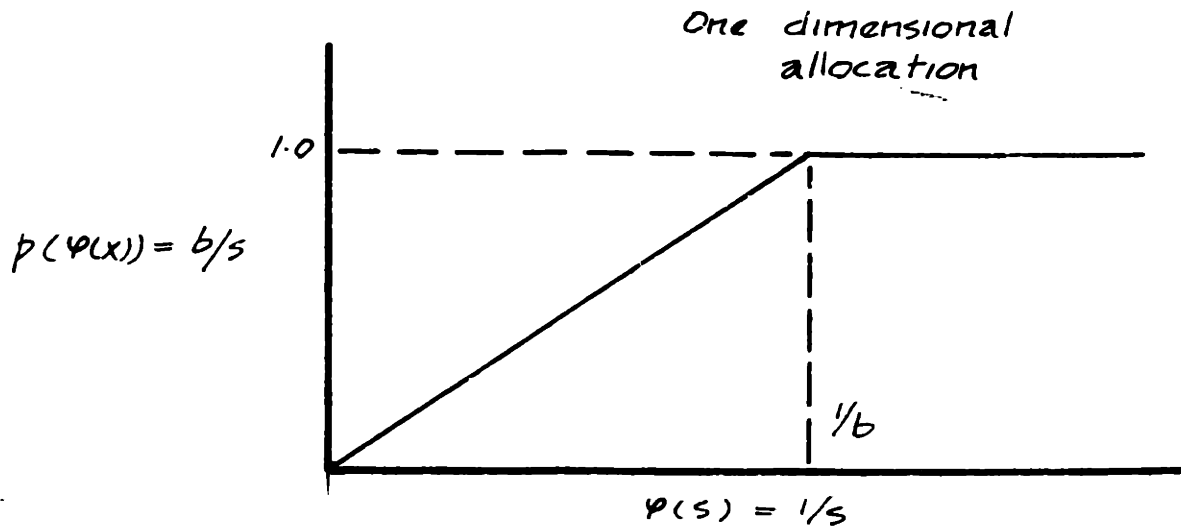


Fig. 30

However, since  $p(\varphi(x_0)) \leq 1.0$ , once  $p(\varphi(x_0))$  reaches 1.0 further increases in  $\varphi(x_0)$  do not cause additional increases in  $p(\varphi(x_0))$  (Figure 30). So, in reality, the optimal allocation is obtained by determining the number of elemental areas over which the total effort can be distributed such that  $p(\varphi(x_0)) = 1.0$ , and allocating the total effort in that density over the most probable elements.\* An example will help to clarify this process.

### Example

An engineer is to search a site (Figure 31) for solution cavities, and is to use systematically spaced borings as the exploration

---

\*A point worth noting is that although the probability of a find for square and triangular grids is the same, 23% fewer borings are required to achieve  $p(\text{find}) = 1.0$  with triangular grids than with square grids.

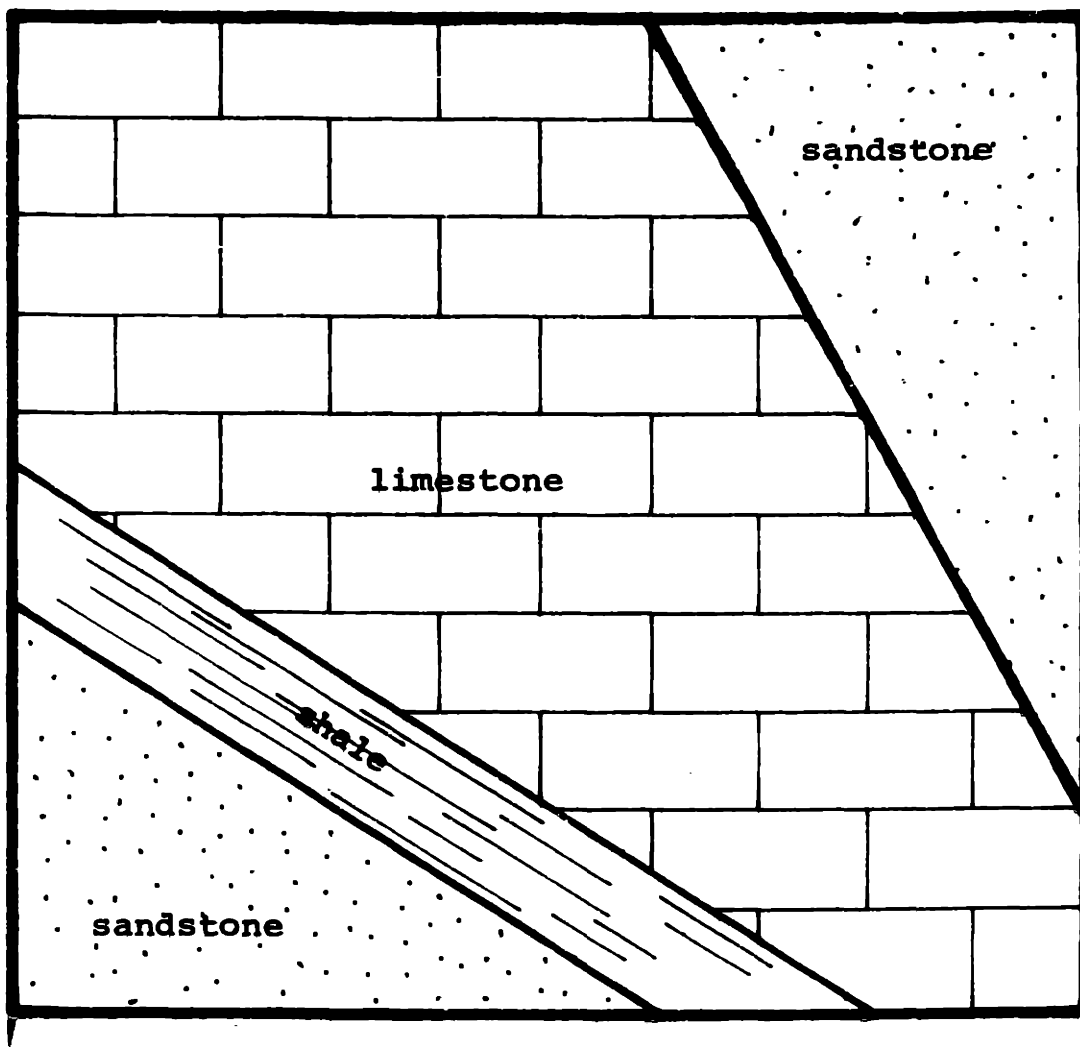


Figure 31

Geological map of the site  
showing sandstone, limestone, and shale beds

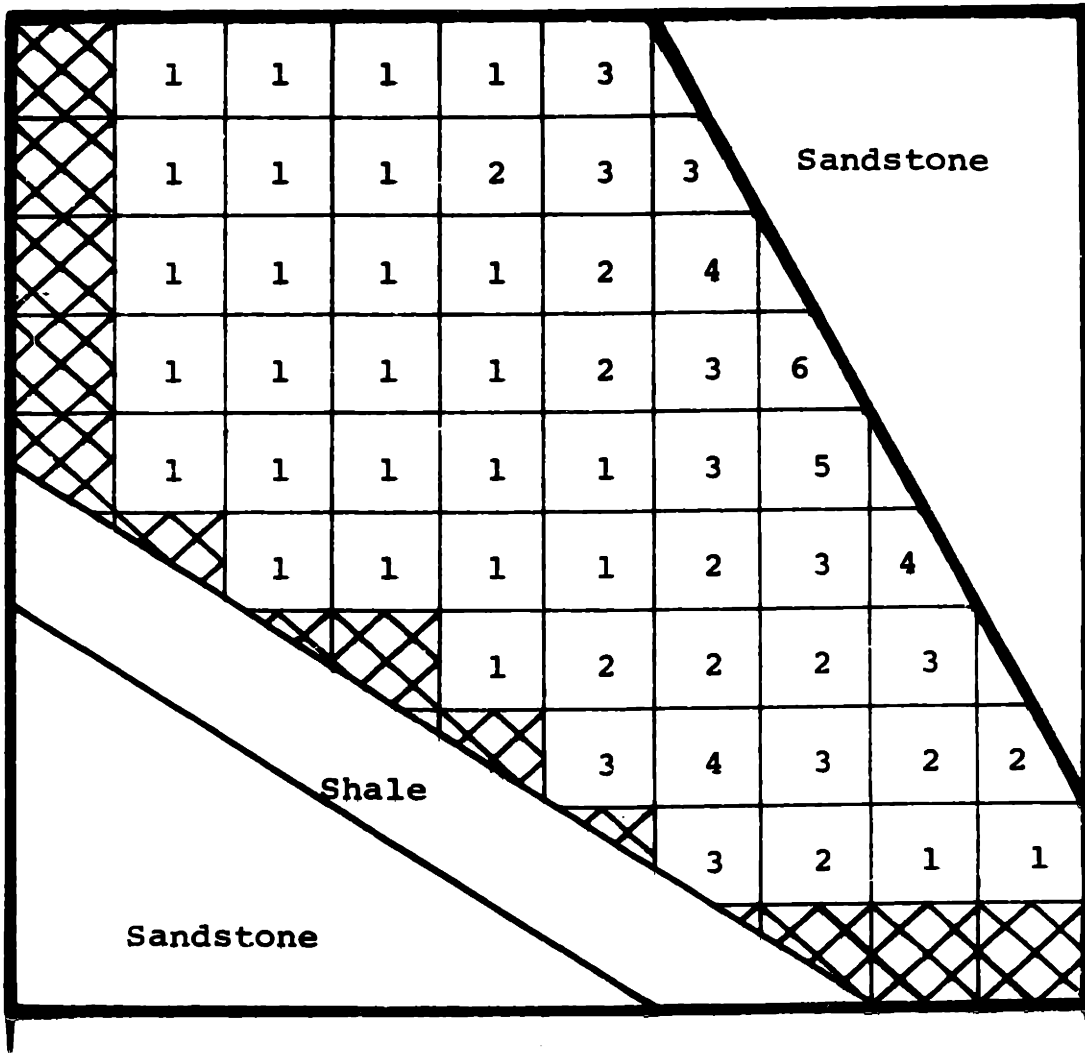


Figure 32

Numbers represent relative probabilities that sinkholes will be located in quadrates. Black-ed-in quadrates are not of interest to the engineer

tool. He has quantified his subjective probabilities by dividing the site into quadrants and evaluating the relative probabilities of each containing cavities (Figure 32). The size cavity of interest is 30'. From Figure 4, for  $p(\varphi(x))$  to equal 1.0,  $s$  must equal

$$s = 0.707 b = 21.2 \text{ ft.}$$

#### Allocation

1. From the text we know that for techniques with linear detection functions the optimal allocation is to exert enough effort to insure  $p(\varphi(x)) = 1.0$ , in the most probable region.
2. With  $\Phi = 100$  borings at 21.2', an area of  $(21.2 \times 10)^2 = 45,000$  sq. ft. can be explored with  $p(\varphi(x)) = 1.0$
3. Therefore, allocate borings in square grids at 21 ft. spacing in the

$$\frac{45,000}{10^4} = 4.5 \text{ (say. 5.0)}$$

most probable regions (Figure 33).

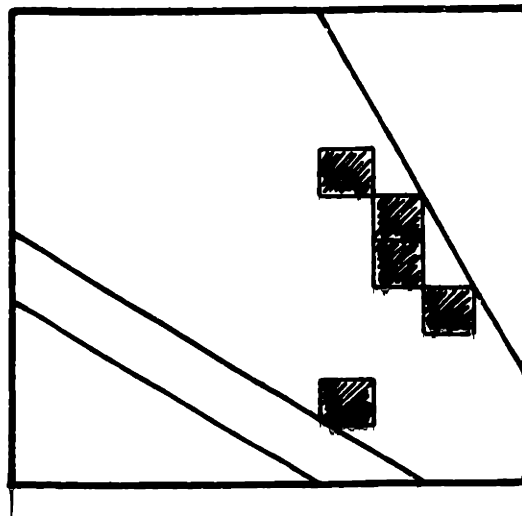


Fig. 33

Probability of a find | Existence:Opt. Allocation

$$\begin{aligned}
 P(\text{find}) &= P_1 + P_2 + \dots \\
 &= p_1 (p(\varphi)) + p_2 (p(\varphi)) + \dots \\
 &= \{ 6.0(1.0) + 5.0(1.0) + 3(4.0)(1.0) \} \times 10^{-2} \\
 &= 0.23
 \end{aligned}$$

Unif Allocation

for 100 borings  $\longrightarrow$   $S \cong 71$  ft.

$$\begin{aligned}
 p(p(x)) &= \frac{\pi}{4} \frac{b^2}{S^2} \\
 &= 0.142 \\
 p(\text{find}) &= \sum_i p_i (0.142) \\
 &= 0.14
 \end{aligned}$$

Exponential Saturation Detection Functions: Koopman's Allocation

The classic work on search theory, developed by B. O. Koopman and others under the sponsorship of the Office of Naval Research is predicated on the so-called "exponential-saturation" detection function.

Consider a target whose location can be specified by one

parameter  $x$  (i.e., search along a line), and let the prior pdf of the target's location be denoted  $f_x(x_0)$ .<sup>\*</sup> If one quantum of effort is exerted in searching elemental region  $\Delta x$  we assume the probability of finding a target, given that it exists within  $\Delta x$ , is constant and independent of  $x$ . Let this probability be  $p$ . If  $n$  quanta are exerted in the  $\Delta x$  the probability of a find is

$$p(\text{find} \mid \text{exists in } \Delta x) = 1 - (1-p)^n \quad 41a$$

$$\cong 1 - e^{-np} \quad \text{for } p \leq 0.1 \quad 41b$$

$$= 1 - e^{-\varphi(x)} \quad 41c$$

where  $\varphi(x)$  is the search density at location  $x$  (i.e., effort per unit distance). Equation 41c is known as the "exponential-saturation" form of the law of detection, and is a supposition of the Koopman theory. Many search techniques can be modeled by exponential laws (e.g., passes with an airplane, allotted time in field reconnaissance, so-called independent "glimpses" by an observer (Koopman, 1957; DeGuenin, 1961)); thus Koopman's theory is applicable to a host of practical cases; however, since  $p$  must be assumed small (i.e.,  $p \leq 0.1$ ) exponential saturation functions are not applicable universally.<sup>\*\*</sup>

---

<sup>\*</sup>This is a subjective probability assessment which can be evaluated by the methods of chapter 2.

<sup>\*\*</sup>E.g., were borings being used to search for a clay-filled seam which is washed out of a recovered core 30% of the time (i.e.,  $p = 0.7$ ), an exponential-saturation model could not be used.

The prior probability of a target existing between  $x_0$  and  $x_0 + dx$  is  $f_x(x_0)dx$ .

Assume for the present that a target is known to exist somewhere at the site; in other words, the probability of existing  $\beta$  is

$$\beta = \int_{-\infty}^{\infty} f_x(x_0) dx = 1.0 \quad 42$$

The only restriction on  $f_x(x_0)$  is that it be continuous, with the exception of simple infinities and finite jumps.

The amount of search effort allocated to the interval  $(x_0, x_0 + dx)$  equals

$$\varphi(x_0) dx_0 \quad 43$$

The function  $\varphi(x_0)$  has the properties

$$\varphi(x_0) \geq 0 \quad \text{for all } x_0 \quad 44$$

and

$$\int_{-\infty}^{\infty} \varphi(x_0) dx_0 = \Phi = \text{total search effort} \quad 45$$

By equation 41c the probability of detecting a target between  $x_0$  and  $x_0 + dx$  is

$$\text{Pr} \{ \text{detection in } (x_0, x_0 + dx) \} = f_x(x_0) [1 - e^{-\varphi(x_0)}] \quad 46$$

and the total probability of a find under the allocation  $\varphi(x)$  is

$$p(\text{find}) = \int_{-\infty}^{\infty} f_x(x_0) \text{Pr} \{ \text{detection in } (x_0, x_0 + dx) \} dx_0 \quad 47$$

$$= \int_{-\infty}^{\infty} f_x(x_0) [1 - e^{-\varphi(x_0)}] dx_0 \quad 48$$



The optimum allocation function,  $\psi^*(x_0)$ , is the one which maximizes equation (48) while conforming to the constraints of equations (44) and (45).

Koopman (1957) shows that for all  $x_0$  such that  $\psi(x_0) \geq 0$  the optimum search density function satisfies

$$f_x(x_0) e^{-\psi^*(x_0)} = \lambda \quad 49$$

where  $\lambda$  is a constant. Taking the natural logarithm of both sides of equation (49)

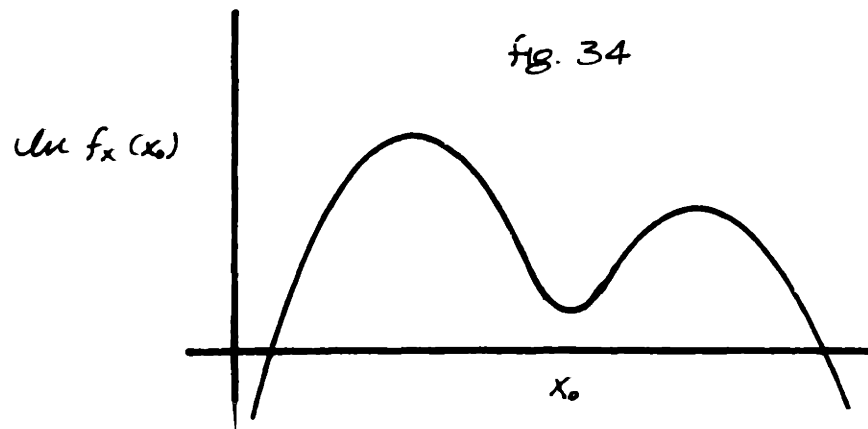
$$\psi(x_0) = \ln f_x(x_0) - \ln \lambda \quad 50$$

and since the total amount of search effort is  $\Phi$

$$\Phi = \int_{-\infty}^{\infty} \psi^*(x_0) dx_0 = \int_{-\infty}^{\infty} \{ \ln f_x(x_0) - \ln \lambda \} dx_0 \quad 51$$

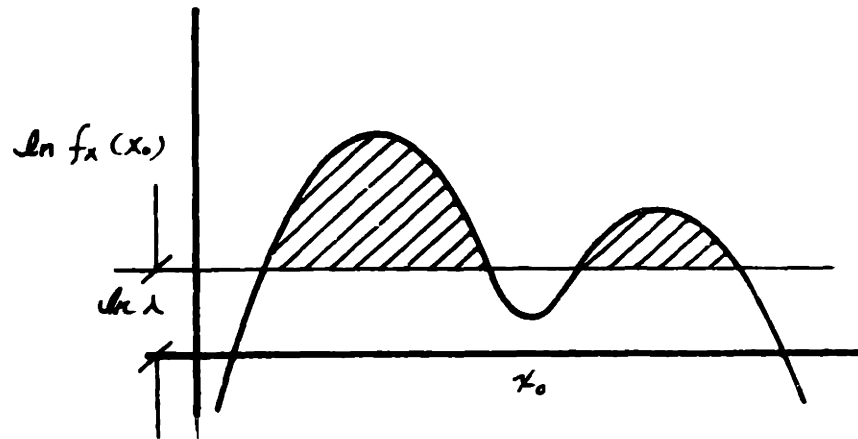
This relationship may be used to determine  $\psi^*(x_0)$  by a simple graphical construction (the mathematical basis of which is to be found in Koopman, 1957), which is quite fortunate.

STEP 1: Graph the natural logarithm of the prior pdf,  $\ln f_x(x_0)$



STEP II: Draw a line parallel to the x-axis truncating the graph of  $\ln f_x(x_0)$ . This line is a plot of  $\ln \lambda$ , where  $\lambda$  is the constant of equation 45.

Fig. 35



STEP III: Determine the area under the curve in  $f_x(x_0)$  and above the line  $\ln \lambda$ . This area is

$$\int_{-\infty}^{\infty} \{ \ln f_x(x_0) - \ln \lambda \} dx$$

the r.h.s. of equation (47).

Under optimal allocation this should equal the total search effort,  $\Phi$ .

By moving the horizontal line up and down (i.e., by varying  $\lambda$ ), this area can be made to equal  $\Phi$  and the value of  $\lambda$  thus determined satisfies the criterion for optimal allocation of equation (49).

The optimal allocation of effort is, therefore, to exert effort only to those locations for which

$$\{ \ln f_x(x_0) - \ln \lambda \} > 0$$

and in proportion to  $(\ln f_x(x_0) - \ln \lambda)$  (Figure 26).

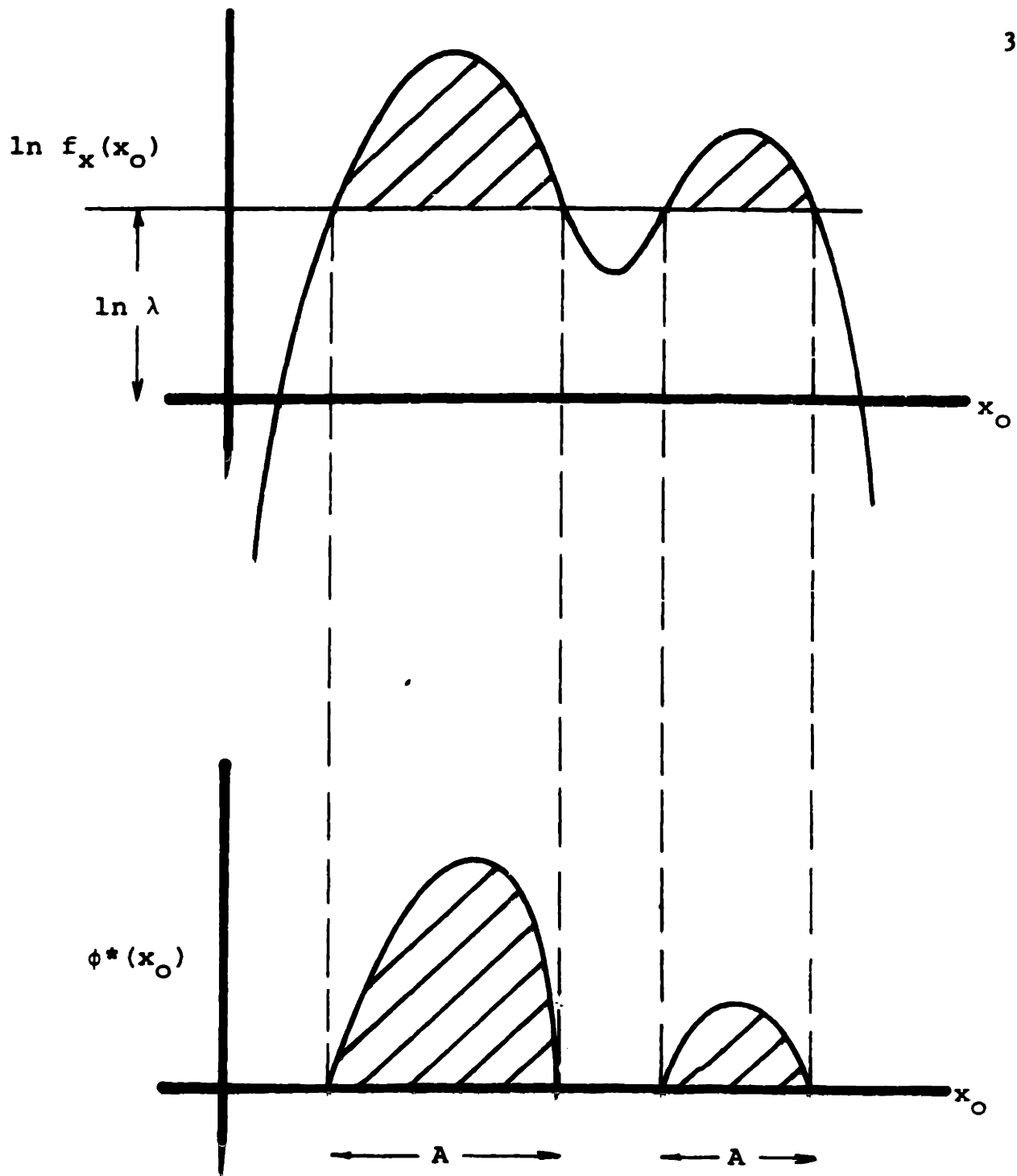


Figure 36

The probability of finding the target given that it exists somewhere at the site is, from equation (48)

$$P = \int_{-\infty}^{\infty} f_x(x_0) [1 - e^{-\varphi(x_0)}] dx_0 \quad 52$$

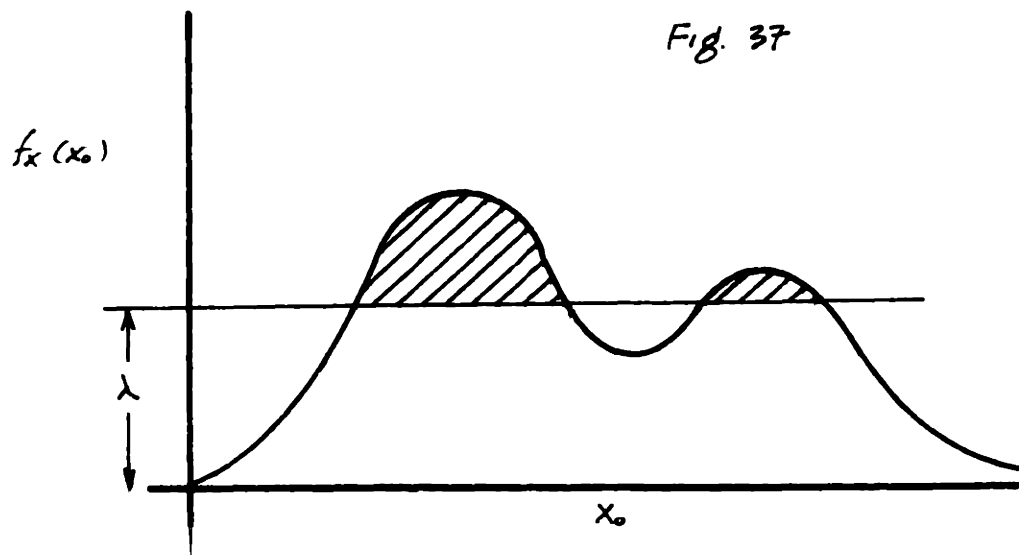
by equation (46)

$$= \int_{-\infty}^{\infty} f_x(x_0) [1 - e^{-(\ln f_x(x_0) - \ln \lambda)}] dx_0 \quad 53$$

$$= \int_{-\infty}^{\infty} f_x(x_0) dx_0 - \int_{-\infty}^{\infty} f_x(x_0) e^{\ln(\lambda/f_x(x_0))} dx_0 \quad 54$$

$$= \int_{-\infty}^{\infty} (f_x(x_0) - \lambda) dx_0 \quad 55$$

which is just the area between the curves  $f_x(x_0)$  and  $\lambda$  (Figure 37).



In the event that the target is not found by the allocation of  $\Phi$  amount of effort, a posterior pdf of the target location can be determined by direct application of Bayes' Theorem.

$$f_x'(x_0) = \frac{f_x(x_0) p(\text{no find} | x_0)}{\int f_x(x_0) p(\text{no find} | x_0) dx_0} \quad 56$$

$$= \frac{f_x(x_0) e^{-\varphi(x_0)}}{\int f_x(x_0) e^{-\varphi(x_0)} dx_0} \quad 57$$

The denominator of the right hand side of equation (57) reduces to

$$\int_{-\infty}^{\infty} f_x(x_0) e^{-\varphi(x_0)} dx_0 = \int_{-\infty}^{\infty} \lambda dx \quad 58a$$

and from equation (55)

$$= \left[ \int_{-\infty}^{\infty} f_x(x_0) dx_0 \right] \quad 58b$$

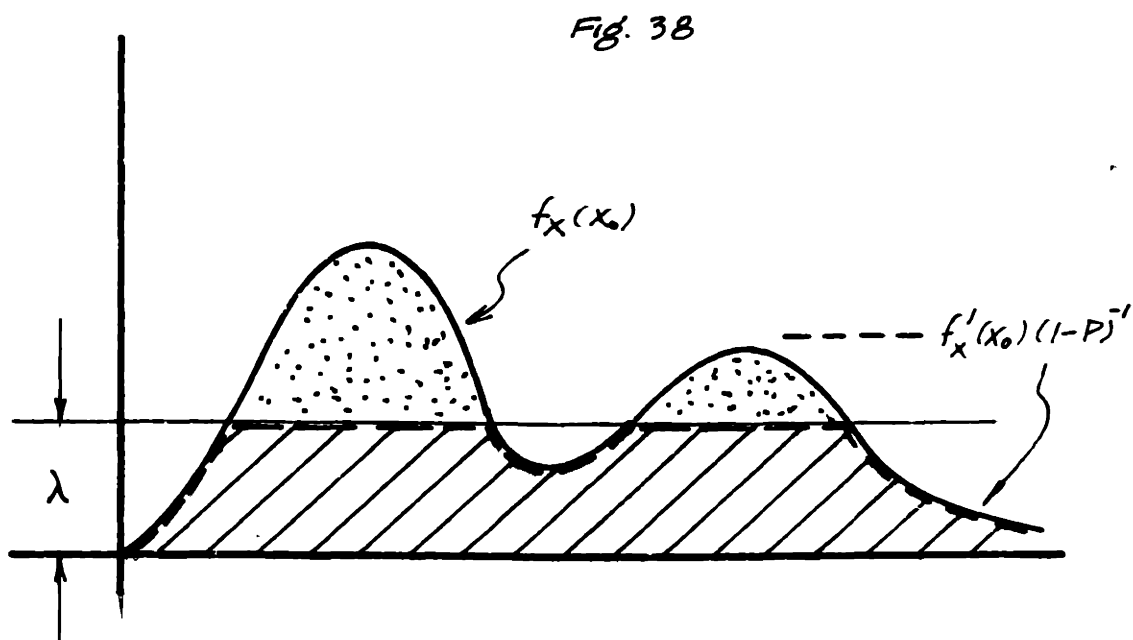
$$= 1 - P \quad 58c$$

For locations  $x_0$  outside of the regions A (Figure 36) (i.e.,  $x: x_0 \notin A$ )  $\varphi(x) = 0$ , and

$$e^{-\varphi(x_0)} = 1 \quad 59$$

so the right hand side of equation (57) equals  $f_x(x_0)/(1-P)$ . For locations  $x$  inside the regions A (i.e.,  $x: x_0 \in A$ ) the numerator of the right hand side of equation (57) equals  $\lambda$  (equation 49).

Therefore, the posterior pdf of location for any  $x_0$  is proportional (i.e., re-normalized to unit area) to the distance between the x-axis and either the curve  $f_x(x_0)$  or  $\lambda$ , whichever is least, by the factor  $(1-P)^{-1}$  (Figure 38).



In the event that the prior probability of the target existing was less than 1.0, and no find has been made, the posterior probability,  $\beta'$ , can be determined by Bayes' Theorem exactly as described in dealing with grid search (Sec. 2.2).

The conditional probability of a find  $P$  under the allocation equals the area between the curves  $f_x(x_0)$  and  $\lambda$  (the stippled area in Figure 38). The conditional probability of a miss is the complement of  $P$ , or the crosshatched area in Figure 38.

**Problem:**

An engineer is scanning airphotographs for evidence of sinkholes. The site is the proposed location of a new highway and the engineer has arbitrarily divided the site into ten intervals (corresponding to ten photographs of the site). He knows the general geology of each of the ten intervals and has quantified his subjective probabilities that sinkholes will be found in each of the ten. His normalized "fellings" are shown in Figure 40.

Each of the photographs is 10" by 10", and the engineers "scanning rate" is 2 square inches per minute. Koopman (1957) has shown the detection function for scanning to be the exponential,

$$p(\text{find at } x_0 | \text{target at } x_0, \phi(x_0)) = 1 - e^{-t\rho}$$

where  $t$  is the allotted time and  $\rho$  is the scanning rate. If one day is available for this task, how should the engineer allocate his effort between photographs?

**Solution:**

$$\text{total time} = T = 8 \text{ hours} = 480 \text{ minutes} \\ \text{(say 500)}$$

$$\text{total effort} = \phi = T\rho = 1000 \text{ square inches}$$

**Step 1:** Plot  $\log p(x_0)$  vs. photograph (Figure 41)

**Step 2:** In step 2 several trial values of the constant  $\lambda$  are computed and the correct one determined graphically.

Guess:  $\ln \lambda = -3.0$   
 $\phi =$  area cross-hatched in Figure 41.  
 $= 540$  square inches, which is too small.

Guess:  $\ln \lambda = -3.8$   
 $\phi = 12.6$ , too large.

Guess:  $\ln \lambda = -3.4$   
 $\phi = 9.0$ , too small

Guess:  $\ln \lambda = -3.6$   
 $\phi = 10.8$ , too large

From Figure 42 the correct  $\lambda$  is  $-3.5$

**Step 3:** Taking  $\lambda = -3.5$ , the optimal distribution of time is shown in Figure 43.

**Problem extension:**

What happens if the allotted time for the entire task is cut in half? That is, the total time  $T$  is assumed to be 250 minutes rather than 500. For  $T = 250$ ,  $\phi = 500$  square inches. Referring to Figure 41,



Figure 42

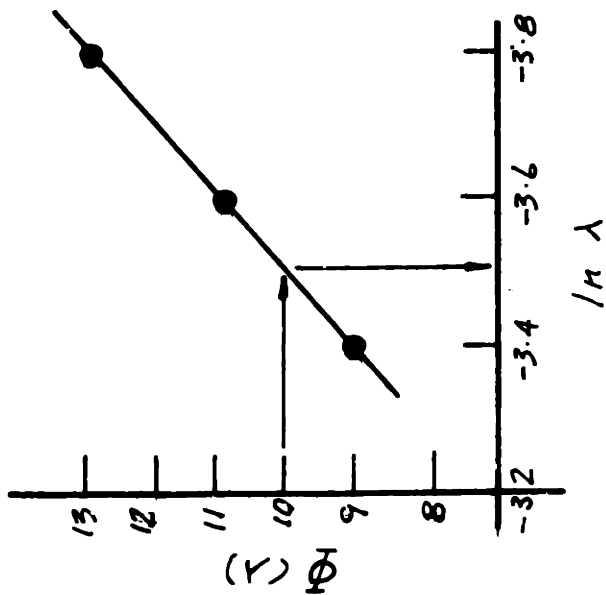


Figure 41

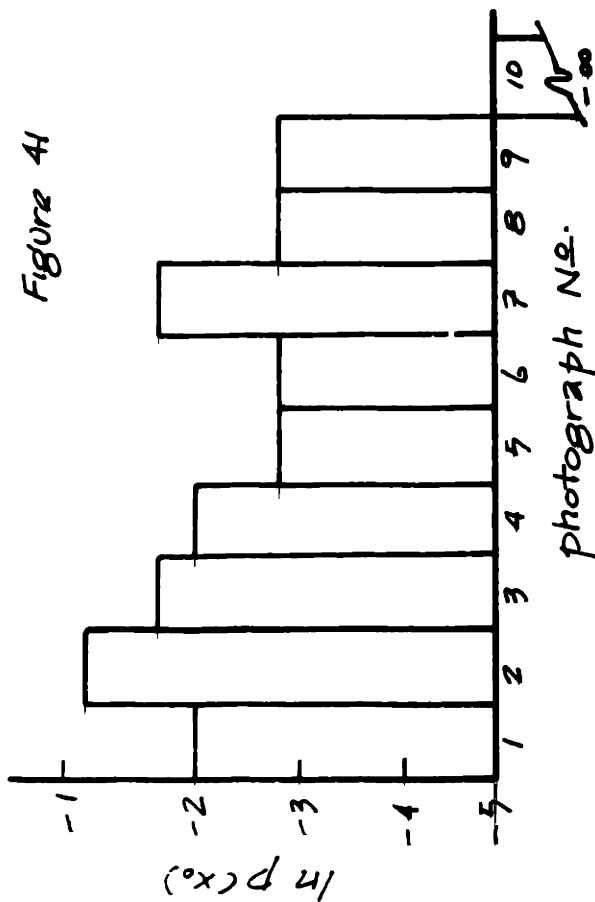


Figure 43

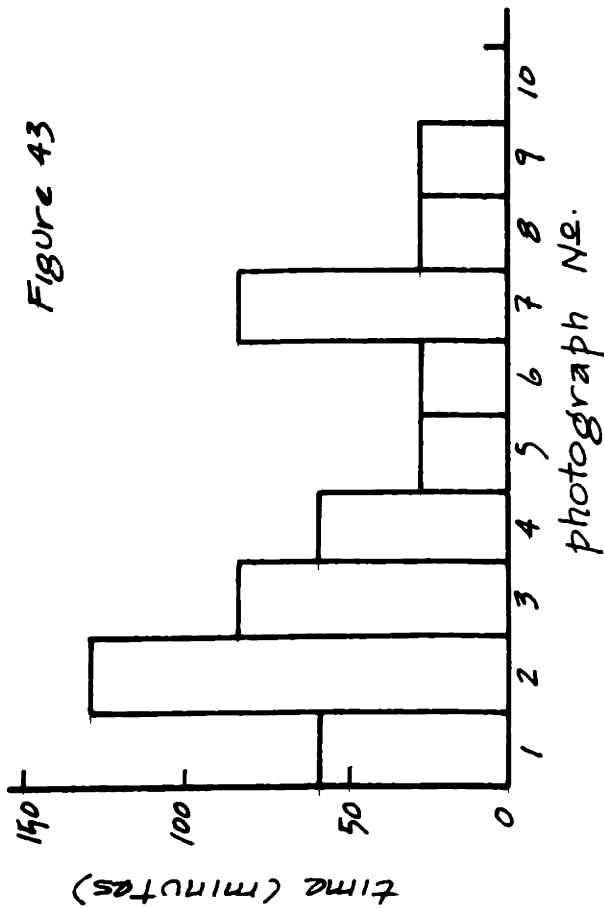


Figure 46

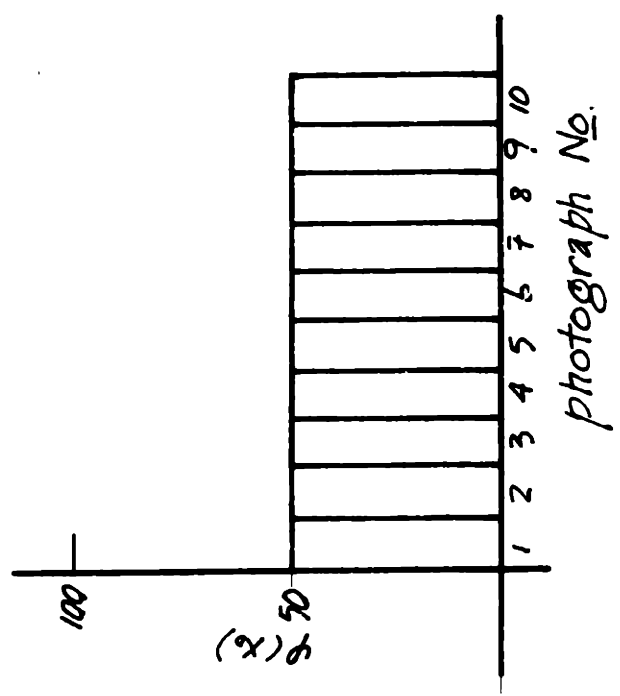


Figure 44

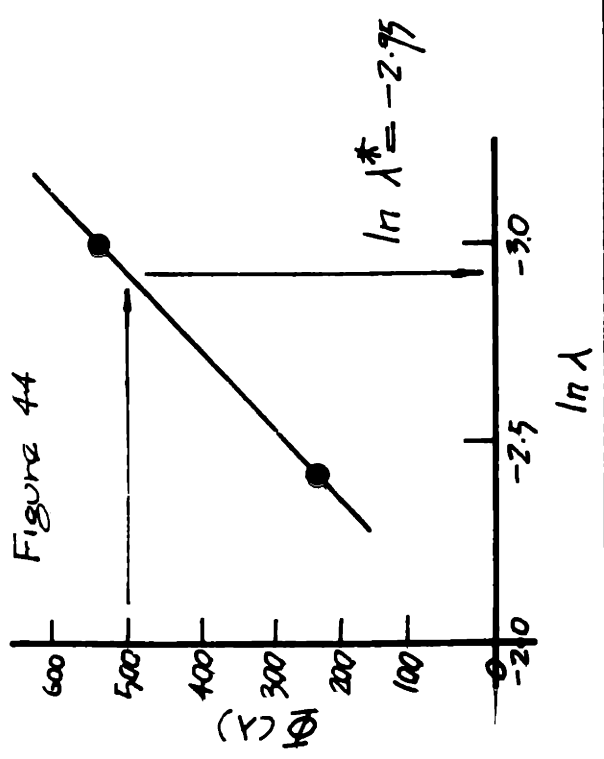
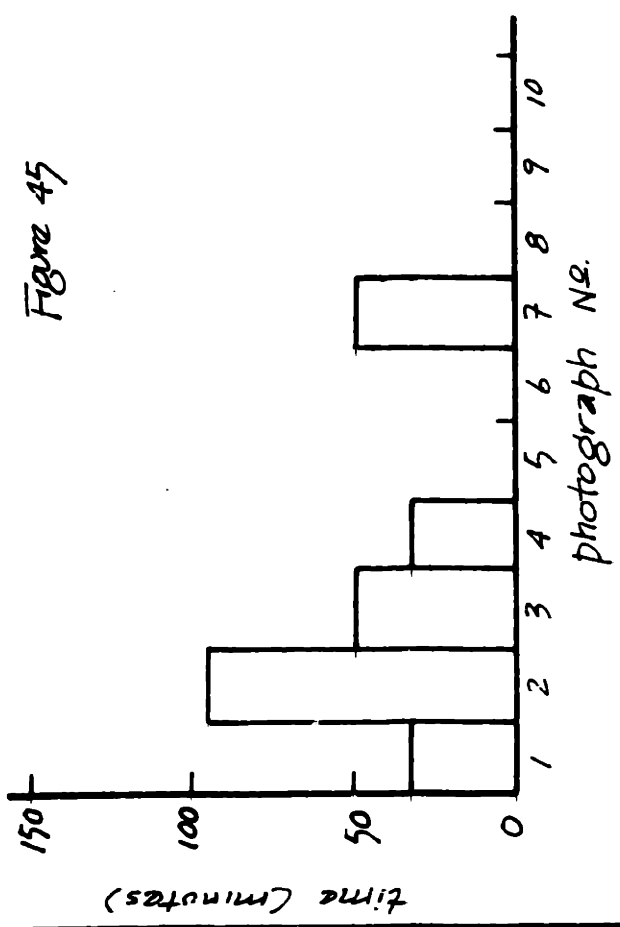


Figure 45



Guess:  $\ln \lambda = -3.0$   
 $\phi = 540$  square inches, too large.  
 Guess:  $\ln \lambda = -2.4$   
 $\phi = 240$ , too small.

From Figure 44 the correct value of  $\lambda$  is seen to be  $\lambda = -2.95$ .

So, with  $T = 250$  minutes and  $\lambda = -2.95$ , the optimal distribution of effort is that shown in Figure 45. Under the optimal allocation photographs 5, 6, 8, 9, and 10 are not searched at all!

We can now compare the probability of finding a solution cavity using the common uniform allocation with the optimal allocation -- assuming a cavity exists.

Probability of a find, given that solution cavity exists:

1--Optimal allocation:

$$\begin{aligned}
 P(\text{find}) &= P(\text{find in photograph 1}) \\
 &\quad + P(\text{find in photograph 2}) \\
 &\quad + \dots \\
 &= P_1 + P_2 + \dots
 \end{aligned}$$

where  $P_i = p_i (1 - e^{-\rho t_i / A_i})$  in which  $p_i$  is the subjective probability of a cavity in the  $i$  th photograph,  $t_i$  is the time spent looking in the  $i$  th photograph, and  $A_i$  is the area of the  $i$  th photograph.

Photo.	$P_i$	$t_i$	$\rho t/A$	$e^{-\rho t/A}$	$1-e^{-\rho t/A}$	$\overset{316}{x P_i}$
1	0.1	61	0.61	0.54	0.46	0.046
2	0.3	174	1.74	0.18	0.82	0.246
3	0.15	102	1.02	0.36	0.64	0.096
4	0.1	61	0.61	0.54	0.46	0.046
5	0.05	00	0.00	0.00	0.00	0.000
6	0.05	0	0.00	0.00	0.00	0.000
7	0.15	102	1.02	0.36	0.64	0.096
8	0.05	0	0.00	0.00	0.00	0.000
9	0.05	0	0.00	0.00	0.00	0.000
10	0.00	0	0.00	0.00	0.00	0.000
					sum	0.53

2--Uniform allocation:

$$P(\text{find}) = \sum_{i=1}^{10} P_i (1-e^{-\rho t/A_i})$$

Photo.	$P_i$	$\rho t/A$	$(1-e^{-\rho t/A_i})$	$x P_i$	
1	0.1	0.5	0.39	0.039	
2	0.3	"	"	0.117	
3	0.15	"	"	0.059	
4	0.1	"	"	0.039	
5	0.05	"	"	0.020	
6	0.05	"	"	0.020	
7	0.15	"	"	0.059	
8	0.05	"	"	0.020	
9	0.05	"	"	0.020	
10	0.00	"	"	0.000	
				sum	0.313

Another interesting allocation would be to allocate all of the available effort (i.e., time) to the photograph with the largest probability of containing a sinkhole. This is photograph #2. The probability of a find, given that a sinkhole exists, with this allocation is,

$$\begin{aligned} P(\text{find}) &= p_2 (1 - e^{-250\rho/A_1}) \\ &= 0.298 \end{aligned}$$

**Conclusion:**

Optimal allocation	P (find)	=	53%
Uniform allocation	P (find)	=	31%
All effort to most probable	P (find)	=	29.8%

Note that in this case, by spending all of our time searching the most probable photograph we obtain almost the same probability of a find as when we search each photograph the same amount of time. This occurs because the conditional probability of a find approaches the probability that a target exists in the photograph.

Figure 39

Plan view of proposed highway right-of-way showing locations of airphotographs. Number above each photograph location indicates the probability of sinkholes existing in that elemental area.

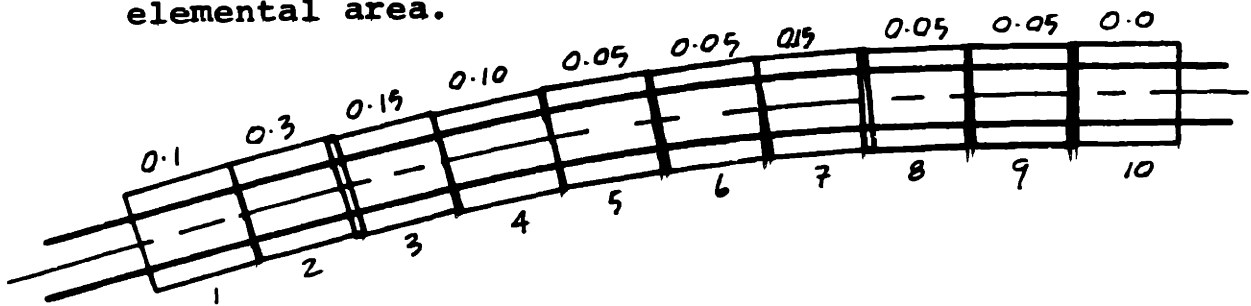
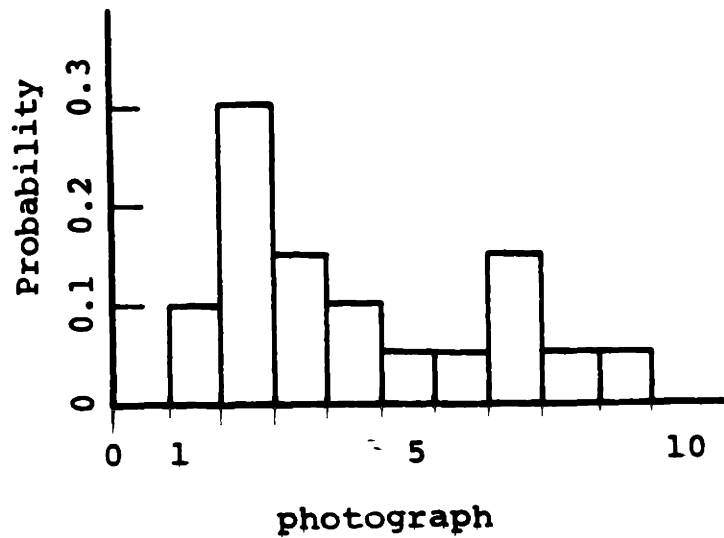


Figure 40

pmf of sinkholes in each photograph (i.e., elemental area)



Higher Dimensions

Extension of Koopman's theory to higher dimensions is made by substituting the multivariable pdf

$$f(x_1, x_2, \dots, x_n)$$

for

$$f_x(x_0)$$

and the multivariate integration

$$\int_{x_1} \int_{x_2} \dots \int_{x_n} f(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n \quad 60$$

for

$$\int_{x_0} f_x(x_0) dx_0 \quad 61$$

in equations 45 through 58c. The construction to determine  $\varphi^*(x_1, x_2, \dots, x_n)$  is analogous to the construction already discussed to determine  $\varphi^*(x_0)$ . In the two-dimensional case, the volume between the surface  $\ln(f(x_1, x_2))$  and  $\ln \lambda$  is varied by varying the value of the constant  $\lambda$  until this volume equals the total amount of search effort. The optimum is then to allocate effort only to those elemental areas for which

$$\ln f(x_1, x_2) > \ln \lambda \quad 62$$

and to expend effort in each in proportion to

$$\varphi^*(x_1, x_2) \propto |\ln f(x_1, x_2) - \ln \lambda| \quad 63$$

In other words, the construction for multidimensional spaces is

strictly analogous to the construction for one-dimensional space.

### General Diminishing-return Functions

Koopman's theory is limited by the assumption of an exponential detection function, which only applies to search techniques with a small conditional probability of a find,  $p$ . DeGuenin (1961) extended Koopman's theory to cover all detection functions which obey diminishing returns, that is, all detection functions for which

$$\frac{d}{d(\varphi(x))} \{P(\varphi(x_0))\} \quad 64$$

is a decreasing function of  $\varphi(x_0)$

DeGuenin has shown that a more general statement of equation (49) for the optimal allocation  $\varphi^*(x)$  is

$$f_x(x_0) \frac{d}{d\varphi(x)} \{P(\varphi^*(x_0))\} = \lambda, \text{ a constant} \quad 65$$

which, in the case of an exponential law of detection reduces to equation 49

$$f_x(x_0) e^{-\varphi(x_0)} = \lambda \quad 66$$

Based on this modification, deGuenin proposes an iterative graphical procedure, somewhat more involved than Koopman's, by which the optimal allocation  $\varphi^*(x_0)$  can be determined.



STEP I: Select an initial value of  $\lambda$ , and evaluate the quantity  $\lambda / p'(0)$  where

$$p'(0) = \left. \frac{d}{d\varphi(x)} p(\varphi(x_0)) \right|_{\varphi(x_0)=0} \quad 67$$

( $p'(0)$  = the derivative of the detection function with respect to  $\varphi(x)$  evaluated at  $\varphi(x_0) = 0$ ).

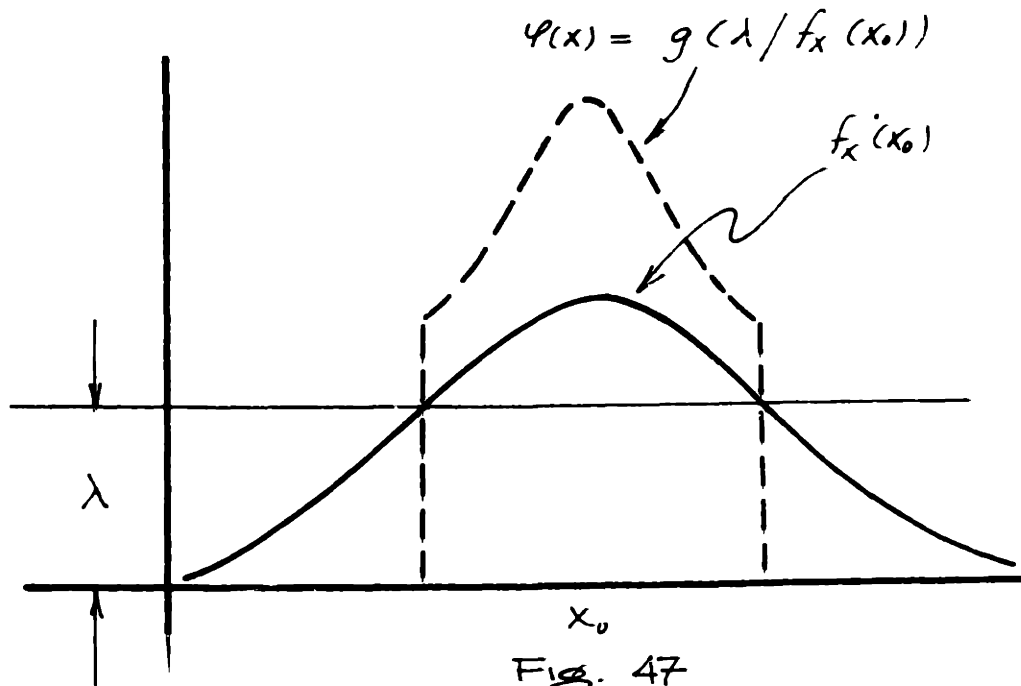
STEP II: Rearranging equation (65)

$$\frac{d}{d\varphi(x)} p(\varphi(x_0)) = \lambda / f_x(x_0) \quad 68$$

By limiting our discussion to detection functions for which the derivation with respect to  $\varphi(x)$  is continuous, an inverse function

$$\varphi(x_0) = g(\lambda / f_x(x_0))$$

is known to exist. From the value of  $\lambda$  of step I and  $f_x(x_0)$ , determine and graph  $\varphi(x_0)$  as shown in Figure 47.



STEP III: Since  $\int \varphi(x) dx = \Phi$  the area under the curve  $\varphi(x)$  should equal the total exploration effort. If  $\int \varphi(x) dx > \Phi$ , repeat from step I with a larger value of  $\lambda$ . If  $\int \varphi(x) dx < \Phi$  repeat with a smaller value of  $\lambda$ . DeGuenin has shown that this procedure is convergent to the optimal allocation function.

#### EXAMPLE

A field geologist has allotted 1 day to reconnoiter a proposed site for some specific target. Due to several factors the probability that he will miss seeing the target even though he is looking at the right location has been estimated to be 0.4. The detection function is, therefore,

$$p(\varphi(x)) = 1 - (1-p)^{\varphi(x)} = 1 - (1-p)^{0.4n}$$

For values of  $p$  greater than 0.1, this detection function is not exponential-saturation, although it does obey diminishing returns. Determination of the optimal allocation  $\varphi^*(x)$  can be made by deGuenin's method.

#### Solution

The pdf of location of the target along a line  $x$ , 1000 m in length is shown in Figure 49. The total effort is one 8 hour day, or 480 min.

Trial #1

1--Guess  $\lambda = 0.5 \times 10^{-3}$

$$\lambda/p'(0) = \frac{0.5 \times 10^{-3}}{0.51}$$

2-- $\phi(x) = \frac{\ln(\lambda/0.51 f(x))}{-0.51}$  (min./m)

3--Area 1 (Figure 49):

$$A_1 \approx 1/2(500)1.38 = 345 \text{ min} < 480 \text{ min}$$

Therefore, chosen  $\lambda$  is too large

Trial #2

1--Guess  $\lambda = 0.4 \times 10^{-3}$

$$\lambda/p'(0) = 0.59 \times 10^{-3}$$

2--

$f(x)$	$\lambda/0.51f(x)$	$\ln(\cdot)$	$\div(-0.51)$
$0.59 \times 10^{-3}$	1.0	0	0
1.00	0.59	-0.53	1.04
1.50	0.39	-0.94	1.78
2.00	0.29	-1.24	2.42

3-- $A_2 = 1/2(700 \times 2.42) + 432(0.05 \times 10)$   
 $= 850 + 216$   
 $= 1066 > 480$

Therefore, chosen  $\lambda$  is too small

Trial #3

1--Guess  $\lambda = 0.4 \times 10^{-3}$

$$\lambda/p'(0) = 0.78 \times 10^{-3}$$

2--

$f(x)$	$\lambda/0.51f(x)$	$\ln(\cdot)$	$\div(-0.51)$
$0.78 \times 10^{-3}$	1.0	0	0
1.00	0.78	-0.24	0.47
1.50	0.52	-0.65	1.28
2.00	0.39	-0.94	1.85

$$\begin{aligned}
 3--A_3 &= 1/2(580 \times 1.85) + 115(0.5) \\
 &= 540 + 58 \\
 &= 598 > 480
 \end{aligned}$$

Optimal Allocation,  $\phi^*(x)$ 

1-- $\lambda = 0.43 \times 10^{-3}$

$$\lambda/p'(0) = 0.84 \times 10^{-3}$$

2--

$f(x)$	$\lambda/0.51f(x)$	$\ln(\cdot)$	$\div(-0.51)$
$0.84 \times 10^{-3}$	1.0	0	0
1.00	0.84	-1.75	0.34
1.50	0.56	-0.58	1.14
2.00	0.42	-0.87	1.71

3-- $A_4 = 480 \text{ min} = \phi_t$

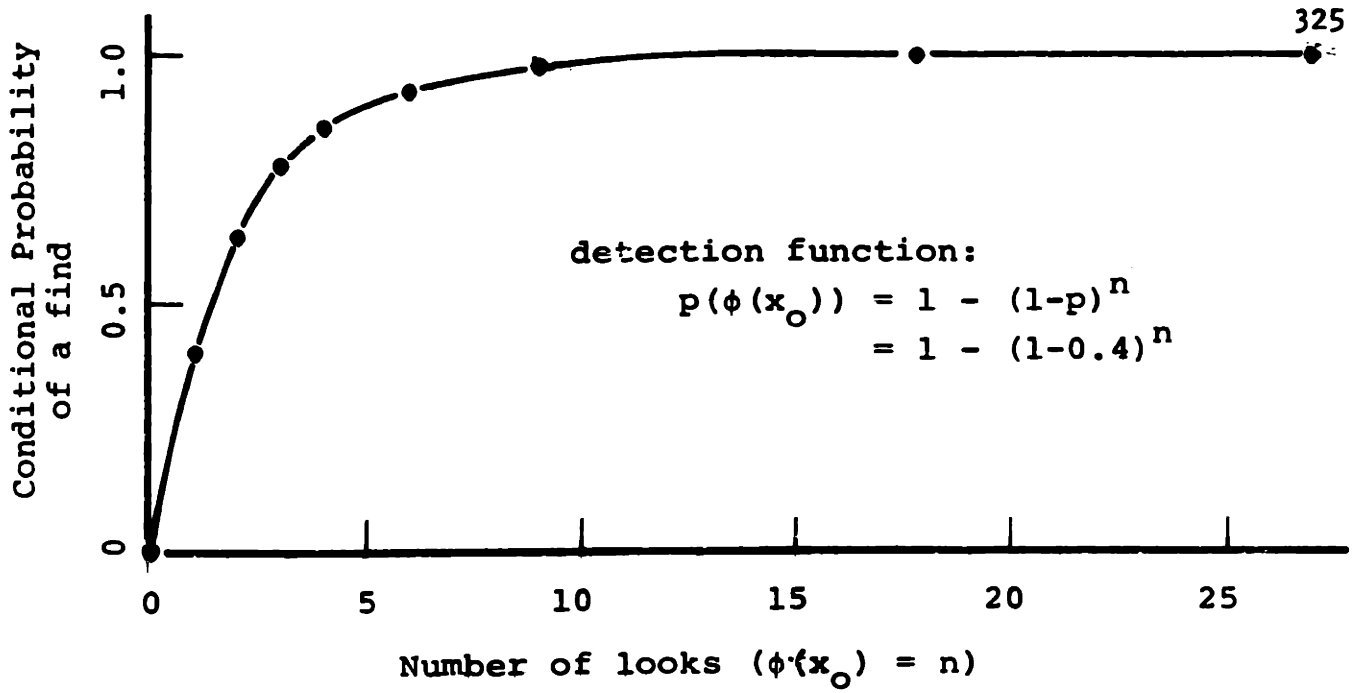


Figure 48

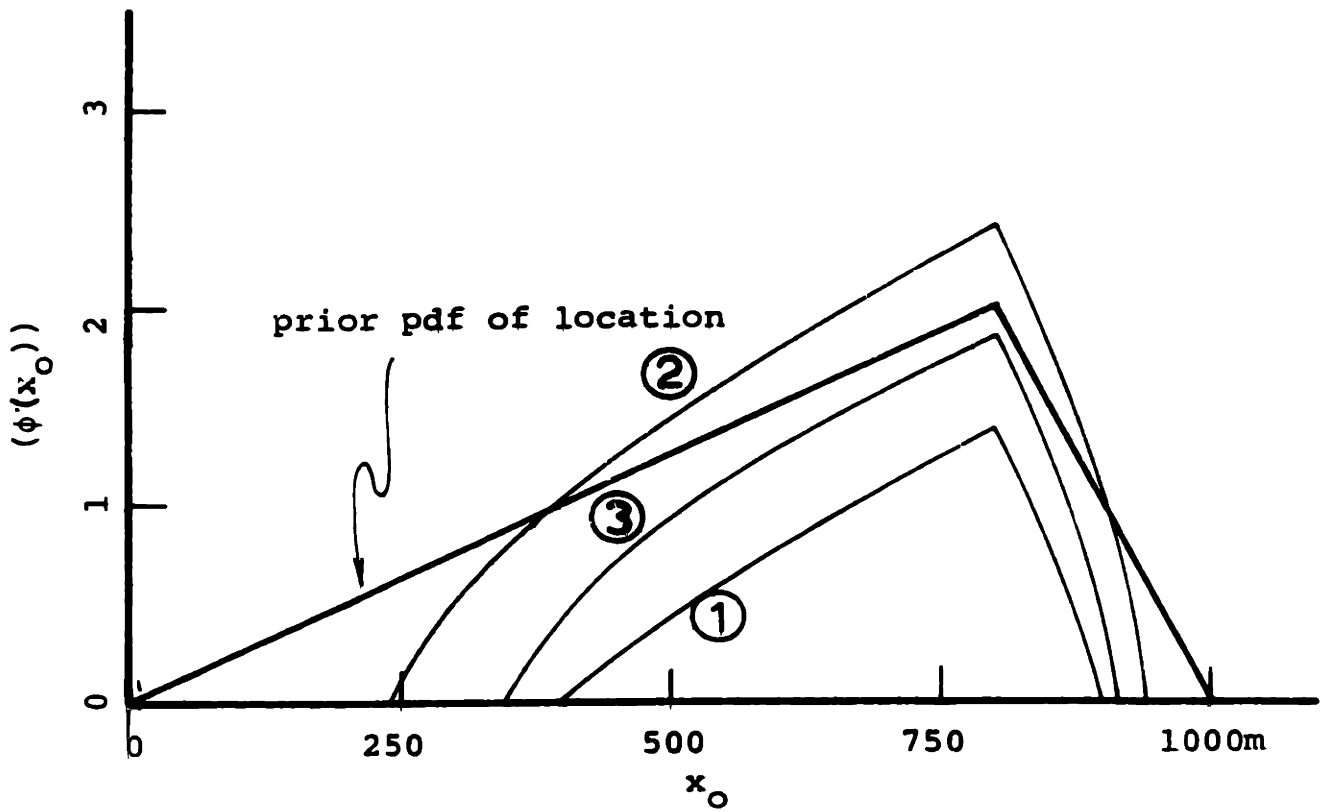


Figure 49

Figure 50

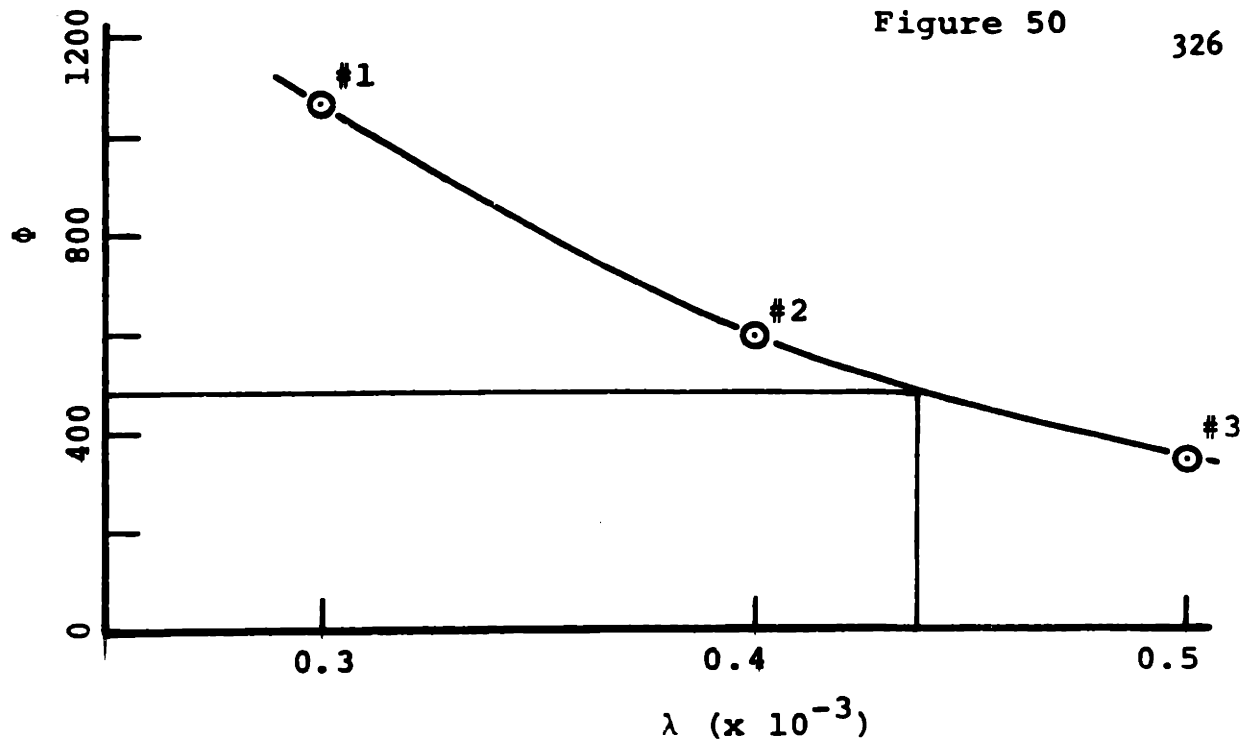
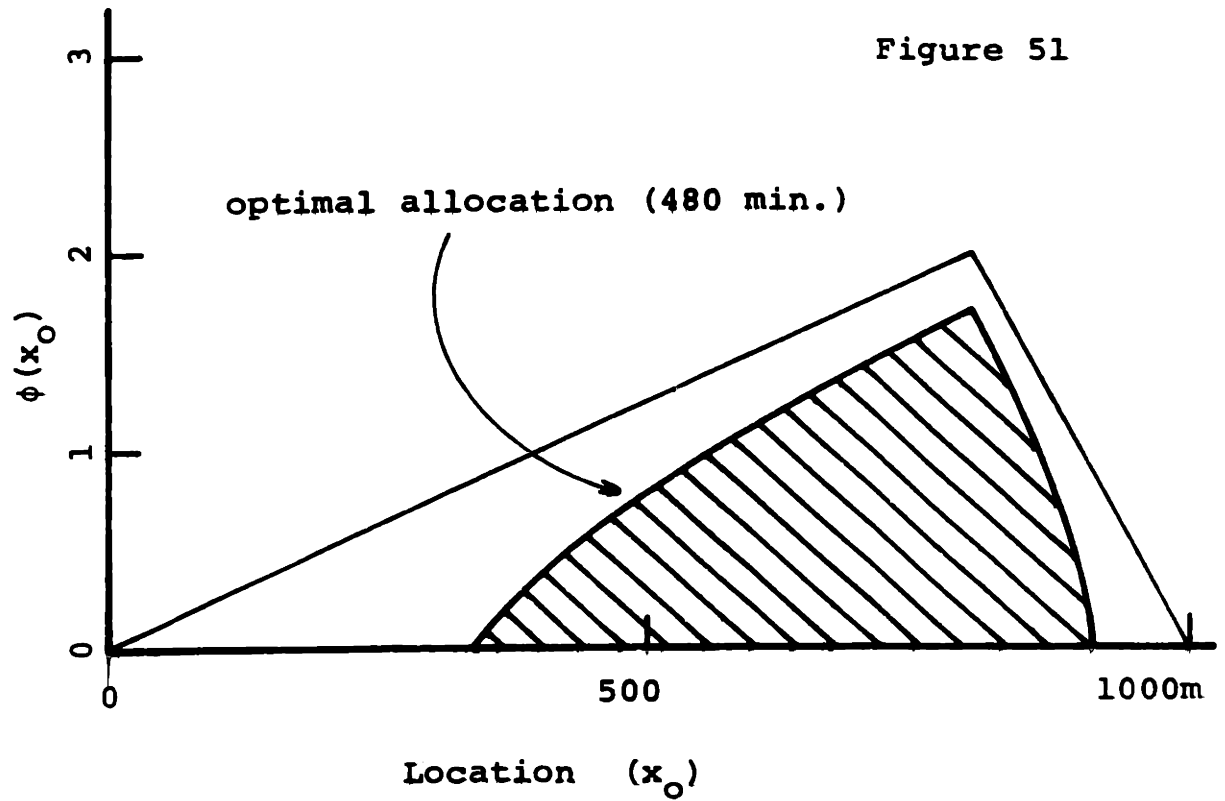


Figure 51



### Multiple Target Types

In some cases more than one target type is of interest. For example, at one site both faulting and weathered zones (unrelated to faulting) may be of concern. There are utilities associated with finding each type target and these utilities are usually not the same. Furthermore, the detection functions for both targets are usually not the same. In such cases we seek an optimal allocation  $\varphi^*(x)$  which maximizes expected total utility of search.

Let the utility of finding a type 1 target be  $u_1$ , and a type 2 target,  $u_2$ . For simplicity we assume both detection functions to be exponential-saturation

$$p_1(\varphi(x)) = 1 - e^{-k_1 \varphi(x)} \quad 69a$$

$$p_2(\varphi(x)) = 1 - e^{-k_2 \varphi(x)} \quad 69b$$

The expected utility of an allocation  $\varphi(x)$  is

$$\begin{aligned} E[u | \varphi(x)] &= u_1 p(\text{finding type 1} | \varphi(x)) \\ &\quad + u_2 p(\text{finding type 2} | \varphi(x)) \\ &= u_1 \int_{-\infty}^{\infty} f_1(x) (1 - e^{-k_1 \varphi(x)}) dx + u_2 \int_{-\infty}^{\infty} f_2(x) (1 - e^{-k_2 \varphi(x)}) dx \end{aligned} \quad 70$$

which we wish to maximize.

The problem can be greatly simplified by assuming  $p(\varphi(x)) = p_1(\varphi(x)) = p_2(\varphi(x))$ . \* Equation 70 may then be rewritten as

$$E[u|\varphi(x)] = \int_{-\infty}^{\infty} [u_1 f_1(x) + u_2 f_2(x)] (1 - e^{-\varphi(x)}) dx \quad 71$$

which has the same form as that optimized by Koopman's construction.

---

\* By making this assumption we reduce the number of distributions to be dealt with from two to one, and may use optimization methods already developed. If we do not make this assumption, we must either find another way to reduce the number of integrals or solve the more difficult problem of maximizing the sum.

One assumption we could make to reduce the number of distributions is that  $k_1 \varphi(x)$  and  $k_2 \varphi(x)$  are small. Then

$$1 - e^{-k_1 \varphi(x)} \approx k_1 \varphi(x) \quad \text{for } k_1 \varphi(x) \ll 1$$

$$1 - e^{-k_2 \varphi(x)} \approx k_2 \varphi(x) \quad \text{for } k_2 \varphi(x) \ll 1$$

and 70 becomes

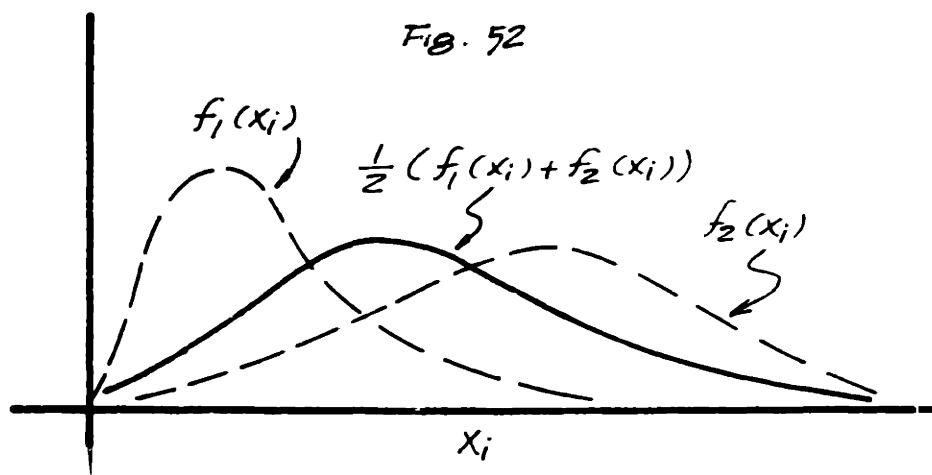
$$E[u|\varphi(x)] = \int_{-\infty}^{\infty} [k_1 u_1 f_1(x) + k_2 u_2 f_2(x)] \varphi(x) dx$$

The problem reduces to optimizing the allocation of search effort with a linear detection function over the weighted distribution  $(k_1 u_1 f_1(x) + k_2 u_2 f_2(x))$ , and has the somewhat trivial solution of allocating all of the available effort to the maximum of the weighted distribution.

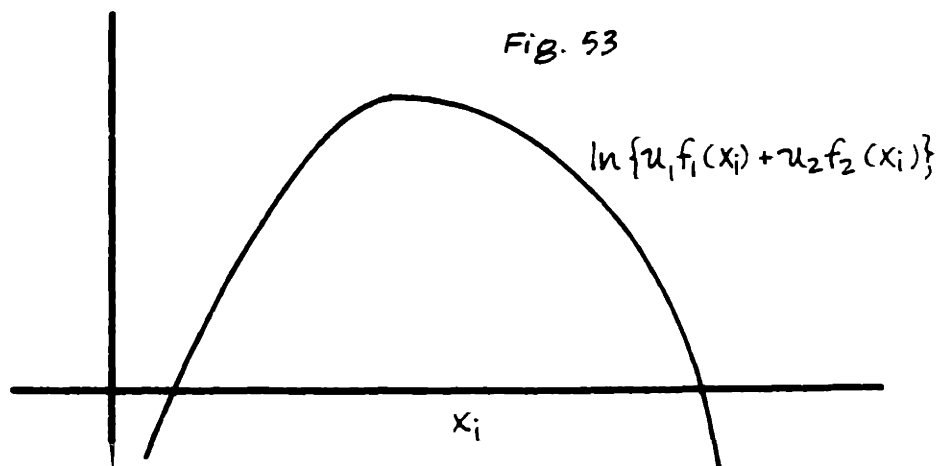


The optimal allocation function  $\varphi^*(x)$  may be determined by a construction analogous to Koopman's.

STEP I: Graph the weighted sum of the prior location pdf's of the two targets, and normalize the area under the curve to 1.0.



STEP II: Graph the logarithm of the normalized sum of Step I.



STEP III: Continue as the basic Koopman construction to find  $\varphi^*(x)$ .

The expected utility of an allocation  $\varphi^*(x)$  can also be found graphically due to the analogous relationship to equation 55. The

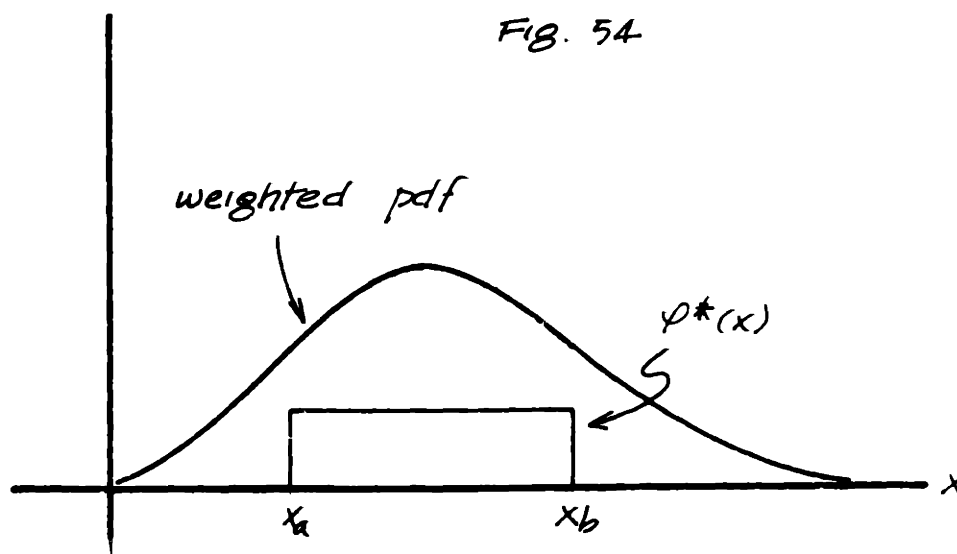
expected utility equals the area below the weighted distribution and above the horizontal line (from Step III) when plotted to an arithmetic ordinate (as in Step I).

The probabilities of finding a type 1 or a type 2 target cannot be determined by a simple graphical construction, but they can be evaluated analytically.

$$p(\text{find, type 1}) = \int_{-\infty}^{\infty} f_1(x_i) p(\psi(x_i)) dx_i \quad 72$$

$$p(\text{find, type 2}) = \int_{-\infty}^{\infty} f_2(x_i) p(\psi(x_i)) dx_i \quad 73$$

The extension of the technique to higher dimensions follows as the extension of Koopman's basic procedure.



Linear detection functions are handled in an analogous way. From the weighted sum of Step I, allocate effort in sufficient density for  $p(\psi(x))$  to just equal 1.0, to the modal region of the weighted

distribution (Figure 54). The expected utility of allocation equals the area underneath the weighted distribution between  $x_a$  and  $x_b$ . The probabilities of finding a type 1 and a type 2 equal the probabilities that each lies between  $x_a$  and  $x_b$  (Figure 55) or

$$P_1 = \int_{x_a}^{x_b} f_1(x) dx$$

$$P_2 = \int_{x_a}^{x_b} f_2(x) dx$$

74

Fig. 55a

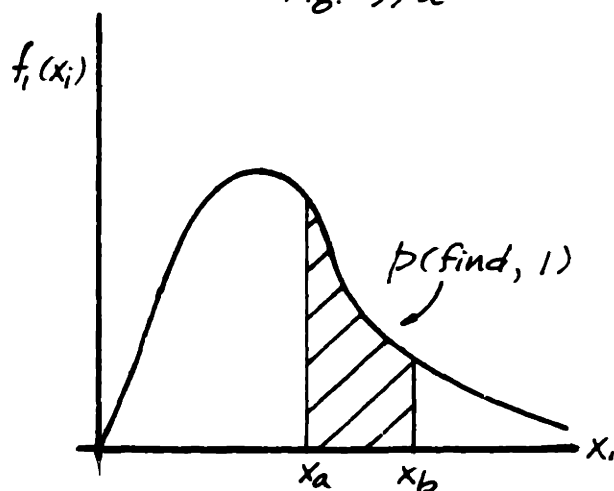
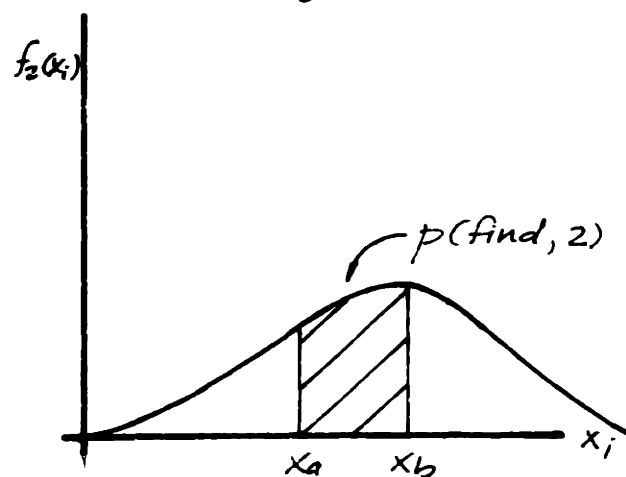


Fig. 55b



## Section 6.2 Two-Stage Search

Site exploration sometimes makes use of inexpensive, imperfect search techniques to assess the likely locations of geological targets through a preliminary search. These techniques indicate where targets are likely to be, although final verification of the target's existence must be made through other, more expensive methods. Since these first-stage techniques are imperfect, they are subject to spurious signals --

or noise -- which indicate high target likelihoods where actually no targets exist and low likelihoods where targets do exist. The preliminary methods also cannot distinguish between the targets of concern and other geological formations with similar geophysical properties.

An example of a two-stage search is the exploration for serpentine bodies along a proposed tunnel route. Due to high magnetite content, serpentine bodies display magnetic anomalies which can be mapped by geophysical surveys. There are other sources of magnetic anomalies besides serpentine, of course, so the magnetic survey cannot locate the serpentine bodies positively. A second stage of search, using more expensive techniques (e.g., borings), must be undertaken to ascertain the character of the anomalous areas and positively locate any serpentine which does exist. The problem we face is to allocate exploration effort over the three parameters

s: spacing between first stage traverses

n: the number of passes in the first stage

$\Delta^*$ : the magnitude of anomaly to be search in the second stage

such that expected utility is maximized (we will assume throughout that disutility is a linear function of cost).

There are four components of cost in two-stage search.\* The

---

\*Engle (1957) has considered a two-stage allocation of search effort in mineral explorations which maximizes the expected utility over the parameters  $\bar{m}$  and  $\bar{n}$ . He assumes that data from the first-stage search is in the form of point impulses or contacts, and that during the second stage locations are searched which exhibited greater than  $\bar{m}$  contacts within a given radius. The present formulation differs from

search effort cost of the first stage, the search effort cost of the second stage, the cost associated with missing a target, and the gain (or negative cost) associated with finding a target. As a first approximation, the cost of first-stage search is dependent on the technique selected and the total length of traverse. We will not consider effects of terrain, mobilization, and the like; although in a more refined model they might be. Again, as a first approximation, the cost of second stage searching is a function of the number of anomalous locations searched. That is, one quantum of second stage search effort is allocated to each location for which the anomaly  $\Delta \geq \Delta^*$ . The utilities of finding and of missing a target depend on the particular situation. In searching for an economic deposit there is positive utility (or gain) from a find, but zero utility from a miss. In searching for a geological detail which might hinder construction, there is zero utility from a find, but negative utility (disutility or loss) from a miss. In general, only one of the two utilities will be non-zero in a given problem.

The total expected utility of a two stage search, then, is

$$\begin{aligned} \text{utility} &= E[\text{No. of finds}] u_{\text{find}} \\ &+ E[\text{No. of misses}] u_{\text{miss}} \\ &- \text{1st stage cost} \\ &- \text{2nd stage cost} \end{aligned}$$

75

---

Engle's in that it considers the parameter  $s$ , (grid spacing) and postulates a different model for the first stage data (in Engle's work a 1st stage covering consists of "passing through the area in such a manner that every point of the area comes under surveillance...exactly once").

If we assume that targets are independent and that overlap of targets in the same plane may be neglected

$$E [\text{No. finds}] = m_r \text{Pr} [\text{find}] \quad 76$$

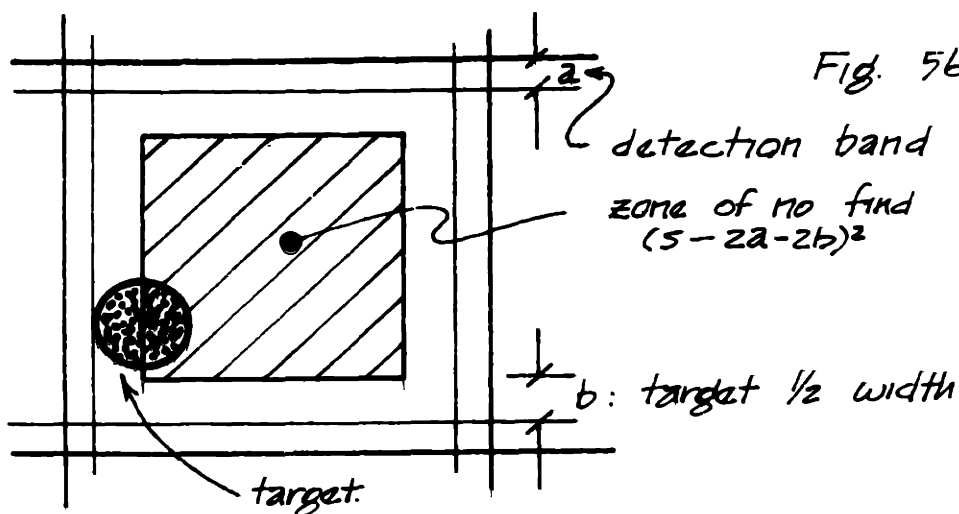
Since we don't know the number of targets at the site we must assume a pmf of  $m_r$  and take the expected value of the right hand side of equation (76).

$$E [\text{No. finds}] = \sum_{m_r=0}^{\infty} m_r p_{m_r}(m_{r0}) \text{Pr} [\text{find}] dm_{r0} \quad 77$$

By similar argument, and noting that  $\text{Pr}(\text{miss}) = 1 - \text{Pr}(\text{find})$ , the expected number of misses is

$$E [\text{No. misses}] = \sum_{m_r=0}^{\infty} m_r p_{m_r}(m_{r0}) (1 - \text{Pr} [\text{find}]) \quad 78$$

Assuming a square line grid of spacing  $s$  (Section 2.2), and a detection half-width of  $a$  (i.e., the target has a possibility of being seen if it lies within  $\pm a$  of the grid line), the probability of locating a circular target of half-width  $b$  is (Figure 56).



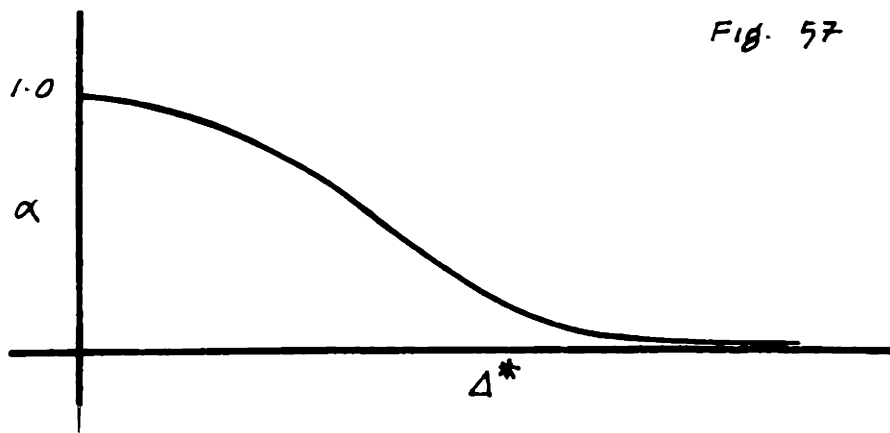
Pr { find | s,b,a,n }

$$= \left[ 1 - \left( \frac{(s-2a-zb)^2}{s^2} \right) \right] (1 - (1-\alpha)^n) + \frac{4a^2}{s^2} (1 - (1-\alpha)^{2n}), \text{ for } p \leq s-2a \quad 79$$

where  $n$  is the number of passes over the grid and  $\alpha$  is the conditional probability of a find, given a target existing within the detection band. Because the small areas which lie in two detection bands are searched twice, the conditional probability of a find in these areas is larger than in other areas. However, for  $a \ll s$  equation (79) becomes,

$$\text{Pr \{ find | s,b,a,n \}} = \left[ 1 - \left( \frac{s-2a-zb}{s} \right)^2 \right] (1 - (1-\alpha)^n), \text{ for } a \ll s, b \leq s-2a \quad 80$$

$\Delta^*$  is the magnitude above which an anomaly is searched in the second stage. The higher  $\Delta^*$ , the more chance of missing a target. Therefore, as  $\Delta^*$  is increased,  $\alpha$  is decreased and the relationship between  $\Delta^*$  and  $\alpha$  might be as shown in Figure 57 where  $\alpha$  approaches 1.0 for small  $\Delta^*$  and zero for large  $\Delta^*$ .



Substituting equation (80) into (77)

$$E [ \text{No. finds} ] \\ = \sum_{m_r=0}^{\infty} m_r p_{m_r}(m_{r0}) \left[ 1 - \left( \frac{s-2a-2b}{s} \right)^2 \right] (1-(1-\alpha)^n) \quad 81$$

Since the probability of a miss is the complement of the probability of a hit

$$E [ \text{No. miss} ] \\ = \sum_{m_r=0}^{\infty} m_r p_{m_r}(m_{r0}) \left[ 1 - \left( 1 - \left( \frac{s-2a-2b}{s} \right)^2 (1-(1-\alpha)^n) \right) \right] \quad 82$$

We have assumed that the cost of the first stage searching is dependent only on the total length of traverse (we neglect other costs to simplify the discussion),

$$\text{cost of } i^{\text{st}} \text{ stage} = c_1 L(n/s) \quad 83$$

where  $n$  is the number of passes,  $s$  is the grid spacing,  $L$  is a constant dependent on the site dimensions, and  $c_1$  is the unit cost.

The expected cost of second stage search is directly related to the expected number of locations which exhibit anomalies,

$$\text{cost of } 2^{\text{nd}} \text{ stage} = c_2 E(\text{no. } \Delta' \geq \Delta^* ) \quad 84$$

Anomalies arise from three sources: targets, real geological formations other than targets, and spurious signals or noise. The expected number of real targets displaying  $\Delta \geq \Delta^*$  has been derived in equation (81). The expected number of geological formations other than targets which display a  $\Delta \geq \Delta^*$  can be determined in a similar manner. If  $m_n$  is the number of non-target formations at the



site which are also anomalous with respect to the property measured in the first stage, and  $p_{m_n}(m_{no})$  is the pmf of  $m_n$ , then

$$E(\text{no. non-target finds}) = \sum_{m_n=0}^{\infty} m_n p_{m_n}(m_{no}) \left(1 - \left(\frac{s-2a-2b'}{s}\right)^2\right) (1 - (1-\alpha')^{m_n}) \quad 85$$

where  $b'$  and  $\alpha'$  are the target half-dimension and detection probability, respectively.

We will assume the magnitude of the maximum spurious signal within a unit length of traverse to be standardized Gamma distributed, and that when more than one traverse is made  $\Delta$  is taken to be the average of the maximum signal from each traverse.\* The distribution of  $\Delta$  for one traverse is,

$$f_{\Delta}(\Delta|r) = \frac{e^{-\Delta} \Delta^{r-1}}{\Gamma(r)} \quad 86$$

where  $r$  is the expected value of  $\Delta$ , and  $\Gamma(r)$  is the gamma function (for integer values of  $r$ ,  $\Gamma(r) = (r-1)!$ ). An advantage of the Gamma distribution is that the sum of  $n$  Gamma distributed random variables is also Gamma distributed, with parameter  $nr$ .

That is,

$$f_{\Sigma\Delta}(\sum_i \Delta_i|r) = \frac{e^{-\Delta} (\sum\Delta_i)^{nr-1}}{\Gamma(nr)} \quad 87$$

---

\*We choose the standardized Gamma distribution only for convenience. In actual applications the form of this distribution must be empirically determined.

Applying a change of variable transformation to obtain the pdf of  $1/n$

$$f_{\frac{1}{n} \sum \Delta_i} (\Delta_o | r) = \frac{e^{-n\Delta_o} (n\Delta_o)^{nr-1}}{\Gamma(nr)} \quad 88$$

Assuming the maximum  $\Delta$ 's to be independent

$$E [\text{No. random anomalies } \Delta \geq \Delta^*]$$

$$= \int_0^{\Delta^*} \frac{e^{-n\Delta_o} (n\Delta_o)^{nr-1}}{\Gamma(nr)} d\Delta_o \quad 89$$

Substituting into equation 75 gives an expression for total expected utility,

$$\begin{aligned} E[\text{utility}] &= u_{\text{find}} \sum_{m_r=0}^{\infty} m_r P_{m_r}(m_{ro}) \left[ 1 - \left( \frac{s-2a-2b}{s} \right)^2 \right] (1-(1-\alpha)^n) \\ &+ u_{\text{miss}} \sum_{m_r=0}^{\infty} m_r P_{m_r}(m_{ro}) \left[ 1 - \left[ 1 - \left( \frac{s-2a-2b}{s} \right)^2 \right] (1-(1-\alpha)^n) \right] \\ &- c_1 L(n/s) \\ &- c_2 \sum_{m_r=0}^{\infty} m_r P_{m_r}(m_{ro}) \left[ 1 - \left( \frac{s-2a-2b'}{s} \right)^2 \right] (1-(1-\alpha')^n) \\ &- c_2 \sum_{m_r=0}^{\infty} m_r P_{m_r}(m_{ro}) \left[ 1 - \left( \frac{s-2a-2b}{s} \right)^2 \right] (1-(1-\alpha)^n) \\ &- c_2 \int_0^{\Delta^*} \frac{e^{-n\Delta} (n\Delta)^{nr-1}}{\Gamma(nr)} d\Delta \end{aligned}$$

**Example: Multi-stage search optimization**

site dimensions	1500 x 1500
target half-width	30
first stage detection band half-width	10
number of real targets	
Poisson distributed	lamda = 10
number of false targets	
Poisson distributed	lamda = 10
conditional probability of first stage find	
real targets	alpha = 0.01*delta
false targets	alpha = 0.01*delta
cost of first stage search	\$15.00/m
cost of second stage search/unit	1.500
utility of finding a target	0.000
utility of missing a target	-200.00

all lengths in meters

all costs in thousands of dollars (except as noted)

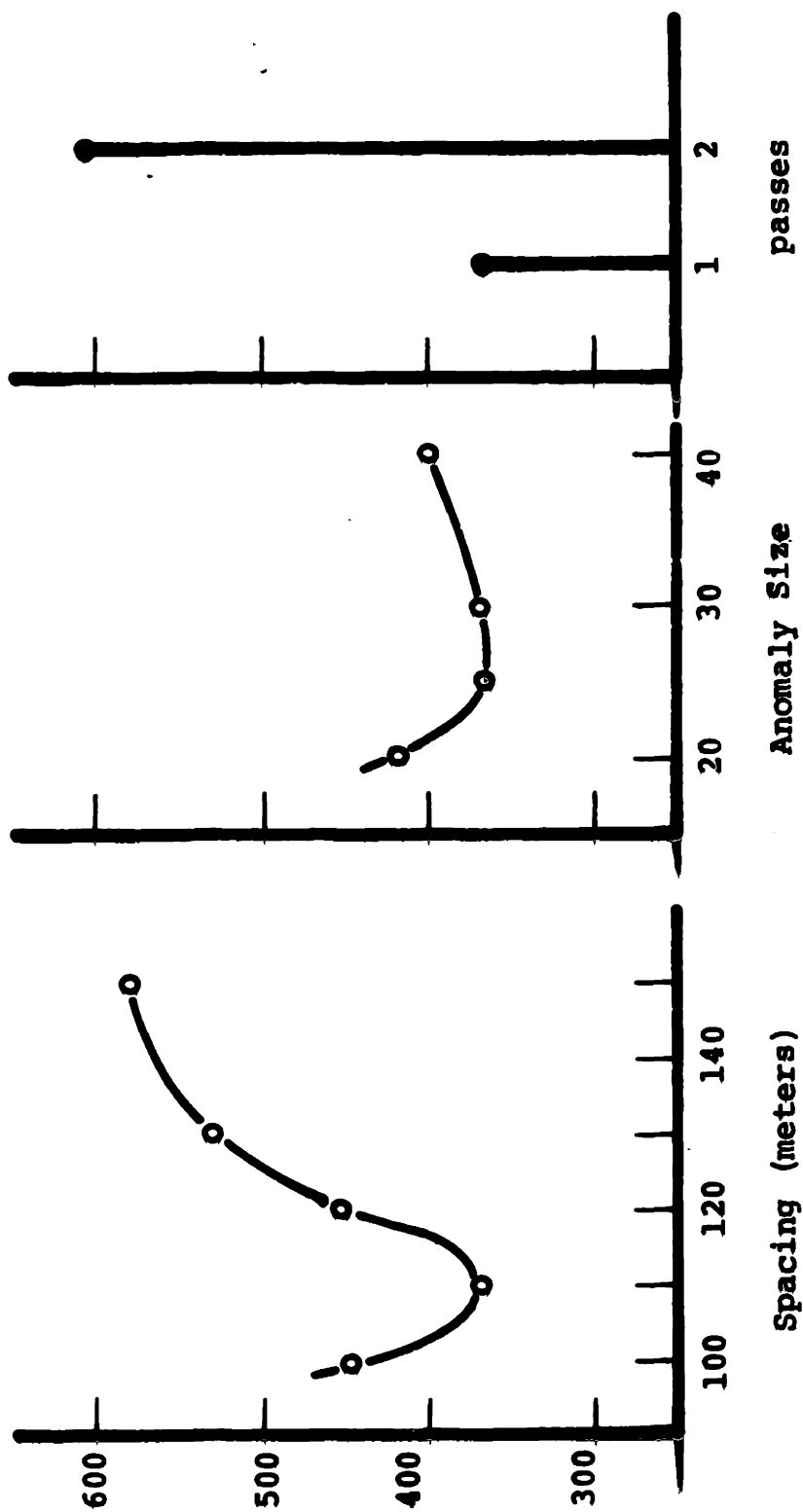


Figure 58  
 Optimization of two stage search over spacing, number of first stage  
 passes, and anomaly size searched

where  $c_1$  and  $c_2$  are the unit costs of first and second stage search, respectively. The optimal values of the parameters  $(s, n, \Delta^*)$  maximize equation 90 (i.e., give the maximum expected utility).

The models we've assumed in this example are simplistic, but they illustrate the process of maximizing expected utility. Obviously, the validity of the parametric optimum is directly dependent upon the validity of the functional relationships which compose equation (90).

The cost relationships as a first approximation are probably satisfactory. The two relationships which may be difficult to specify with sufficient validity until empirical data are evaluated (for a specific search tool) are the correspondence between detection probability and anomaly size and, less importantly, the frequency of random readings.

### Section 6.3 Quantity of Total Effort

We have so far considered optimizing search allocations for a given total effort. In this chapter we extend our considerations to the quantity of total effort expended.\*

If too little effort is expended the probability of an undetected target existing is high and the associated risk manifests itself in greater design and construction costs. If too much effort is expended its marginal value is small and does not decrease design and construction

---

\*This section treats only the single-stage case. The sequential case is treated under "optimal stopping."

cost enough to be justified. Figure 59 shows a decision tree for the total effort decision when a single target is searched for. A box indicates a decision fork, a circle indicates a chance fork, and a gate (i.e., a slash across a branch) indicates a fixed cost. The curved arrow across the effort decision indicates that  $\Phi$  is a continuous variable.

The three utilities of interest are

$u_1$ : The utility of finding a target.

$u_2$ : The utility of not finding a target which exists.

$u_3$ : The utility of not finding a target and no target existing.

The utilities depend on the cost of effort allocation and therefore on  $\Phi$ , but in addition,  $u_2$  and  $u_3$  also depend on design costs associated with the probability of an undetected target existing (which itself is a function of  $\Phi$ ). In application, these utilities must be ascertained from the cost of remedial work if a target exists and is not designed for, and additional construction costs when a conservative design is adopted because of uncertainty.

The expected utility of a branch equals the weighted sum of its possible terminal utilities multiplied by their respective probabilities. Assuming utilities to be additive, we can factor out  $c_i$  to obtain,

$$E[u | \Phi = i] = c_i + \beta [p_i u_1 + (1-p_i) u_2(c_i)] + (1-\beta) u_3(c_i) \quad 91$$

where  $c_i$  is the fixed cost of search,  $\beta$  is the prior probability of the target existing, and  $p_i$  is the conditional probability of a find when  $\Phi = i$ .

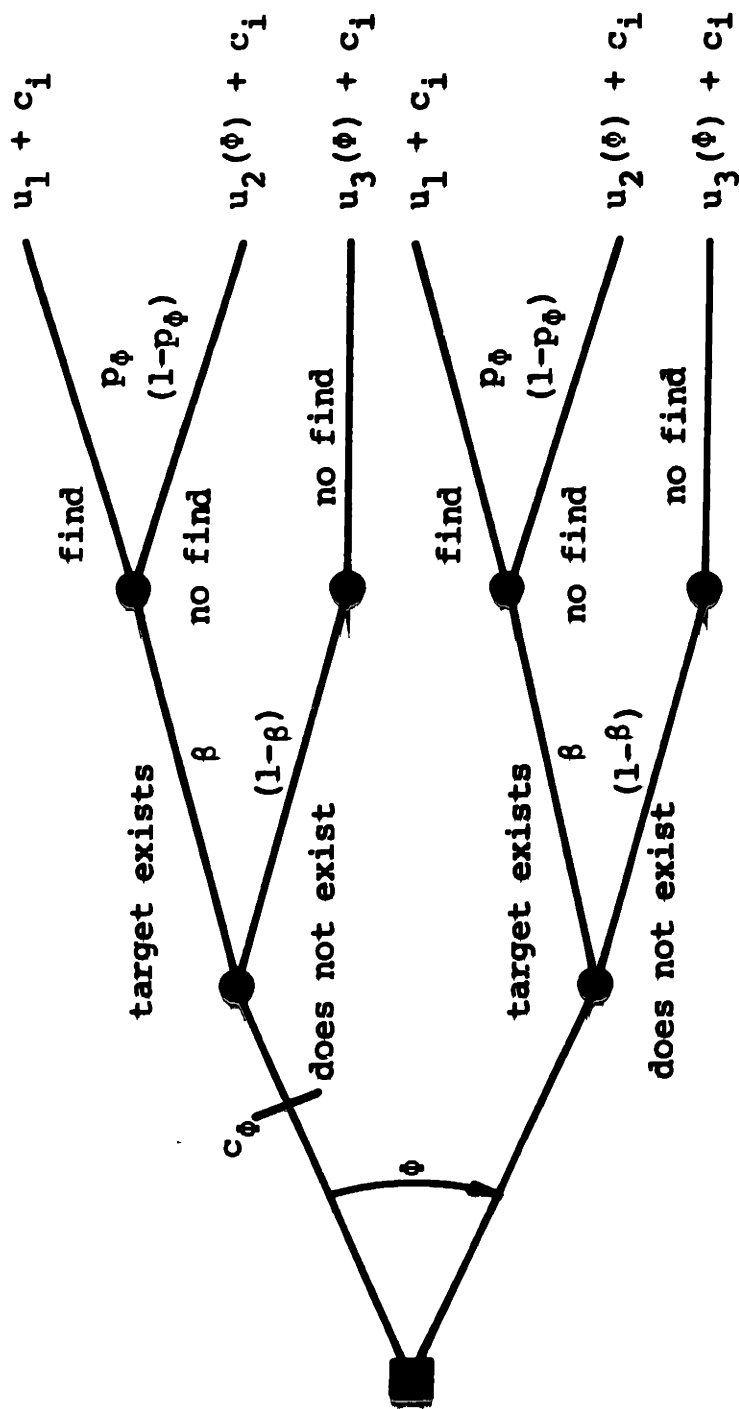


Figure 59: Decision tree for decision on total effort expended in searching for a single target ( $\phi$  = total effort)

Example:

(Grid search with uniform priors)

A large complex of structures is to be built at a site composed of residual soils. Should a zone of deep weathering exist beneath part of the complex, differential settlements may result which would require extensive remedial work. The engineer would like to know of the existence of such a zone beforehand so that he can design the structures to compensate for it (perhaps by relocating).

The utility to the engineer of finding a deeply weathered zone is zero since the structures will be moved to other locations. The expected cost of remedial work should a weathered zone go unfound has been estimated to be  $5 \times 10^5$  dollars.

Through a separate decision analysis the engineer has decided that should the probability of an undetected weathered zone be greater than 0.05 he must increase his in-situ instrumentation for during and post construction performance monitoring (this assumes that optimal instrumentation cost depends on search effort only through the probability of an undetected target). He has estimated the present worth of these activities to be as shown in Figure 60.



Fig. 60



Therefore,

$$u_1 = 0$$

$$u_2 = (4 \times 10^5)(\gamma - 0.05) + \gamma(5 \times 10^5)$$

$$u_3 = 4 \times 10^5 (\gamma - 0.05)$$

where  $\gamma$  is the probability of an undetected target.

The prior probability of location is uniform over the site, so the engineer has chosen to use a grid pattern of geophysical traverses to search for weathered zones. He would like to make a decision on the spacing at which to place the grid lines, if his prior probability of the target existing is 0.3.

As a first approximation, the cost of search will be taken proportional to the length of traverse

$$c_i = k/s$$

where  $k$  is a constant depending on site dimensions and unit traverse cost.

In Figures 61a, b, and 62 are shown the functional relationships between the parameters of the problem and grid spacing; in Figure 63 is shown the separate utilities and expected total cost. The spacing minimizing expected cost is about 50 meters.

Note that above  $s=200\text{m}$  the expected total cost is essentially independent of grid spacing.

Figure 61a

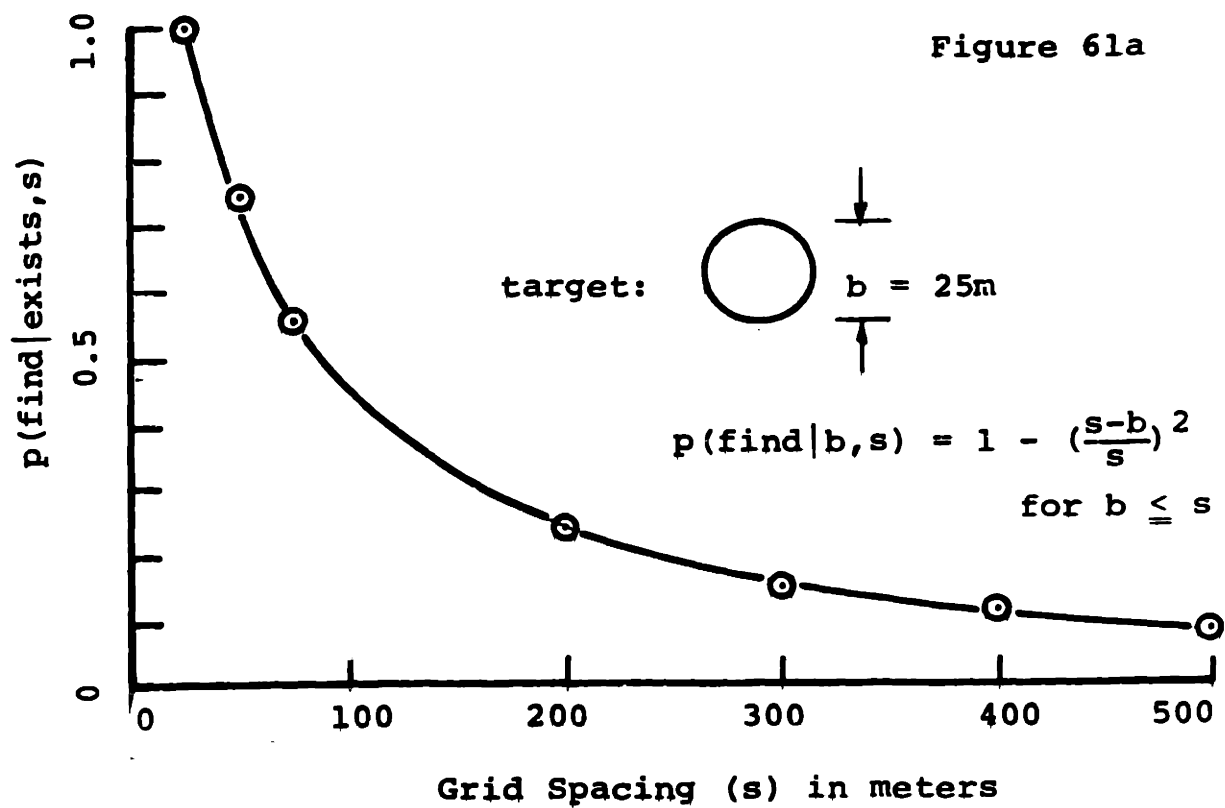
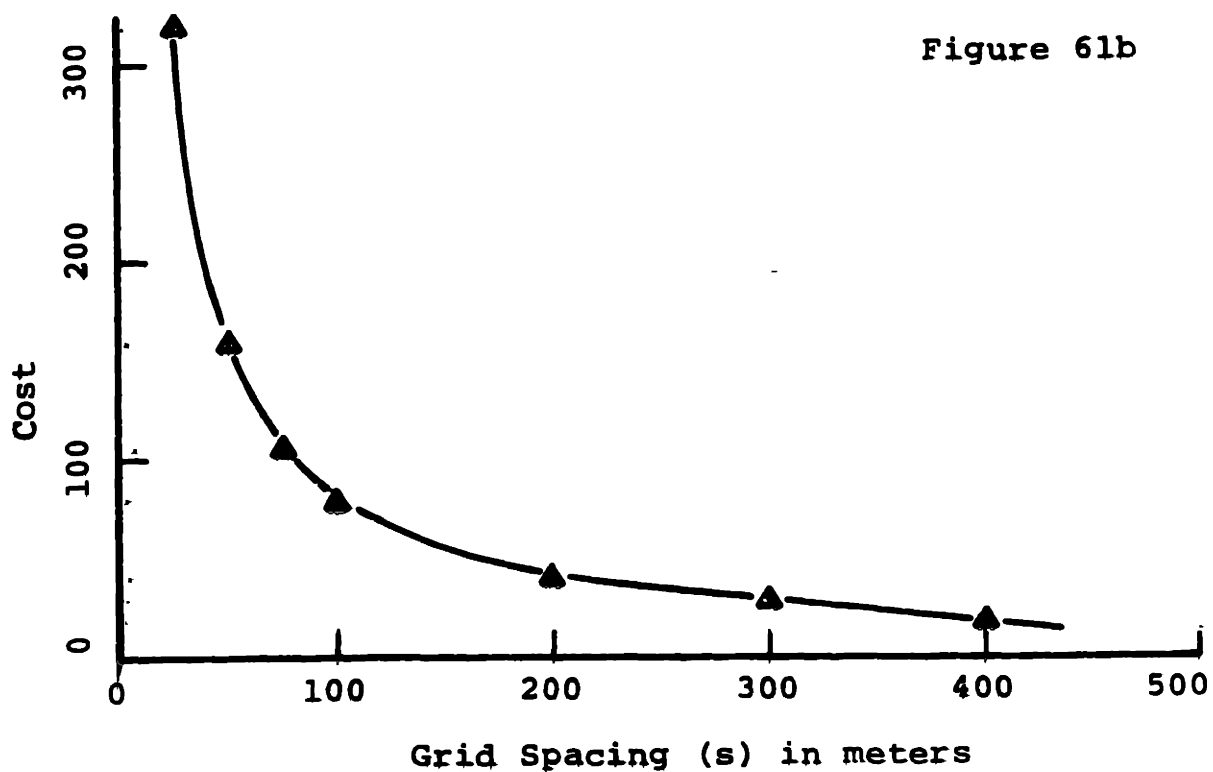


Figure 61b



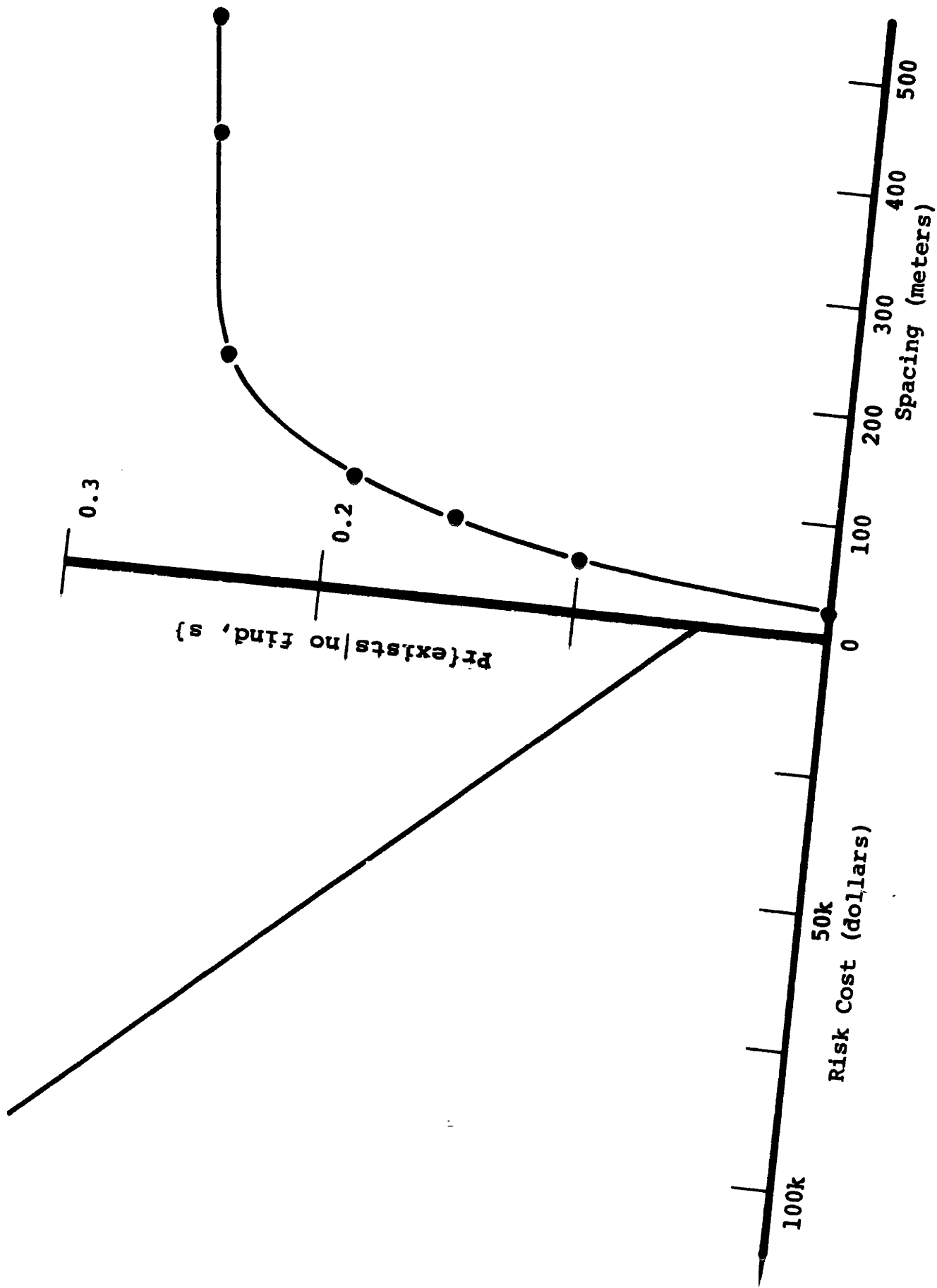


Figure 62

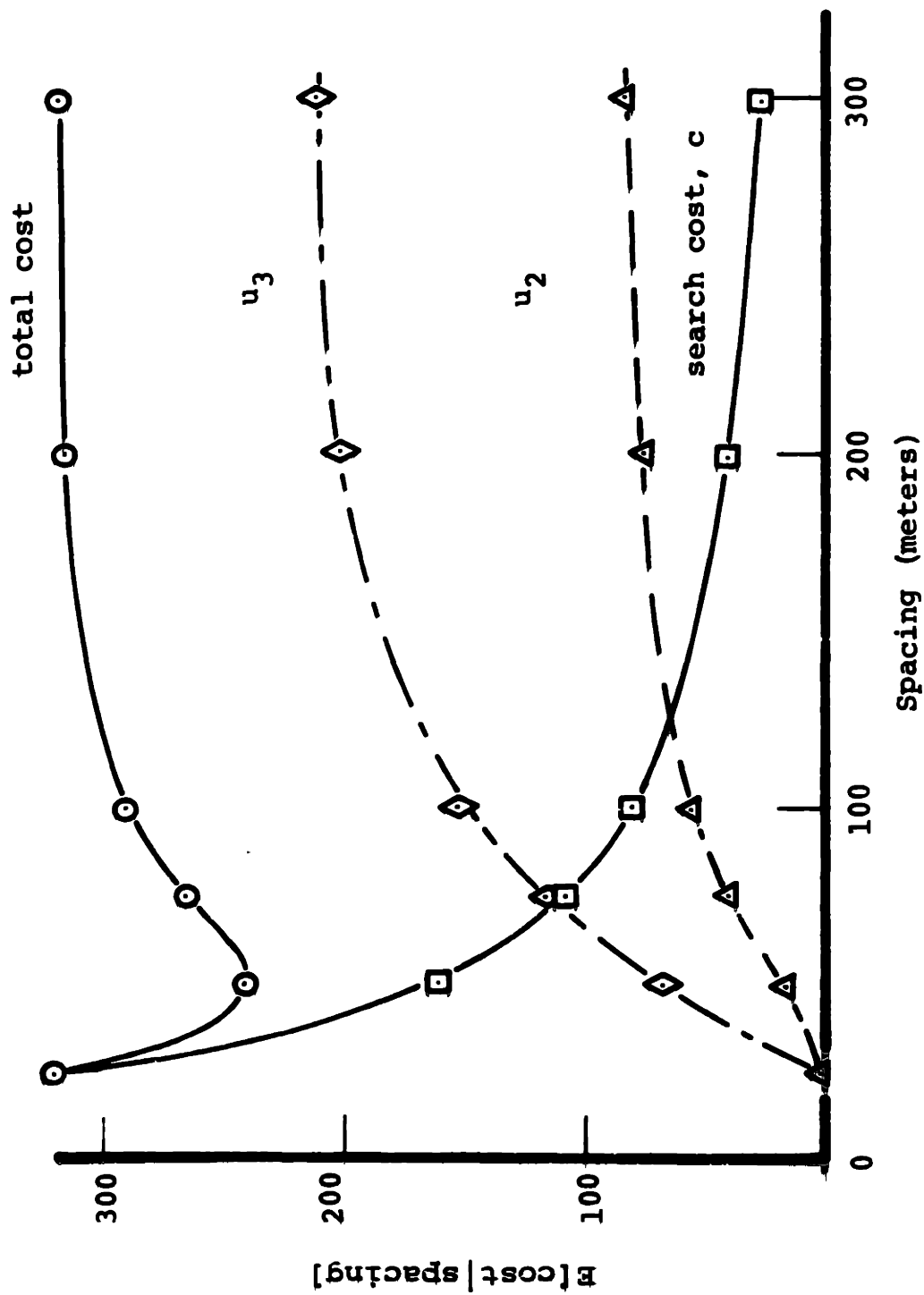


Figure 63

## Section 6.4 Sequential Search

A sequential search procedure is one in which the pdf of location is made current and a new optimal strategy determined after each individual quantum of effort is allocated, as opposed to single stage procedures in which all of the effort is expended before updating. Since each allocation is made on the basis of all past information, sequential procedures frequently produce a greater probability of a find than do single-stage procedures. However, the increased probability of a find may not be sufficient to compensate for discontinuities in field work and for the increased effort required by constant updating.

We will formulate the problem of sequential search in a manner similar to the single and multi-stage problems we have already considered. We are searching for a target which is thought to exist somewhere along a line (extensions to higher dimensions will be considered later), and the prior pdf of the target's location will be denoted by  $f(x)$ .

We break from the single-stage formulation here in that we specify that discrete quantities of effort are expended in searching. These discrete quantities may be located at any point along the continuous  $x$  axis within the region of search. We once again assume that the search technique is not perfectly efficient, that there exists a positive conditional probability of the target being missed in a search of location  $x_i$ , even though it exists at  $x_i$ . This conditional probability of a miss will be denoted  $\alpha_i$ .

$$\alpha_i = \Pr \{ \text{miss @ } x_i \mid \text{target exists @ } x_i \}$$

We now seek a "strategy" of search -- a rule telling us where to look at each stage -- which is in some sense optimal. To judge the "goodness" of each strategy we will consider four criteria which are commonly of interest in site exploration:

1. Minimizing the expected amount of effort required to find the target.
2. Maximizing the probability of finding the target with a given amount of effort.
3. Minimizing the conditional probability that the target exists somewhere at the site when no find is made.
4. Minimizing the expected amount of effort required to lower the conditional probability that the target exists to a specified level.

We will demonstrate our finding that each of these criteria leads to the same optimal strategy, that is, a strategy which optimizes any one criterion also optimizes the others. The importance of this property is that a unique optimum strategy exists.

#### Section 6.4.1 General Solution to Sequential Strategy

The subject of optimal sequential decisions forms a major component of the theory of statistical decisions. The basic functional equations for the problem of sequential search are difficult to solve directly (DeGroot, 1970), but for several cases the optimal strategy can be found by other analytical means (Chew, 1967; Ellis and Blackwell, 1959).

We will first consider optimality in the sense of criterion 1, minimizing the expected effort to find a target.

For any strategy  $\delta$  let the cost of the search at the  $n$ th stage be  $C_n$  and the probability that the target is found on the  $n$ th stage be  $\lambda_n$ . We assume that the cost of searching location  $x_i$  is  $c_i$ . From equation (92) the probability of finding the target at location  $x_i$  is

$$f_x(x_i) (1 - \alpha_i) \quad 93$$

The expected cost of finding the target using strategy  $\delta$  is

$$\begin{aligned} E[C(\delta)] &= \sum_{n=1}^{\infty} C_n \Pr\{N \geq n\} \quad 94 \\ &= C_1 + C_2(1 - \lambda_1) + C_3(1 - \lambda_1 - \lambda_2) + \dots \quad 95 \end{aligned}$$

where  $N$  is a random variable equalling the number of searches until a find is made. For the present we shall only consider prior distributions for which  $\int f_x(x_i) dx_i = 1.0$ , since  $E(C) = \infty$  when there is a positive probability of continuing the search indefinitely. However, we will return to this problem. We can show that the strategy  $\delta^*$  which minimizes  $E(C(\delta))$  must satisfy the relation (DeGroot, 1970).

$$\frac{\lambda_1}{c_1} \geq \frac{\lambda_2}{c_2} \geq \dots \quad 96$$

Suppose  $\delta$  is a strategy for which  $E(C(\delta)) < \infty$ , and for which

$$\frac{\lambda_n}{c_n} < \frac{\lambda_{n+1}}{c_{n+1}}$$

for some value of  $n$ . Let  $\delta^*$  be the strategy which is similar to at every stage, except that stages  $n$  and  $n+1$  are executed in reverse order (i.e., the location searched in stage  $n+1$  of  $\delta$  is searched before stage  $n$  in  $\delta^*$ ). In other words,  $\delta^*$  is the strategy such that

$$\left\{ \begin{array}{l} C_n^* = C_{n+1} \\ \lambda_n^* = \lambda_{n+1} \end{array} \right\}$$

$$\left\{ \begin{array}{l} C_{n+1}^* = C_n \\ \lambda_{n+1}^* = \lambda_n \end{array} \right\}$$

$$\left\{ \begin{array}{l} C_m^* = C_m \\ \lambda_m^* = \lambda_m \end{array} \right\} \quad \begin{array}{l} \text{for all } m \neq n \\ m \neq n+1 \end{array}$$

From equations 93, 94 and 95 we know

$$\begin{aligned} E[C(\delta)] - E[C(\delta^*)] &= C_n(1 - \lambda_1 - \dots - \lambda_{n-1}) \\ &\quad + C_{n+1}(1 - \lambda_1 - \dots - \lambda_n) \\ &\quad - C_{n+1}(1 - \lambda_1 - \dots - \lambda_{n-1}) \\ &\quad - C_n(1 - \lambda_1 - \dots - \lambda_{n-1}) \\ &= C_n \lambda_{n+1} - C_{n+1} \lambda_n > 0 \end{aligned}$$

Therefore, the expected total cost of  $\delta^*$  is lower than that of  $\delta$ .

Since this relation is valid for all  $n$ , the expected total cost of any strategy  $\delta$  not satisfying relationship 96 can be reduced.

Therefore, the optimal strategy  $\delta^*$  must conform to 96. If at any stage two or more values of  $f(x_i) (1 - \alpha_i) / c_i$  are equal, the ordering of these locations may be made arbitrarily.



The optimal strategy, therefore, specifies that at any stage in the search, the next location to be searched is the one that maximizes the probability of a find per unit cost.\* In the case when all of the  $c_i$ 's are equal (as is frequent in site investigation) the optimal strategy specifies searching the location with the highest probability of a find  $f_x(x_i)(1 - \alpha_i)$ . At any given stage,  $n, \dots$ ,  $f^{(n)}(x_i)$  is the current posterior pdf of location.

We are now in a position to show that the optimal strategy in the sense of criterion 1 also maximizes the probability of a find for a given amount of effort (criterion 2). The probability of a find in  $n$  searches is

$$p(\text{find}) = f(x_i)(1 - \alpha_i) + f^1(x_j)(1 - \alpha_j) \\ + f^{(2)}(x_k)(1 - \alpha_k) + \dots \\ + f^{(n)}(x_l)(1 - \alpha_l)$$

98

Using the optimal strategy of criterion 1, this equals

$$p(\text{find}) = \max_i f(x_i)(1 - \alpha_i) + \max_j f^1(x_j)(1 - \alpha_j) \\ + \dots + \max_l f^{(n)}(x_l)(1 - \alpha_l)$$

99

Therefore, the optimal strategy (in the sense of min.  $E(N)$ ) simultaneously maximizes, for all values of  $n$ , the probability of a find in  $n$  searches.\*\*

\*A sequential strategy which at each stage specifies the short-term best decision is said to be "myopic."

\*\*Furthermore, this strategy maximizes any non-decreasing function of the expected number of searches,  $g(E(N))$ .

We will now show that a strategy  $\delta^*$  which is optimal in the sense of criteria 1 and 2 is also optimal in the sense of criteria 3 and 4.

The strategy  $\delta^*$  is optimal in the sense of criterion 3 if it minimizes

$$\beta'_n = \Pr \{ \text{target exists} \mid \text{no find, } n \text{ searches} \}$$

Let  $\delta^*$  be a strategy which maximizes the probability of detection in  $n$  searches. If the prior probability of the target's existence is denoted by

$$\beta = \int_{-\infty}^{\infty} f_x(x_i) dx_i \quad 100$$

then by Bayes' Theorem

$$\beta' = p(\text{exists} \mid \text{no find, } n) = \frac{\beta p(\text{no find} \mid n, \text{exists})}{\beta p(\text{no find} \mid n, \text{exists}) + (1 - \beta) p(\text{no find} \mid n, \text{not exist})}$$

$$\frac{\beta [1 - p(\text{find}) \mid n, \text{exists}]}{\beta [1 - p(\text{find}) \mid n, \text{exists}] + (1 - \beta)}$$

101

Since  $\beta$  is a constant, any strategy which maximizes the term  $p(\text{find} \mid n, \text{exists})$  also minimizes  $\beta'$ .  $\delta^*$  is such a strategy, and is, therefore, optimal in the sense of criterion 3.

To be optimal in the sense of criterion 4 the strategy  $\delta^*$  must minimize the expected number of searches  $N$  required to lower the posterior probability of existence below a certain level. By Bayes' theorem, requiring the posterior probability  $\beta'$  to be below a certain

level is equivalent to requiring the quantity

$$[1 - p(\text{find} \mid n, \text{exists})]$$

to be below a certain value.

$$[1 - p(\text{find} \mid n, \text{exists})] \leq \xi \quad 102$$

or

$$p(\text{find} \mid n, \text{exists}) \geq 1 - \xi \quad 103$$

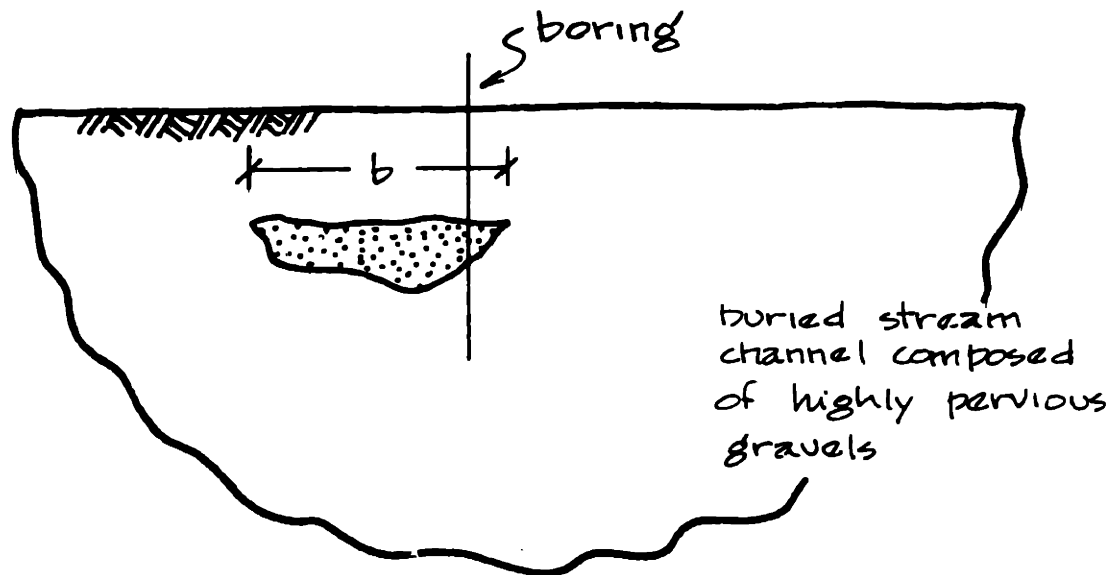
Hence a strategy  $\delta^*$  which maximizes  $p(\text{find} \mid n, \text{exist})$  also minimizes the probability that the random variable  $N$  required for inequality (103) to hold will be greater than  $n$ . In so doing, the strategy  $\delta^*$  minimizes the expected value of  $N$ .

So we have shown that a strategy which is optimal in the sense of criterion 1 is also optimal in the sense of criteria 2, 3, and 4; and, further, we have shown that this optimal strategy conforms to relation 96.

Example:

An extended embankment is to be constructed across a wide valley through which a buried stream channel composed of highly pervious material is thought to exist. The engineer's problem is to locate the buried channel, or lower his subjective probability of its existence to a certain low level.

A schematic cross-section of the stream channel is shown in Figure 64; in Figure 65 the engineer's subjective prior probability assessment of the channel's location.



The only tool at the disposal of the engineer is boring and evaluating the core logs. We will assume that the costs of boring at each location are equal, and that the probability of failing to identify the channel even though it has been hit by a boring is the same for each location.

$$c_i = c_j \quad \text{all } i, j$$

$$\alpha_i = \alpha_j \quad \text{all } i, j$$

The optimal strategy  $\delta^*$  specifies that the first boring should be placed at  $x_0$  since the quantity

$$\lambda_i / c_i = f(x_i) (1 - \alpha_i) / c_i$$

is maximized at that location. If a find is made, the search is terminated. If no find is made then the posterior pdf of location, given the channel exists, is

$$f'(x_i | \bar{x}_0) = \frac{f(x_i) L(\bar{x}_0 | x_i)}{\int f(x_i) L(\bar{x}_0 | x_i) dx_i}$$

where,

$$L(\bar{x}_0 | x_i) = \Pr \{ \text{no find @ } x_0 \mid \text{target @ } x_i \}$$

Letting  $d = |x_0 - x_i|$

In the present case,

$$L(\bar{x}_0 | x_i) = \begin{cases} \alpha_0 + (1 - \alpha_0)(d/b) & |d| \leq b \\ 1.0 & |d| \geq b \end{cases}$$

Therefore

$$\begin{aligned} \int_{-\infty}^{\infty} f(x_i) L(\bar{x}_0 | x_i) dx_i &= \int_{-a}^{x-b} f(x_i) (1.0) dx_i \\ &+ \int_{x-b}^{x+b} f(x_i) (\alpha_0 + (1 - \alpha_0)(d/b)) dx_i \\ &+ \int_{x+b}^{\infty} f(x_i) (1.0) dx_i \end{aligned}$$

and,

$$f'(x_i | \bar{x}_0) = \begin{cases} \frac{f(x_i) (1.0)}{\int_{-\infty}^{\infty} f(x_i) p(\bar{x}_0 | x_i) dx_i} & |d| \geq b \\ \frac{f(x_i) [\alpha_0 + (1 - \alpha_0)d/b]}{\int_{-\infty}^{\infty} f(x_i) p(\bar{x}_0 | x_i) dx_i} & |d| \leq b \end{cases}$$

Figure 65

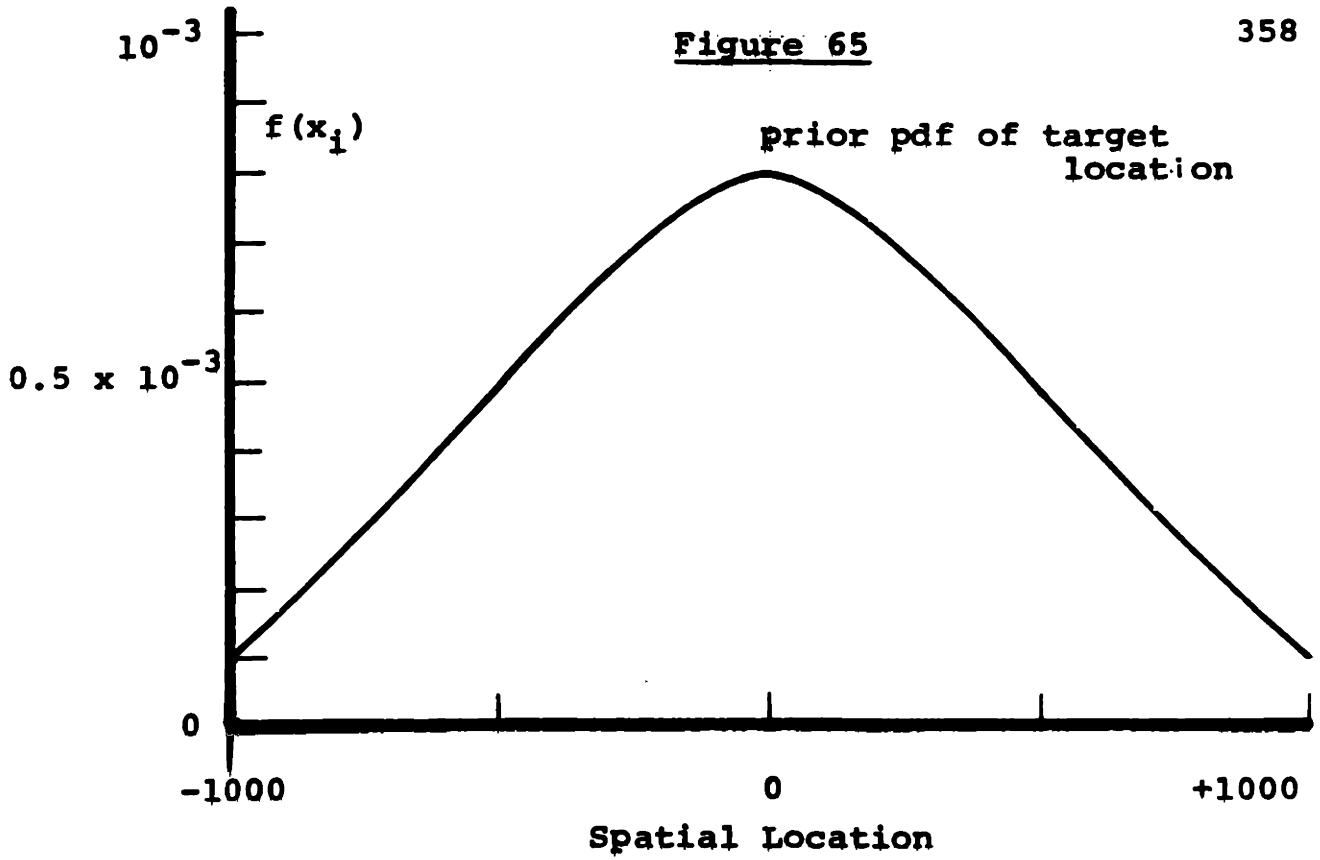
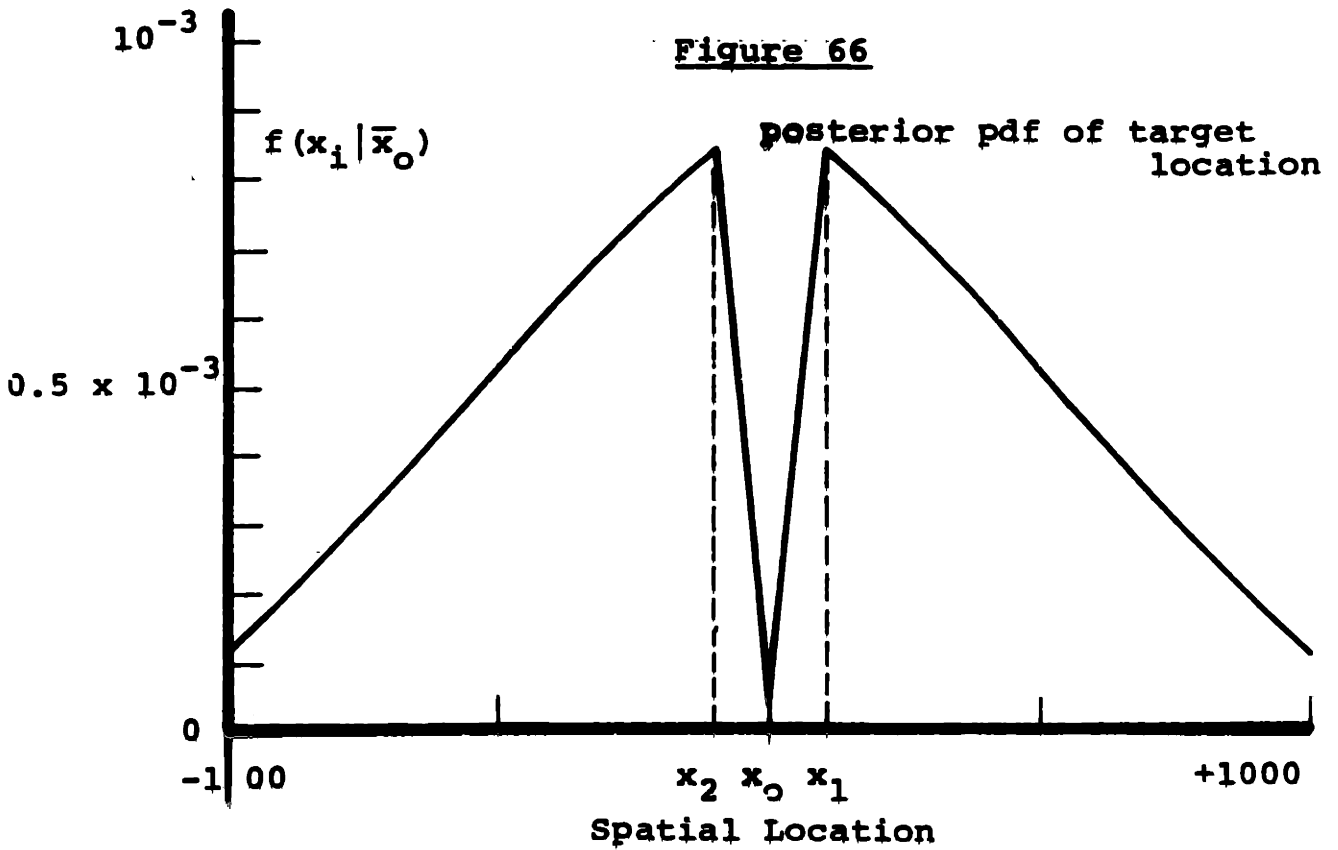


Figure 66



If no find is recorded at  $x_0$  the second boring should be placed at the location with the largest probability of a find under the posterior idistribution. Since a boring may record a find if the center of the deposit is within  $\pm b$  of the boring, the second boring should be placed either slightly to the right of  $x_1$  or slightly to the left of  $x_2$  (both locations have the same probability of a find) due to the local asymmetry of the posterior distribution about  $x_1$  and  $x_2$ .

If no find is made on stage 2 the new posterior pdf of location should be evaluated and the third stage should again search at

$$x_i: \lambda_i / c_i = \max_j \lambda_j / c_j$$

The procedure should continue until either a find is made or the posterior probability of a find is less than some acceptable value. The posterior probability that the target exists where no find has been made on the previous stage can be evaluated by a direct application of Bayes' Theorem

$$\beta' = \frac{\beta p(\text{miss @ } x_0 \mid \text{exists})}{\beta p(\text{miss @ } x_0 \mid \text{exists}) + (1-\beta)(1.0)}$$

where  $\beta$  is the current probability that the target exists prior to the present stage, and  $\beta'$  is the new posterior probability.

The evaluation of posterior pdf's of location can be made by hand, but because the calculations are tedious the evaluation is more appropriately made with an electronic computer.

### Section 6.4.2 Optimal Stopping

At each stage in the sequential-search procedure the engineer must decide whether to continue searching, in the hope of finding the target, or terminate searching and accept the risk associated with the current posterior probability that the target exists. The problem of deciding at which state the current risk associated with stopping is exceeded by the expected future costs of continuing the search is called the problem of "optimal stopping."

Blackwell (see Matula, 1964) shows that the optimal strategy  $\delta^*$  is independent of the stopping rules  $s$ , where  $s$  is the number of unsuccessful searches after which the current risk should be accepted. This means that no matter what the stopping rule  $s$ , the strategy  $\delta^*$  of section 5.1 is always optimal. In the terminology of operations research the strategy  $\delta^*$  is said to be "strongly optimal" since

$$\Pr \{ N > n \mid \delta^* \} \leq \Pr \{ N > n \mid \delta \} \quad 104$$

for any strategy  $\delta$ , and  $\delta^*$  insures the greatest probability of finding the target within any fixed number of inspections (Chew, 1967).

The problem of optimal stopping reduces to minimizing the expected cost of the search plus terminal risk, which can be expressed

$$E[C \mid \delta^*, s, f(x)] = \sum_{n=1}^s p(N=n \mid \delta^*, f(x))n + (s + \rho(\delta^*, s))p(N>s \mid \delta^* f(x)) \quad 105$$



where  $c$  = total cost in number of inspections or observations  
 $s$  = stopping rule  
 $f(x)$  = prior pdf of target location  
 $N$  = random variable number of searches until the target is found  
 $\rho(S^*, s)$  = risk associated with not finding the target after  $s$  searches using strategy

This expression assumes that search costs are equal at each stage and at each location.

Chew (1967) has solved a similar problem, in which the risk term is constant, by determining the set  $S_b$  of all probability density functions (in Chew's case, pmf's) for which the current risk of terminating searching without a find is exceeded by the expected risk of continuing. That is,  $S_b$  is the set of pdf's for which the expected cost of continuing the search is greater than that of stopping the search and making decisions on current information. If at any stage the posterior pdf is a member of the set  $S_b$ , searching is terminated. Because of the complexity of the mathematical formulation Chew finds sets satisfying an inclusion relation with  $S_b$  (i.e.,  $S_l \subset S_b \subset S_u$ ) which are "close" to the required set, and compares the distribution of interest with these bounding sets. Chew's solution is a general one, and more sophisticated than is actually required in site exploration.

For a given prior pdf all of the terms of equation (105) can be evaluated by forward induction techniques. The amount of computation involved makes hand solution infeasible, but, machine computation can be used at a cost much lower than the costs of field work (i.e., the

costs of searching beyond the optimal stopping stage greatly exceed the costs of machine computation).

Consider the first term in the right hand side of equation (105),

$$\sum_{n=1}^{\infty} P(N=n | \delta^*, f(x)) n. \quad \text{For } n = 1$$

$$\begin{aligned} \Pr \{N=1 | \delta^*, f(x)\} &= \Pr \{ \text{find in 1}^{\text{st}} \text{ stage} \} \\ &= \max_i f(x_i) (1 - \alpha_i) \end{aligned} \quad 106$$

For  $n = 2$

$$\begin{aligned} \Pr \{N=2 | \delta^*, f(x)\} &= \Pr \{ \text{no find in 1}^{\text{st}} \text{ stage} \} \times \Pr \{ \text{find in 2}^{\text{nd}} \text{ stage} \} \\ &= [1 - \max_i f(x_i) (1 - \alpha_i)] [\max_j f'(x_j) (1 - \alpha_j)] \end{aligned} \quad 107$$

where  $f'(x_j)$  is the posterior pdf of target location, given that no find was made in the first stage.\* For  $n = 3$

$$\begin{aligned} \Pr \{N=3 | \delta^*, f(x)\} &= [1 - \max_i f(x_i) (1 - \alpha_i)] \\ &\quad \times [1 - \max_j f'(x_j) (1 - \alpha_j)] \\ &\quad \times [\max_k f''(x_k) (1 - \alpha_k)] \end{aligned} \quad 108$$

and so forth, each term in the summation can be evaluated. The major computation time involved is in updating the location pdf at each stage.

The term  $\Pr \{N > s | \delta^*, f(x)\}$  is the complement of the probability of a find on one of the first  $s$  stages,

---

\* We implicitly assume a point-target in this discussion.

$$\Pr \{N > s | \delta^*, f(x)\} = \left[ 1 - \sum_{n=1}^s \Pr(N=n | \delta^*, f(x)) \right]$$

109

and can be evaluated as a side product of the evaluation of 29.

The risk term  $\rho(\delta^*, s)$  is a convex function of the probability that the target exists somewhere at the site, given that no find has been made in  $s$  stages of the optimal strategy  $\delta^*$ . The posterior probability that the target exists after  $s$  stages can be determined by induction of the direct application of Bayes' Theorem.

$$\begin{aligned} \Pr \{ \text{exists} | \text{no find}, \delta^*, s=1 \} &= \frac{\beta \Pr \{ \text{no find in 1st stage} | \delta^*, f(x) \}}{\beta \Pr \{ \text{no find in 1st stage} | \delta^*, f(x) \} + (1-\beta)} \\ &= \beta' \end{aligned} \quad 110$$

$$\begin{aligned} \Pr \{ \text{exists} | \text{no find}, \delta^*, s \} &= \beta^{(s)} \quad 111 \\ &= \frac{\beta^{(s-1)} \Pr \{ \text{no find} | \delta^*, f^{(s-1)}(x) \}}{\beta^{(s-1)} \Pr \{ \text{no find} | \delta^*, f^{(s-1)}(x) \} + (1-\beta^{(s-1)})} \end{aligned}$$

The only thing remaining to be shown is that the expected cost (equation 27) is a unimodal function of  $s$  (i.e., the function has a unique minimum). With this property the inductive computation can be terminated as soon as any minimum point is encountered.

A function can be shown to be unimodal if it is the sum of two functions, one with monotonically increasing first derivative (with respect to  $s$ ) and the other with monotonically decreasing first derivative. Unfortunately, equation 105 contains the term

$(s)p(N > s | \delta^*, f(x))$  whose first derivative is monotonic only for large  $s$ . Perhaps future work will prove 105 to be unimodal, but for now evaluations must be made for all  $s$  up to  $n$  such that  $p(n > s)$  is small.

### Section 6.4.3 Exponential Saturation Allocations

We have so far assumed that discrete quantities of effort have been expended, and that after each discrete quantity has been expended in searching one location for the target the probability density function is updated by Bayes' Theorem. In practice this is an improbable situation. A more likely situation is that amounts of effort greater than that necessary to search just one location are expended sequentially. In this case an allocation of effort over space -- similar to that discussed for single stage search (Section 2.3) -- is necessary at each stage of the search. By the same argument as before (Section 5.1) the myopic procedure (i.e., allocate effort in each stage in such a manner as to maximize the probability of a find in that stage) is optimal.

Consider a strategy  $\delta$  for which at any stage  $n$

$$p_n(\text{find}) < p_{n+1}(\text{find}) \quad 112$$

Consider also a strategy  $\delta^*$  exactly the same as  $\delta$ , except in which stages  $n$  and  $n+1$  are reversed

The difference in the expected value of the cost between these two strategies is

$$E[C|\delta] - E[C|\delta^*] = [P_n (1 - \sum_{i=1}^{n-1} p_i) + P_{n+1} (1 - \sum_{i=1}^n p_i)] - [P_{n+1} (1 - \sum_{i=1}^{n-1} p_i) + P_n (1 - \sum_{i=1}^n p_i)] > 0 \quad 114$$

Therefore, for any strategy  $\delta$  which does not conform to

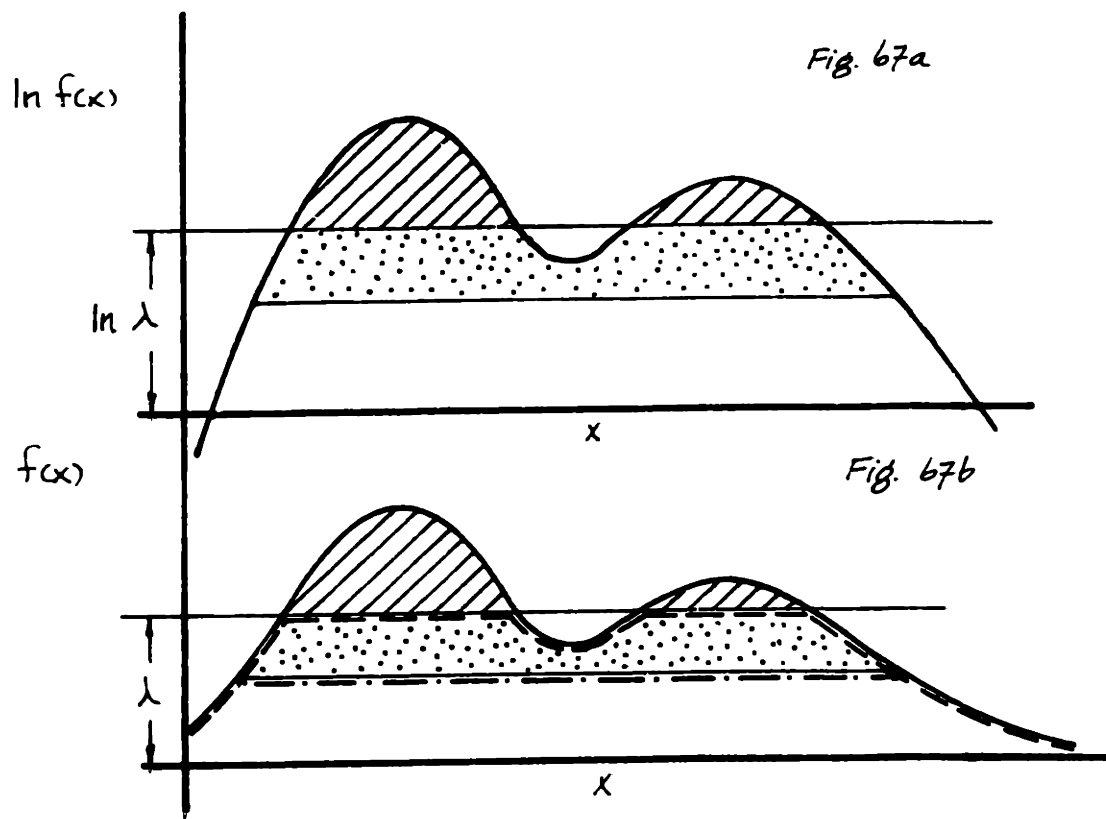
$$p_i \geq p_{i+1} \quad 115$$

another strategy  $\delta^*$  can be found with a lower expected cost. So, the optimal strategy must conform to relationship 115.

We shall now show that in a special, but common, case sequential allocations offer no increase in the probability of a find over single stage search for a given total amount of effort. The special case is the exponential saturation detection function. True, the expected cost of a find is lower for sequential allocations since there is a positive probability that the target will be found before the final allocation. However, if the target is not found after expending the total search effort,  $\Phi$ , the probability that the target exists is the same in both sequential and single stage procedures.

Consider the allocation of . units of effort over the prior location pdf shown in Figure 67. Using the methods of Section 2.3 the allocation is determined by translating a horizontal line parallel to the x axis until the area above the line and below the curve  $\log f(x)$  equals  $\Phi_1$ , then allocating all the effort in the

intervals  $A_1$  and  $A_2$ . The probability of a find equals the cross-hatched area in Figure 67b. The posterior pdf given no find is proportional to the dotted line in Figure 67b.



If an additional amount of effort  $\Phi_2$  is now allocated, the allocation is determined by a similar procedure using the posterior pdf of the target's location as a basis. The stippled area in Figure 67 corresponds to this second allocation of effort; the dashed and dotted line in Figure 67b to the posterior pdf given no find from the 2<sup>nd</sup> allocation. The posterior pdf after the 2<sup>nd</sup> allocation is the same as had the total effort  $\Phi = \Phi_1 + \Phi_2$  been allocated in one step and no target encountered. The probability

of a find during both allocations, represented by the sum of the cross-hatched and stippled areas in Figure 67b, is also the same as had the total effort been allocated in one stage.

Further discussion of the additive nature of effort for exponential saturation detection functions is given by Danskin (1964) and by Dobbie (1963).

### Section 6.5 Response Surfaces and "Optimal Seeking"

The term "response surface" refers to a surface defined by some functional relationship  $z = f(x, y, \dots)$ , and according to Wilde (1964) has its origin in biometrics where the variable  $z$  is often the response of living organisms to environmental variables  $x, y, \dots$ . Techniques of searching response surfaces for optima by sequential experimentation have been developed (primarily in chemical engineering), and recently suggestions have been made for extending their use to problems of mining exploration (Wilde, 1964, p. 109; Koch and Link, 1971, v. 2).

These techniques differ from the majority of those described in the present chapter because they are not directed at finding discrete targets, rather they seek parametric optima (e.g., at what coordinates of site location is a certain highly compressible clay stratum the thickest). Also, strategies of seeking response surface optima are not based on probability theory. Nevertheless, since these techniques are akin to search theory and may find application in selected problems, we will briefly discuss their approach to strategies of sequential search through the medium of an example.

If a large structure were to be built on a site of several acres extent, one might be advised to construct the facility where

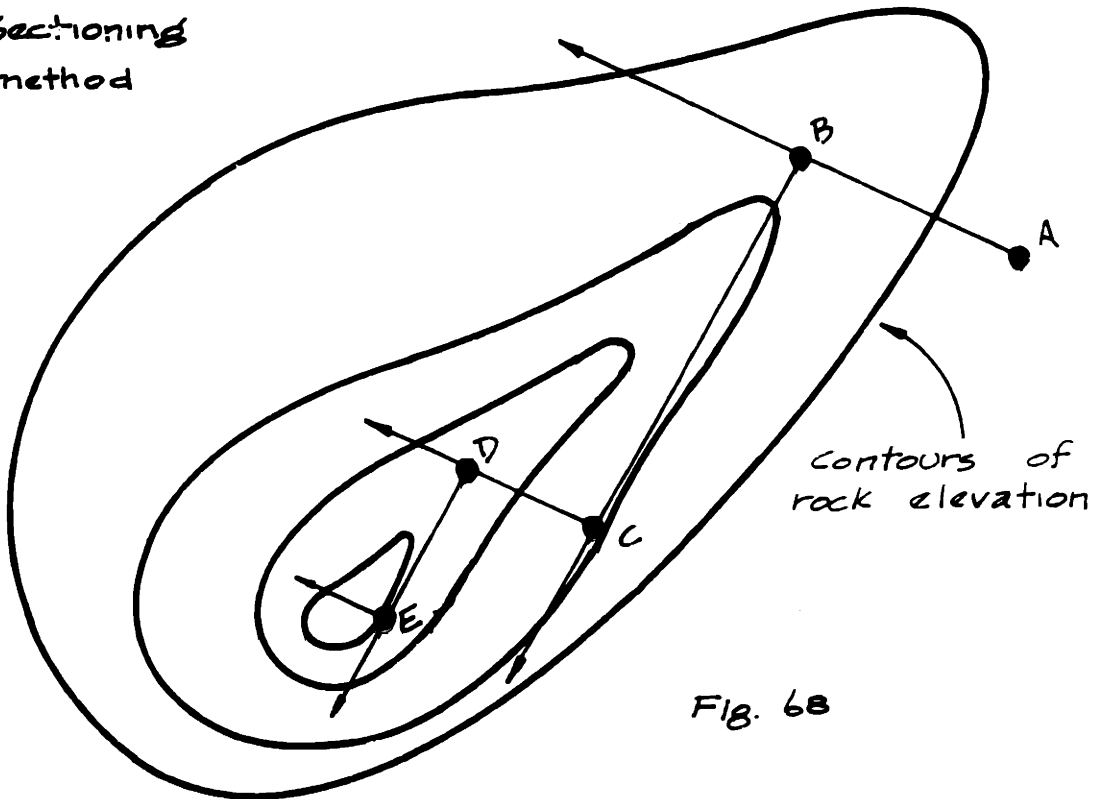
the depth to rock was the least. One purpose of exploration, therefore, would be to locate this point. Consider that Figure 68 shows the actual (and unknown) elevations of rock beneath the site. How can response surface techniques help in constructing a strategy of search?

A simple method is that of sectioning (proposed by Friedman and Savage, 1947). For our discussion we will denote rock surface elevation by  $z$  and site location by the parameters  $x$  and  $y$ . By the sectioning method the independent variables are varied and maximized independently until a point is found at which all of the variables are maximized simultaneously. Assume that seismic refraction surveys are used as an exploration tool which result in a continuous measure of  $z$  along a line. Starting at an arbitrary point (e.g., A in Figure 68) and traversing in the  $x$ -direction a maximum  $z$  along the traverse is found (point B). Now by traversing in the  $y$ -direction through point B a maximum  $z$  along the  $y$ -traverse can be found (point C). Once again traversing in the  $x$  direction, and so on we finally "zero-in" on the point of greatest rock elevation.

The sectioning method works well as long as no ridges (e.g., Figure 69) occur in the response surface. Ridges can cause the separation method to give a false maximum (i.e., a point at which both  $x$  and  $y$  are maximized, but which is not the true maximum) without the explorer's knowledge.



Sectioning  
method



Contours of  
rock elevation

Fig. 68

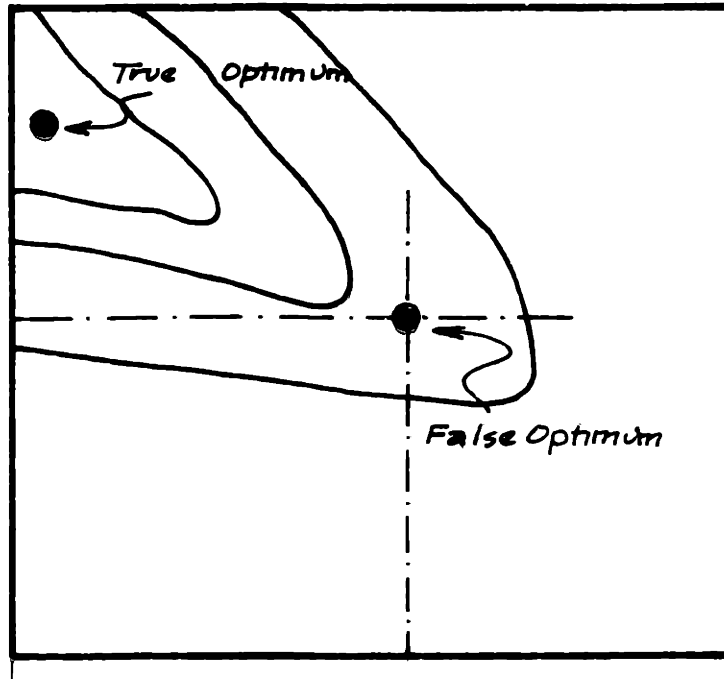
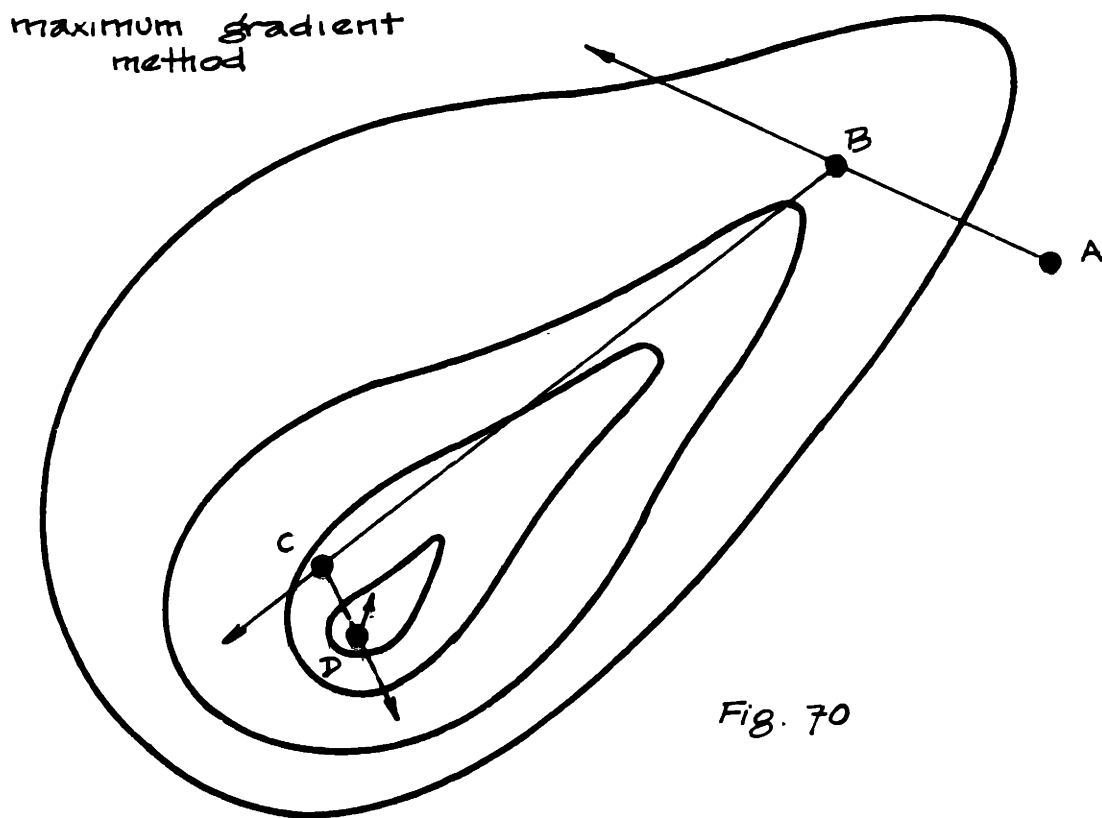


Fig. 69



The "gradient method" (or "method of steepest ascent") overcomes this difficulty (Wilde, 1964) by forming an approximation to the function gradient in a local region and allocating subsequent effort in the direction of steepest ascent. Using the gradient method three non-colinear point observations would be made in the neighborhood of A from which a plane approximately tangent to the response surface at A is found. A traverse is then made parallel to the gradient of this plane until a maximum is reached (point B', Figure 70). Once again an additional point observation is made (non-colinear with the traverse), the approximate tangent plane constructed, a new traverse parallel to the new gradient undertaken, and so on.

If the engineer has a priori information about the response

surface (e.g., elongation parallel to some regional trend), coordinate transformations and other techniques can be used to speed convergence.

We have highlighted only the simplest of response surface techniques to present a concept of their functioning. Finer points, more sophisticated techniques, and special strategies near the optimum are to be found in Wilde (1964). The example we have considered is especially amenable to response surface techniques, but in general geological exploration may not be. Response surface techniques assume that the function  $z = f(x,y)$  is smooth and has a continuous first derivative; in geological formations this is not necessarily the case. Additionally, geological formations frequently are multimodal, that is, they do not have a unique optimum. Response surface techniques are not developed adequately to handle multimodal problems, and will seek the local maximum or minimum they first encounter, which may be far from optimal.

## CHAPTER 7

JOINT SURVEYS

The geometry of joints at a prospective site is an important consideration in the design of structures in rock, since these planes of weakness significantly influence the engineering properties of rock masses.

Joint surveys are sampling programs for collecting information on jointing. They seek to obtain a description of the joint population which includes the distribution of orientations, intensity of jointing, and such other properties as continuity, roughness and state of weathering. In this chapter we treat orientation and intensity alone, not because the other properties are unimportant, but because orientation and intensity are highly amenable to statistical description.

As we have asserted in previous chapters, joint surveys comprise one problem class of site exploration in which the use of relative-frequentist methods rather than degree-of-belief methods is justified. The primary reason for this is that the use of joint survey findings in design can at best be described as qualitative: The state-of-the-art of rock mechanics does not require precise estimates of the distribution of joint orientation and intensity for design decisions. Current classical methods are easy to use, and since sample sizes in joint surveys are large ( $n \geq 100$ ), estimates made

by classical methods have high precision. The increased computational effort which Bayesian methods require is not justified by increased benefits in design.

The orientation of a plane in space can be uniquely defined by two independent parameters; civil engineers and geologists commonly use strike and dip, where strike is the direction of a horizontal line in the plane and dip is the vertical angle between the plane and a horizontal plane (Figure 1). Joints are almost universally represented by planes although in reality the planarity of a joint is only approximate. We will not dwell on this point since it is discussed frequently in the literature.

Intensity of jointing is basically some measure of the number of joints per unit of rock, where the unit of rock may be a volume, an area, or a length. Commonly, intensity is reported separately for each group of sub-parallel joints, called a joint set.

In this chapter we discuss programs for sampling joints and inferences which can be made from the resulting data. These problems must be approached statistically if we are to quantitatively express the uncertainty in our information on local jointing. In the first section we discuss requirements which joint sampling plans must satisfy, and propose a specific plan; in the second section we discuss statistical tests for making inferences from sampling data, and consider the use of analytical models for joint-orientation distributions; in the appendix we review general classes of classical sampling plans.

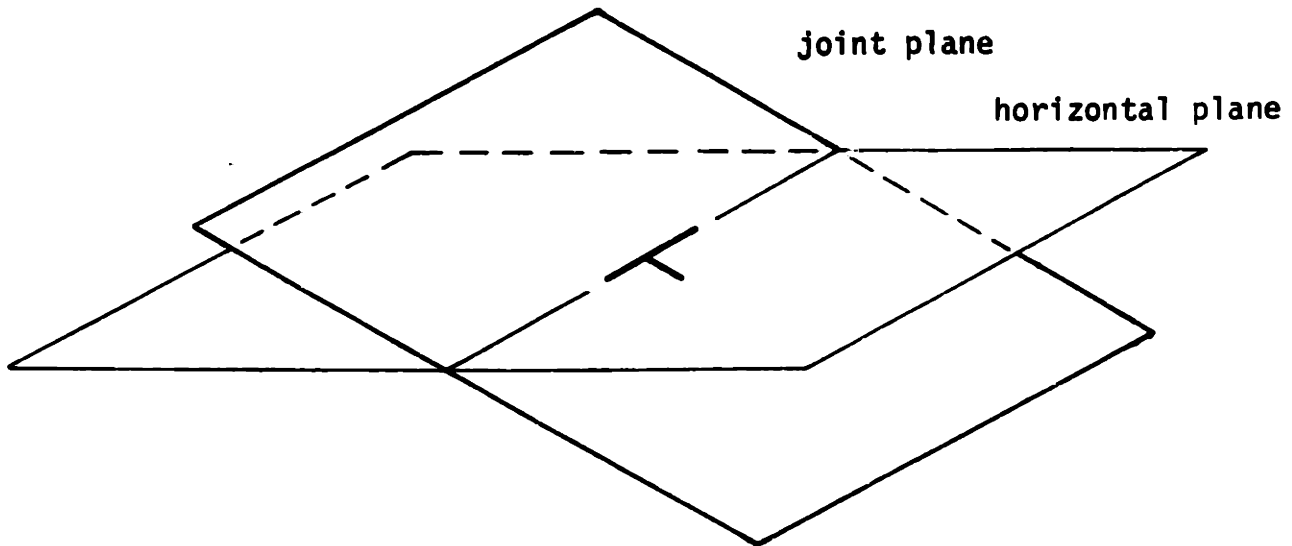


Figure 1

## Section 7.1 Principles of Sampling Joint Populations

### 7.1.1 Statistical Sampling

The objective of sampling is to infer characteristics of a large population without measuring all of its members, since measuring every member of a large population is usually infeasible, and sometimes impossible. A group of elements selected from the large population according to some statistical plan is called a sample. A sample has statistical characteristics (e.g., a sample mean, a sample variance, etc.) that we will call sample statistics, and from which we can estimate characteristics of the large population. As long as samples are taken according to a "statistical sampling plan" the precision of these estimates can be evaluated.

The large population in which we are interested is called the "target" population (Cochran, 1963). However, in many cases only a

sub-set of the target population is available for sampling. In measuring the orientations of joints at a site only the orientations of joints that appear in outcrops, boreholes, or excavations can be measured. The remaining joints are "unavailable". We define the population of available or measurable elements as the "sampled" population (Figure 2).

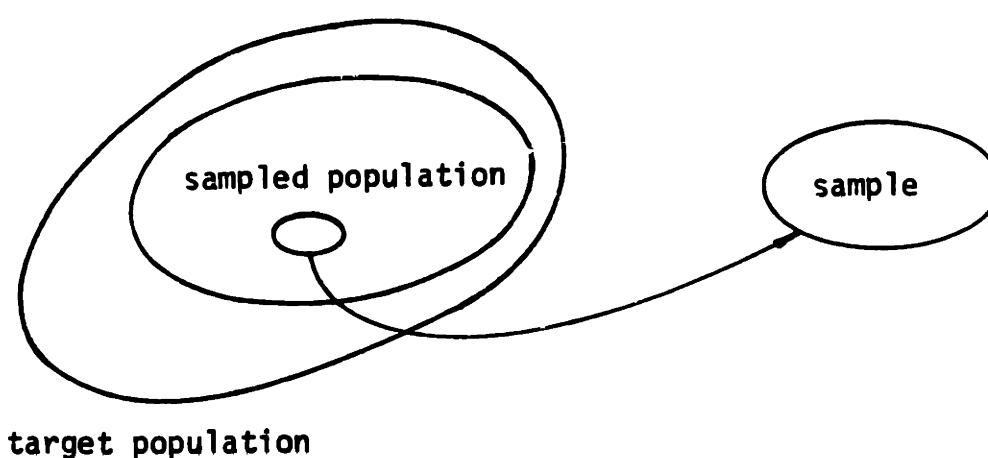


Figure 2

The characteristics of the sampled population can be estimated from the characteristics of the sample. This process is based on mathematical statistics in the form of sampling theory. The reliability of these estimates can be determined quantitatively. On the other hand, the characteristics of the target population can be estimated from the sampled population only by professional judgement. The reliability and precision of these estimates are not within the realm of mathematical statistics. The extent to which the joint orientations at a depth of 500 feet correspond to the orientations at the surface (barring drill holes and the like) is a geological

question, not a statistical one.

### Probability Sampling

A probability sample possesses four characteristics:

1. Each element of the sampled population has a probability greater than zero of entering the sample.
2. For each pair of elements in the sampled population, the relative probability of each entering the sample is known.
3. In analyzing the sample data each element is given weight in proportion to the reciprocal of its probability of being sampled.

$$(\text{Weight}_i) \times (\text{probability}_i)^{-1} = \text{constant}$$

4. For any two possible samples, the sums of the reciprocals of the relative weights of all elements in the samples are the same.

Although concern is frequently voiced about the representativeness of samples, note that an individual probability sample cannot be made representative of the sampled population -- the sampling plan must be made representative. This is accomplished not by giving each element an equal chance of being sampled, but by compensating for the differences in the probability of being sampled through weighting. Those individuals more likely to enter the sample are given less weight; those less likely are given more weight. The net effect is to give each element an equal chance to affect the weighted-sample mean. Such "general probability samples" (Cochran, Mosteller



and Turkey, 1954) are just as honest and legitimate as self-weighting samples (i.e., in which each element has an equal probability of being sampled), and often are considerably easier to use.

The process by which sample statistics are corrected to conform with known characteristics of the sampled population is called "adjusting". Consider the problem of determining the average compressive strength of a shale-sandstone sequence of predicting the lifetime of machine cutting heads. By simple random selection a sample of 100 4' cylinders of rock is obtained. As the sample turns out, 60 of the cylinders are shale, and 40 sandstone. The shale has an average unconfined strength of 250 kg/cm<sup>2</sup>; the sandstone, 1000 kg/cm<sup>2</sup>. The average strength of the sample is, therefore,

$$\begin{aligned} \overline{\text{strength}} &= 0.60 (250) + 0.40 (1000) \\ &= 550 \text{ kg/cm}^2 \end{aligned}$$

However, if by other means the engineer knows that the formation is 50% shale and 50% sandstone, he could "adjust" the sample statistics to arrive at the estimate

$$\begin{aligned} \overline{\text{adj. strength}} &= 0.50 (250) + 0.50 (1000) \\ &= 625 \text{ kg/cm}^2 \end{aligned}$$

for the average profile strength. In this way, the sample statistics are brought into agreement with known characteristics of the sampled population (e.g., relative proportions of shale and sandstone), and

the unknown characteristics of the sampled population are inferred from the adjusted sample statistics.

### Sources of Error in Sampling

Discrepancies between the sample characteristics and the sampled population characteristics can be attributed to three sources:

sampling error

measurement error

estimation error

Sampling error is caused by sampling plans which are not representative of the sampled population. Elements with zero probability of being sampled, or unaccounted-for differences in the probability of occurrence between elements, would be causes of sampling error. By conforming to the four characteristics of probability samples, one eliminates this error. If the sampling plan is representative, differences between samples due to the random way in which they are selected can be evaluated by sampling theory.

Measurement errors are caused by inaccuracies in either the instruments of measurement (e.g., a Brunton compass) or the reading of instruments (e.g., operator variations). A substantial treatment of measurement errors can be found in most surveying or instrumentation textbooks (e.g., Wilson, 1952).

Measurement errors are of two types, random and systematic. Random errors are unpredictable both in magnitude and algebraic sign. They are accidental and cannot be avoided. The treatment of random

errors usually assumes that the magnitude of the error is normally distributed with mean zero, which allows us to place confidence limits on the measurements themselves, prior to considering estimation errors. Fortunately, random measurement errors are usually much smaller than sampling or estimation errors, and can often be neglected.

In circumstances where quantities depend on several measurements, random errors propagate and may be no longer negligible. If the variance in the random error of any quantity  $i$  is  $\sigma_i^2$ , and the desired quantity  $z$  is some function of  $i$ ,

$$z = f_n(x, y, \dots, n) = u(x, y, \dots, n) \quad 1$$

then

$$\sigma_z^2 \cong \sum_i \sum_j \left( \frac{\partial u}{\partial i} \right) \left( \frac{\partial u}{\partial j} \right) \text{Cov}(i, j) \quad 2$$

where  $\text{cov}(i, j)$  is the covariance of  $i$  and  $j$  ( $\text{cov}(i, j) = E(ij)$ ).

Considering some common functions of independent variables:

$$\text{If } z = i + j + \dots + n, \text{ then,} \quad \sigma_z^2 = \sigma_i^2 + \sigma_j^2 + \dots + \sigma_n^2 \quad 3$$

$$\text{If } z = (i)(j), \text{ then,} \quad \sigma_z^2 \cong j\sigma_i^2 + i\sigma_j^2 \quad 4$$

One can determine the variance of random error by making repeated measurements of one element of a sample. Making the common assumption that random error is normally distributed, confidence limits on a measured value can be set using tables of area under the normal curve.

Sometimes the variance of the sampling error is a function of the magnitude of the measurement. This is true for the strike and dip of joints as will be discussed in Section 7.1.2. The random error in measuring the strike of a steeply dipping joint is less than the random error in measuring the strike of a gently dipping joint.

Systematic errors are errors whose mean value is different from zero. Measured values of a quantity will, therefore, be almost consistently high or consistently low. Systematic errors, like random errors, are inevitable; however, unlike random errors, they are not reduced by large sample sizes. The only strategy that can be used against them is holding them to a "reasonable" level -- "reasonable" must be defined on the basis of the cost of further reduction and the cost of inaccuracy. We never know whether the systematic errors have been reduced sufficiently: We can only carefully consider possible sources of error, and search for consistencies (or inconsistencies) in the sample data. The pdf of unknown systematic errors is obviously also unknown; thus, the magnitude of systematic errors cannot be accounted for in assessing the reliability of estimates made from sample data (unless a subjective approach is used).

Estimation errors are caused by differences in sample statistics from sample to sample caused by the random nature of samples. For example, since different samples of sandstone cylinders do not all yield the same average strength, they lead to different estimates of in situ intact strength.

Since representativeness is not inherent to the individual sample, we cannot evaluate the uncertainty in sample results without evaluating the similarity (or lack of similarity) between the obtained sample and samples which the sampling plan might have provided. The degree of similarity between samples is referred to as the "stability" of a sampling plan. The width of the confidence limits of an estimator is a measure of the precision of the sampling plan. From a knowledge of the sampling plan, the precision of an estimator, and hence the confidence limits on the estimate, can be determined.

An analytical determination of precision can sometimes be complicated, and simpler empirical methods are sometimes available. One such method is the use of interpenetrating replicate samples. A large sample of, say, 100 elements can be randomly divided into several smaller samples, say 10 samples of 10 elements each. We can evaluate each of the 10 replicate samples as if they were chosen individually at random from the sampled population, and assess empirically the stability of the 10 estimates. This, in turn, can be used as a yardstick for the stability of the lumped sample.

## Section 7.2 Sampling Theory Applied to Joint Surveys

The correspondence between target and sampled populations in joint surveying can be treated only by geological judgement, not by statistics. For example, the target population might be the joints at the elevation of a proposed galley 1000 feet below the present surface, while the sampled population in the same case might be the

surface joints or those in shallow borings (Figure 3). This is a point that most engineers give lip service to, but it deserves reiteration here.

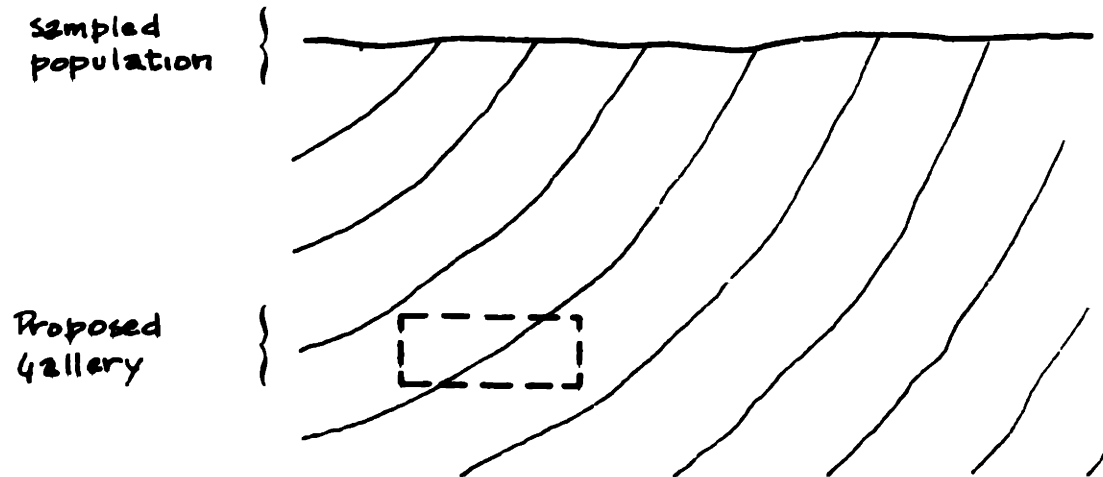
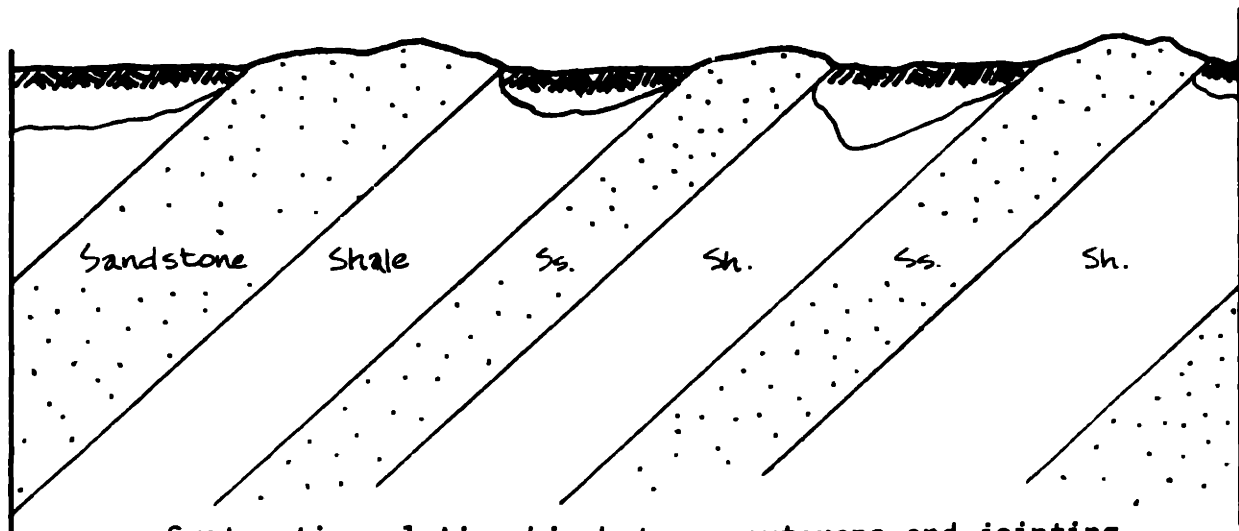


Fig. 3

Some work has been published on the correspondence between target and sampled joint populations (Robinson and Lee, 1965; Wahlstrom, 1964), but additional work is needed.

Outcrops, which are the source of the sampled population, are actually presampled clusters of joints. The bias in this presampling is, of course, unknown: At best, the outcrops are randomly located independent of jointing, and at worst may be functionally related to jointing (Figure 4).



Systematic relationship between outcrops and jointing  
(joint orientation different in sandstone and shale)

If they are functionally related, the outcrops will contain an atypical set of joints, and the sampled population will not be representative of the joints near the surface. In borings, low core recovery may be functionally related to jointing, and the sampled population will again not be representative. These are problems which the engineer must face with little help from statistics.

During construction new information becomes available on the target population. Statistical comparisons of this information with the original survey should be made, and will be discussed in Section 7.2.2.

### Measurement Error

Although random errors in joint orientation measurement arise from a host of sources, several general comments can be made about them.

Random error in the strike direction is greater for "flat" dipping joints than for "steep" dipping joints. The sensitivity of the direction of the line of intersection of two planes to error in the orientation of one or both is a function of the angle between them: the smaller the angle, the more sensitive. Since flatter joints form a smaller angle with the horizontal plane, they are more sensitive to errors in leveling the compass.

Random error in the dip is greater for steeper joints than for flatter joints. This error comes primarily from two sources other than reading errors and round-off errors: inaccuracy in leveling the pendulum inclinometer, and inaccuracy in alligning the Brunton parallel to the dip direction. While the first is independent of the dip, the second is not (Figure 5). The scale of surface roughness to compass size will also contribute to random measurement error (Figure 6).

While one could attempt to determine random measurement error analytically, the simplest and most reliable way is to simply perform several measurements on a single joint and empirically determine the dispersion of values. This was done in a laboratory setting with two simulated joints (i.e., fixed planes) by having several people measure the strikes and dips with a Brunton compass (Figure 7). The variance in the dip measurements was about the same for both planes, but the variance in the strike measurements was almost twice as great for the flat plane as for the steep plane.

Systematic errors in joint orientation measurement can be found



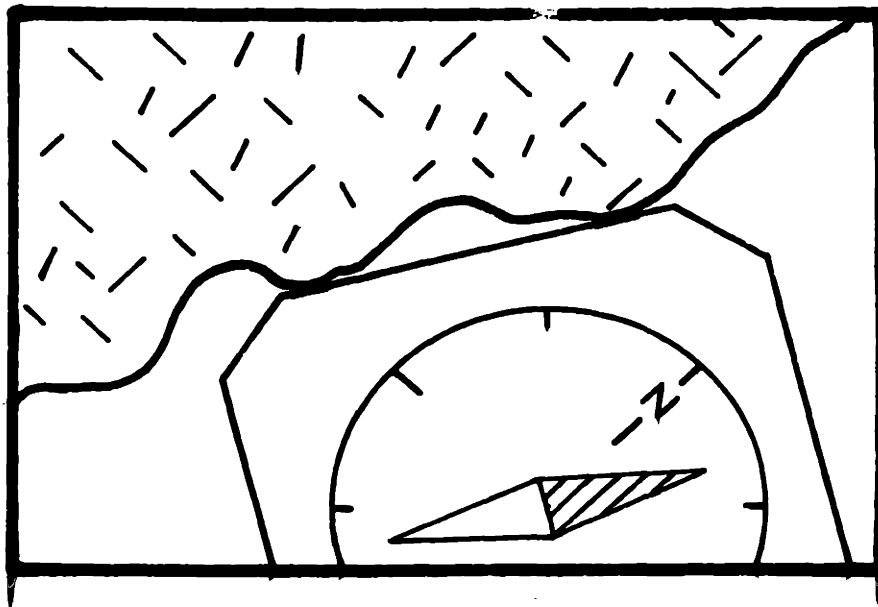
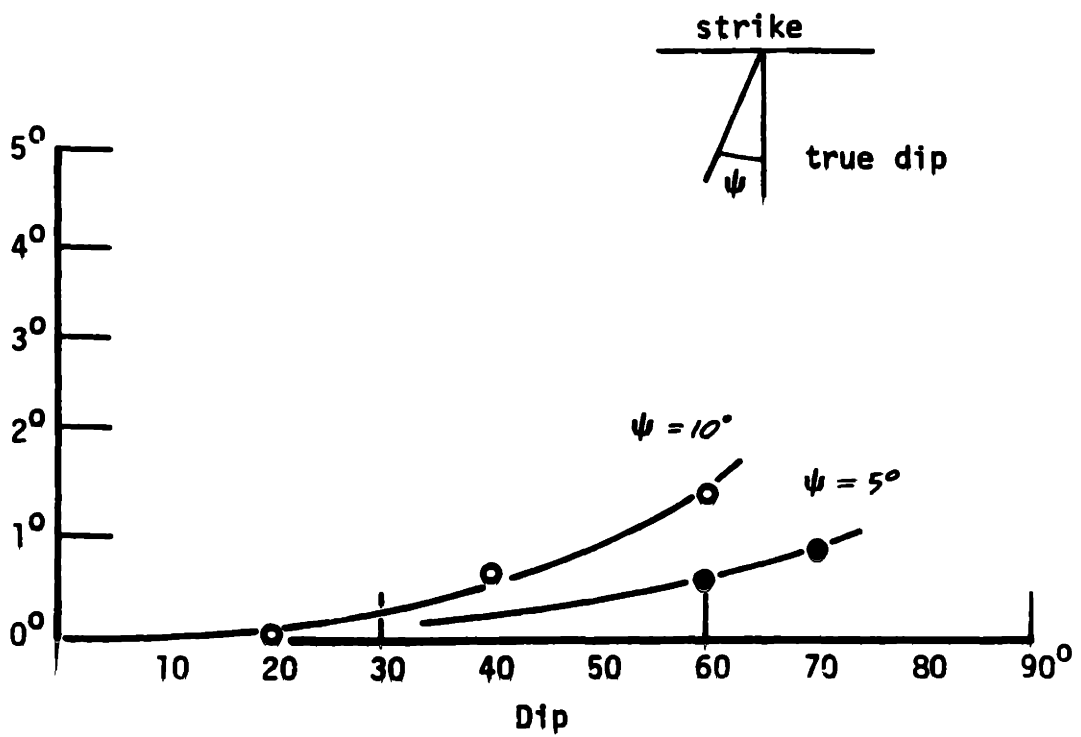


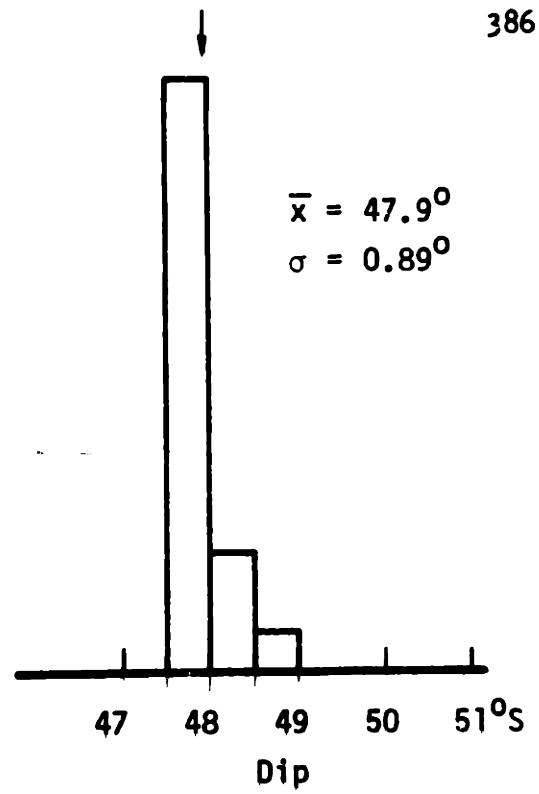
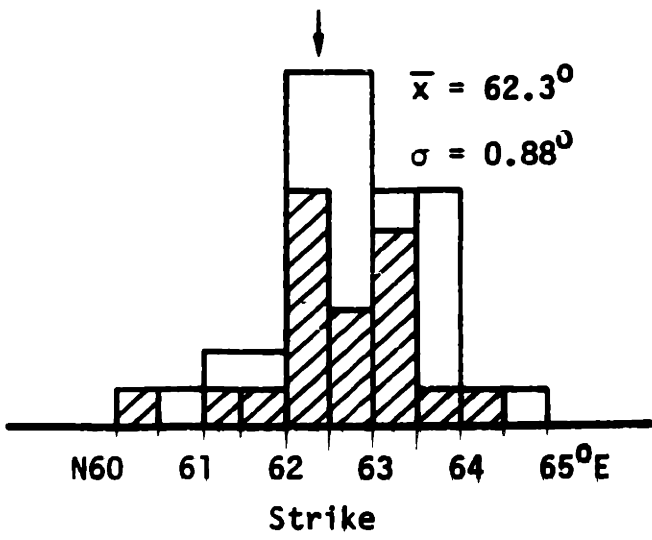
Figure 6

Figure 5

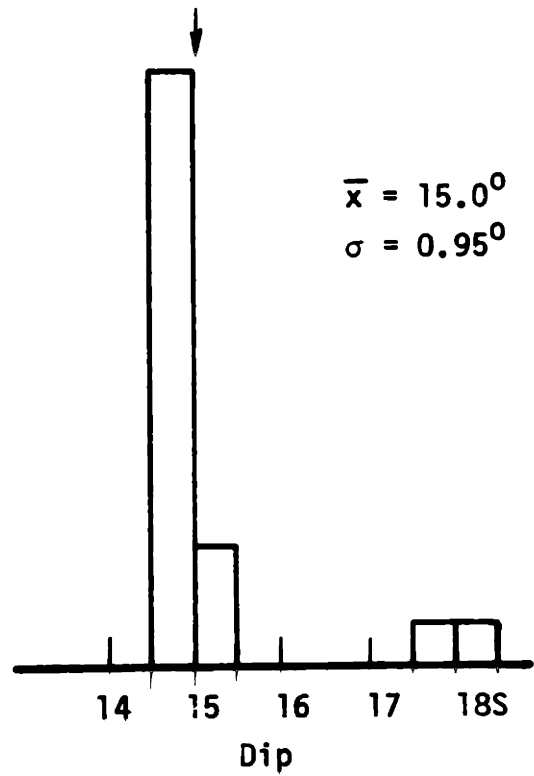
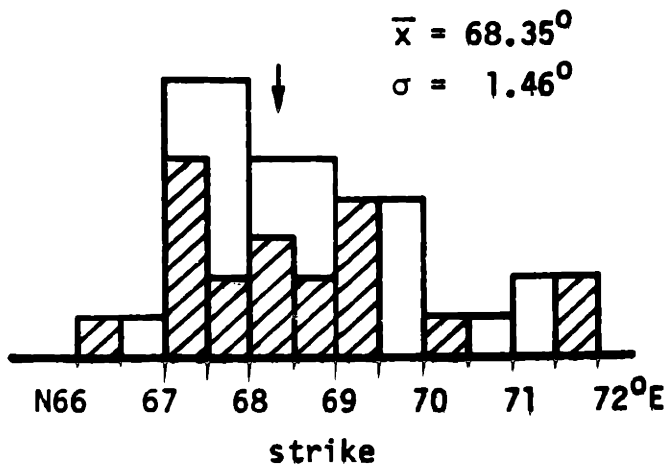


Error in measured dip due to misalignment of Brunton compass by  $\psi^\circ$

Figure 7



"steep" plane



"flatter" plane

only by analyzing the possible sources of error. They cannot be determined empirically by remeasurement as random errors can. Some examples of systematic error -- and, of course, there are others -- are:

1. Incorrect declination adjustment for true north.
2. Misadjusted bubble glass on compass or pendulum incunometer.
3. Holding a geology hammer always in the same hand, near the compass, while making measurements.
4. Always rounding off measurements in the same direction.

### Weighting and Analysis

Probability sampling requires that four conditions be met. Important among them is that each element be given an equal chance to affect the weighted sample mean. This is accomplished through the process of weighting.

In sampling joint orientation data, differences in the probability of being sampled arise, so the sample is generally not self-weighting. Differences in the probability of being sampled are caused by geometric relationships (e.g., the relative orientation of joints and outcrops) and non-geometric relationships (e.g., differences in the degree of weathering between joints). Geometric causes are the subject of this section.

Geometric relationships which cause joints to have a low probability of being sampled were brought to the attention of the

engineering profession by R. Terzaghi (1965), although Sander, et al (1954) and others had earlier considered similar phenomena in the study of thin-sections.

To be sampled a joint must be a member of the sampled population. It must intersect an outcrop, a boring, or an excavation from which samples might be drawn. Joints which do not, cannot be measured. Assuming that the sampling plan by which the joints are sampled from outcrops and borings is self-weighting (Section 7.1.3), the probability of an orientation entering the sample is proportional to the probability of it intersecting an outcrop or boring. Consider a two-dimensional situation as shown in Figure 8. The probability of a joint of a given orientation intersecting the ground surface in interval  $\Delta L$  is

$$P(\text{intersecting } \Delta L) = \frac{\Delta L \sin \alpha}{d} \quad 5$$

So the probability of being sampled is proportional to  $\sin \alpha$ . In other words, given the spacing,  $d$ , joints which are flatter with respect to the surface will appear less frequently in the sample than joints which are steeper. In order to be a probability sample this difference must be accounted for by weighting.

Since the ratio of probability must be constant, if  $w_\alpha$  is the weighting factor for joints at angle  $\alpha$ , and we assume  $w_{90} = 1$ , then,

$$w_\alpha = 1/\sin \alpha \quad 6$$

Although the probability is also proportional to  $1/d$ , this is

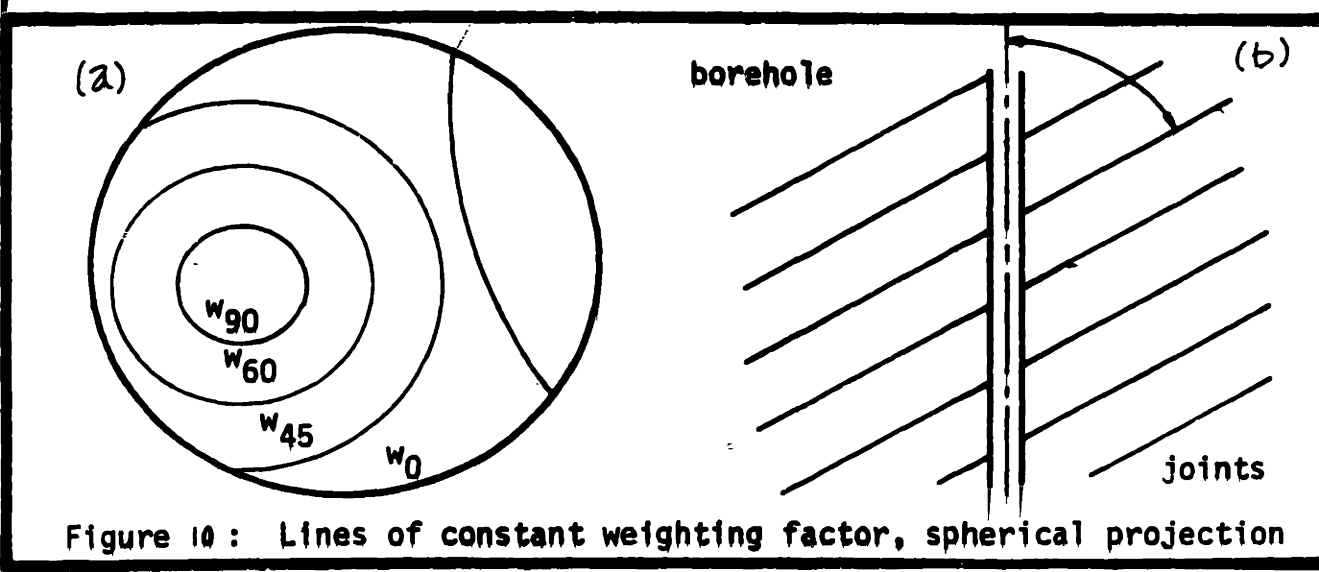
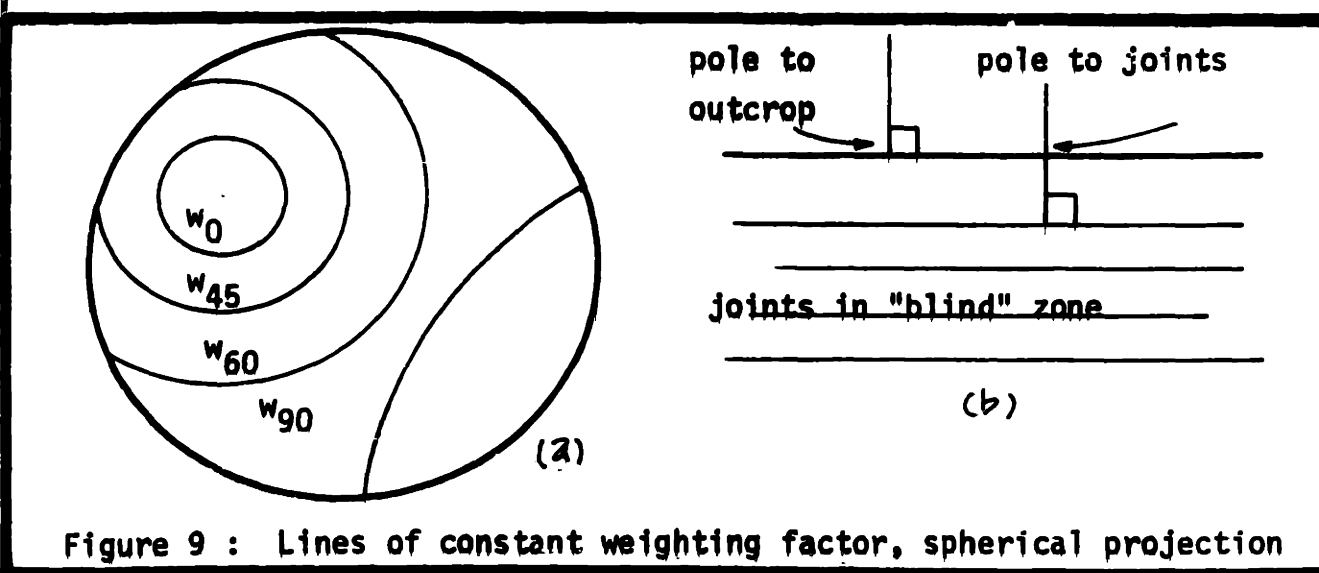
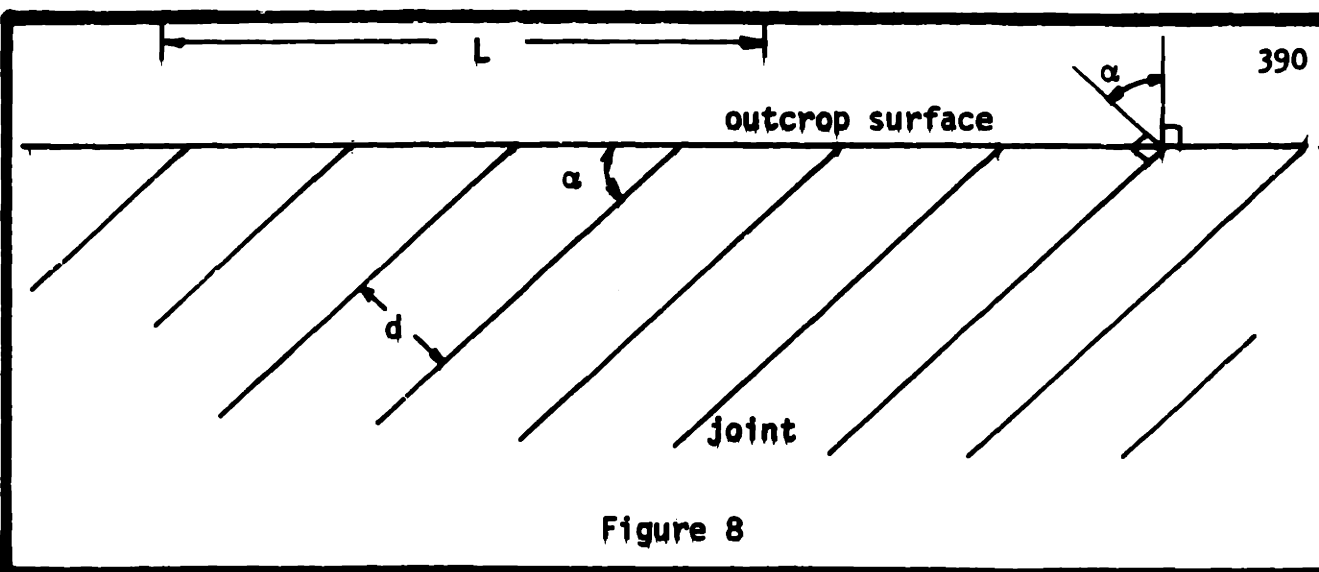
in turn proportional to frequency of occurrence -- which is what the sample is trying to determine, and so must not be compensated for by weighting.

Figure 9a shows an equal area plot of the pole to a planar outcrop, and loci of constant weighting factor. Notice that as  $\alpha \longrightarrow 0$ ,  $w_\alpha \longrightarrow \infty$ . This corresponds to the physical occurrence of a "blind" zone for certain joint orientations (Figure 9b). When the joints and the outcrop surface are parallel, no members of the joint set will be sampled. Exfoliation jointing is a good example.

Weighting factors for borehole sampling are much the same as those for outcrop sampling with the exception that the defining direction of a borehole is parallel to the bore hole axis. (Figure 10).

Terzaghi recommends plotting the data from each outcrop or boring separately, and weighting them independently, then summing the data onto one plot after weighting. For hand computations she recommends using an overlay of the weighting factors (Figures 9 and 10) and weighting all densities within bands by the average weight of the band. This is approximate, but usually sufficient.

The use of electronic computers allows the consideration of more precise methods: methods in which the sample is considered as a whole and in which closer approximations may be made. This can be done by evaluating the probability of a given orientation appearing in the total sample (i.e., in any of the outcrops or borings) and applying weights accordingly. The probability of a joint of a given orientation occurring in outcrop  $i$  is proportional to



$$P(\text{orientation } \alpha \text{ in outcrop } i) \propto B_i \sin \alpha_i, \quad 7$$

where  $B_i$  is some dimension of the outcrop (see Section 6.1.3) and  $\alpha_i$  is the angle which the orientation makes with the normal to the plane that best models the outcrop. Since a single joint may be sampled in more than one outcrop or boring (i.e., sampling "with replacement") the probability of the orientation being measured in the entire set of outcrops sampled is proportional to

$$\text{Pr}(\text{given orientation being seen in the entire sample of outcrops}) \propto \sum B_i \sin \alpha_i \quad 8$$

Similar consideration for boreholes leads to

$$\text{Pr}(\text{given orientation being seen in the entire sample of borings}) \propto \sum L_j \cos \beta_j \quad 9$$

where  $L_j$  is the length of the  $j^{\text{th}}$  boring and  $\beta_j$  is the angle which the joint makes with the  $j^{\text{th}}$  bore hole axis. The differences between equations 8 and 9 reflect differences in the orientation-defining axis of outcrops and borings.\* The probability of a given orientation appearing in the entire sample is

$$\begin{aligned} &\text{Pr}(\text{given orientation in entire sample}) \\ &\propto \sum B_i \sin \alpha_i + \sum L_j \cos \beta_j \quad 10 \end{aligned}$$

and the weighting factor, being proportional to the reciprocal of

---

\*For the present we will concern ourselves only with oriented core borings in which measures of strike and dip are directly measurable.

the probability, is,

$$W(\text{given orientation } \underline{\epsilon} ) \propto \frac{1}{\sum B_i \sin \alpha_i + \sum L_j \cos \beta_j} \quad 11$$

The procedure for determining the weighted densities of joint orientation is, therefore

1. Coalesce the entire sample and evaluate the unweighted density as a function of orientation.
2. For each arbitrarily small element of orientation (i.e., each small interval on a stereoplot) determine the set of angles  $\alpha_i$  and  $\beta_j$ .
3. Weight each element according to equation 11 to obtain the corrected relative densities.

Three wall case:

The case of sampling from three mutually perpendicular walls is so common that special attention will be given to it. This condition arises, for example, in sampling joints from the interior of galleries or adits.

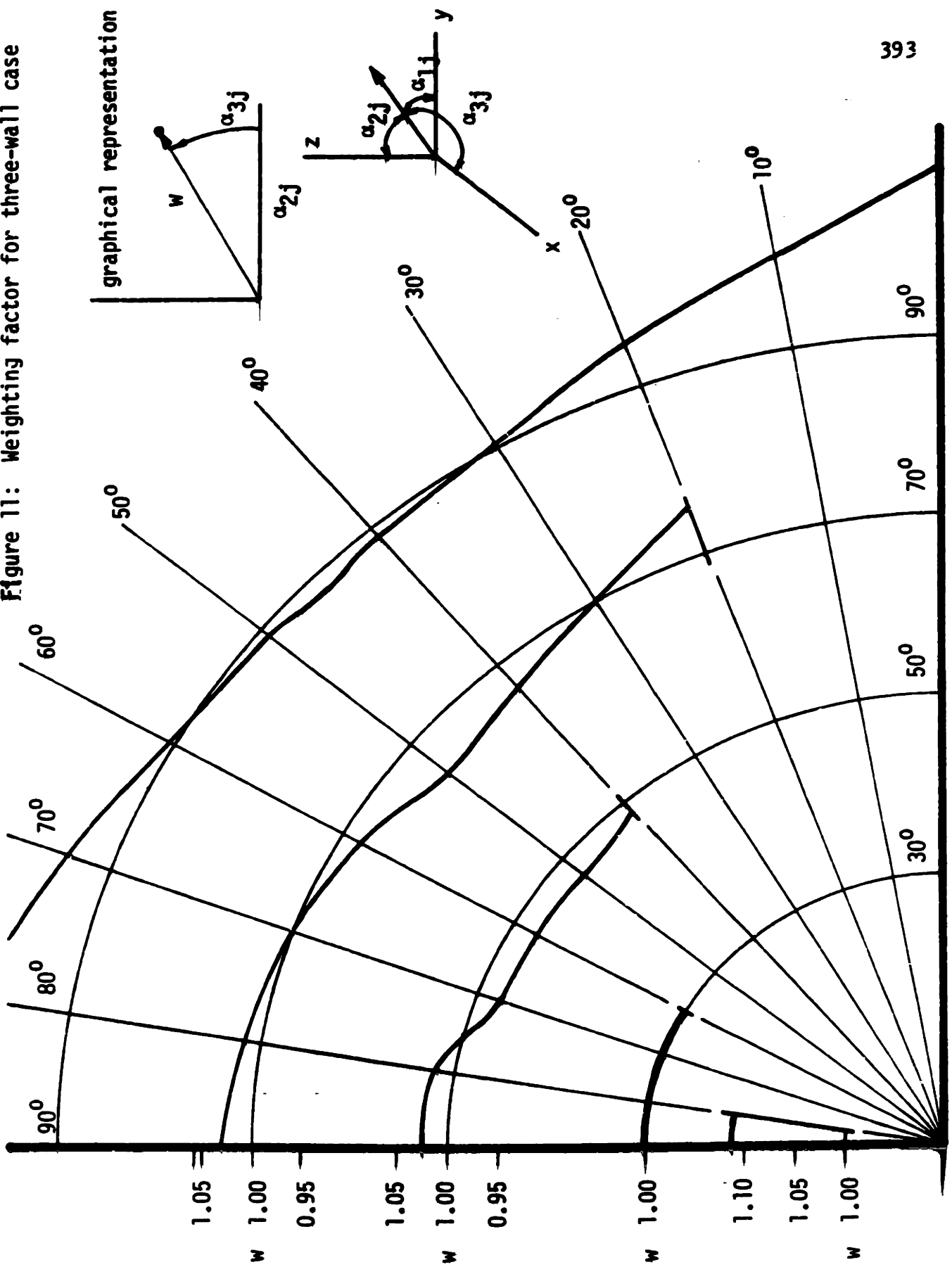
We assume that a single joint may be sampled on more than one wall; however, if the sampling fraction (i.e., sample size/population size) at each wall is small, the probability of this occurring is negligible. The probability of sampling a given orientation  $j$  is

$$P_j \propto \sum_{i=1}^3 B_i \sin \alpha_{ij}$$

where  $B_i$  is the dimension of the sampling area on wall  $i$  and  $\alpha_{ij}$



Figure 11: Weighting factor for three-wall case



is the angle between the pole to joint orientation  $j$  and the pole to wall  $i$ . The weight appropriate to orientation  $j$  is proportional to the reciprocal of the probability

$$W \propto \frac{1}{\sum_{i=1}^3 B_i \sin \alpha_{ij}} \quad 12$$

Since only two of the three angles  $\alpha_{ij}$  are independent ( $\cos^2 \alpha_{1j} + \cos^2 \alpha_{2j} + \cos^2 \alpha_{3j} = 1$ ), graphs of the weighting factor as a function of  $\alpha_{1j}$  and  $\alpha_{2j}$  can be constructed. Figure 11 is such a graph for the common condition  $B_1=B_2=B_3$ . This condition is easily attainable by sampling from an area of the same size on each wall.

### Section 7.1.3 A Sampling Plan for Joint Surveys

Sampling plans for joint surveys must meet two criteria:

1. They must allow valid statistical inferences to be drawn, whose precision can be evaluated (i.e., probability sampling)
2. They must be economical and easy to implement.

In most cases the cost of analysis is much less than the cost of field data collection, so plans which minimize the sampling effort are to be favored.

Simple random sampling of joints is almost always infeasible. These plans require randomly selecting individual joints around the site and measuring their orientations. Measuring 100 different joints by simple random sampling would mean traveling to 100 different

locations at the site and measuring one joint at each.

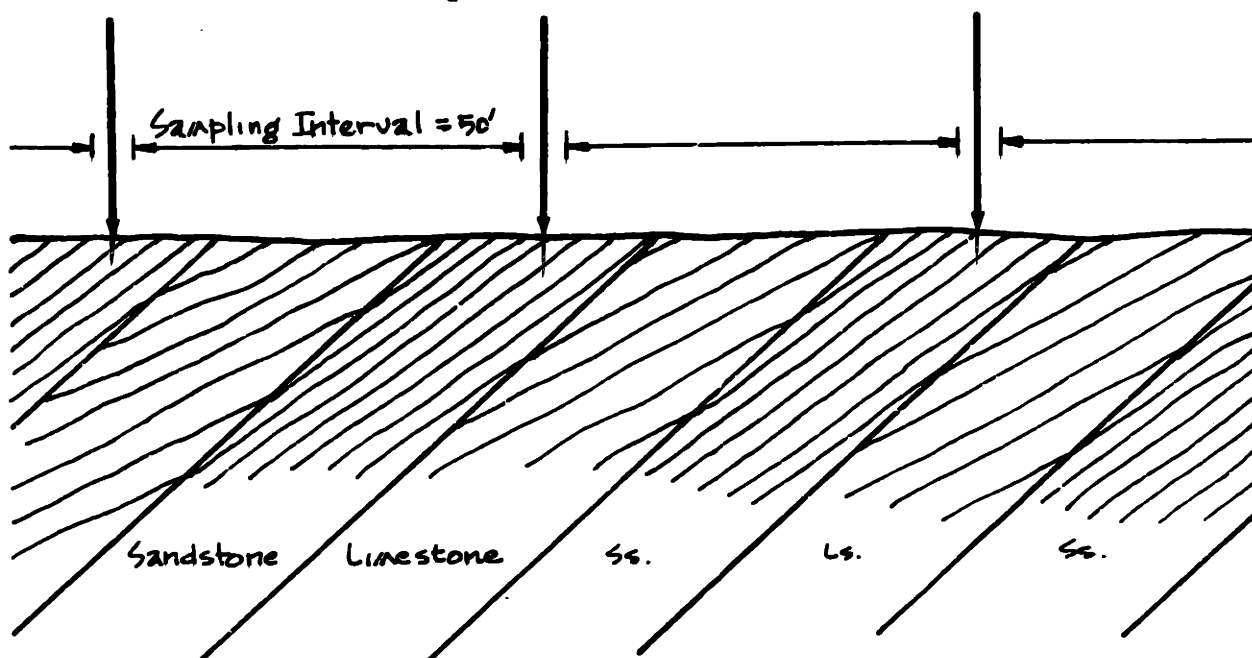
For the same reason, stratified random sampling is infeasible unless strata are small (e.g., the size of outcrops). Stratification is not an innate property of populations in general; if a population is stratified into internally homogeneous subpopulations the strata sizes are a property of the population not just the sampling plan. Joint populations are naturally stratified into joint sets, and joint populations are frequently stratified into geographical or lithological subpopulations (Price, 1966; Ramsay, 1967). Prior stratification by these properties may improve the performance of any sampling plan.

Systematic plans for sampling joints are easier to use than simple or stratified random ones because the joints to be measured are easily (if not quickly) located. A plan which specifies every 100th joint, say, is infeasible, but a plan which specifies the joints within a 6" circle every 50 feet, say, is not. Problems of periodicity in the sampled population might be encountered if stratification by lithology does not precede systematic sampling (Figure 12).

Cluster sampling plans have long been favored for joint surveys because the time required to sample several joints at one outcrop is less than the travel time between outcrops. In cluster plans several outcrops are selected by some random process, and from each selected outcrop a sample is taken (also by some random process).

The sampling plan we suggest for joint surveys is based on cluster sampling, and conforms to the requirements so far discussed. This is but one of many sampling plans which could be devised.

Figure 12: Example of error generated by using systematic sampling plan when regional periodicity exists. Direction of jointing changes due to differences in the internal friction angle of the intact materials under a uniform regional stress field.

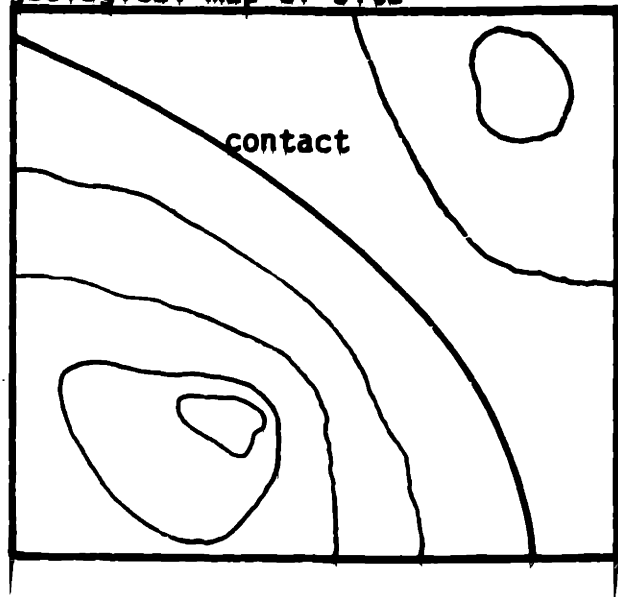


Different geological formations at a site frequently have different patterns of jointing due to differences in rigidity, friction angle, age of jointing, and the like. Therefore, the initial step is to stratify the site by major formations (Figure 13). Since one does not know a priori whether the populations of joints are homogeneous from formation to formation, these data sets are maintained separately.

Next each formation is arbitrarily stratified by superimposing a large regular grid, and the data from each quadrat (i.e., each square element) kept separately (Figure 14). The dimension of this

Figure 13

geological map of site



site stratified by lithology

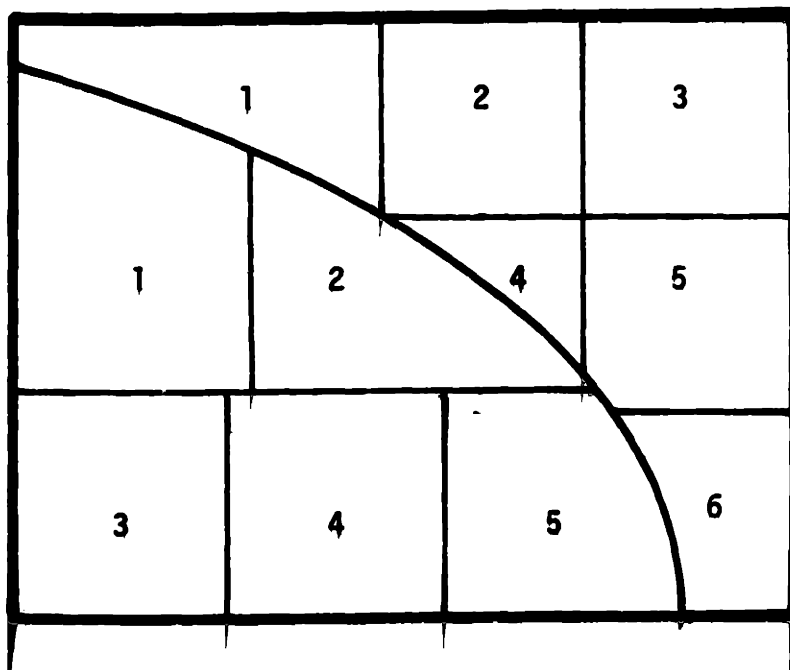
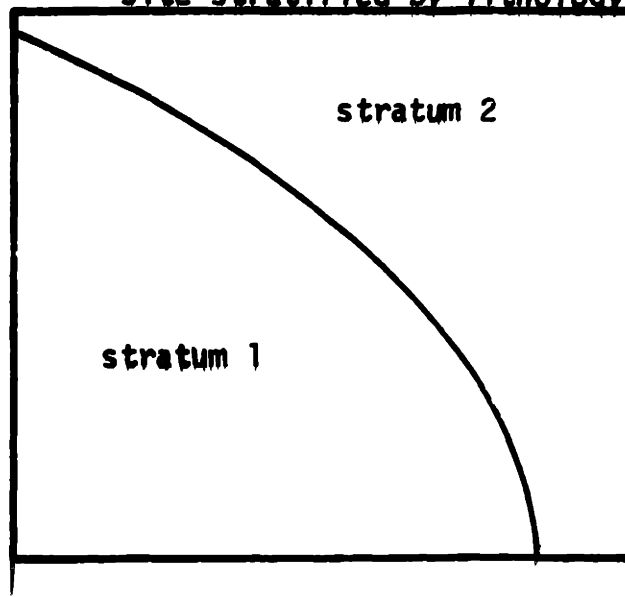


Figure 14: Each lithological unit divided into "sampling units"

grid might be on the order of 1000 feet. This stratification will allow us to perform a "nested analysis" of variance (Section 7A.4) and obtain variances in joint-population properties as a function of the spatial dimensions considered (e.g., within strata, between adjacent strata, etc.). We desire this information because the variance of joint properties generally increases as the volume of rock considered increases. Were we designing a structure which only affected a small volume of rock (e.g., an arch dam abutment) estimates made from the total joint population would overestimate the true local variance, and perhaps either overestimate or underestimate the local mean. The formation is not being stratified to improve the overall estimate of population parameters (the usual reason for stratification) since each stratum is treated identically, but simply to maintain separation of the data sets.

Within each stratum clusters of joints are selected for sampling. If few outcrops exist, all of them should be sampled; if not, a random process should be used for selection. This random selection might be accomplished by numbering the outcrops and choosing on the basis of a table of random numbers taped to the underside of a clipboard. The orientation and size of all selected outcrops should be recorded.

Several joints intersect each outcrop to be sampled. If their number is too large, not all of them may be measured and a second-stage sampling plan is required.

Two second-stage sampling plans which should be avoided, even

though they are frequently used, are systematic plans (say, sampling every tenth joint along a line) and plans randomly locating points on the outcrop and measuring the closest joint. Systematic plans should be avoided because periodicities are likely to exist in the way joints intersect an outcrop (Figure 15a), while sampling by measuring the closest joint to a random point should be avoided because joints whose individual spacing is large have a higher probability of being sampled than ones whose individual spacing is small (and these differences are not accounted for by weighting) (Figure 15b).

Snow (1966) has suggested sampling a single outcrop by randomly locating a line segment and measuring every joint which intersects it -- a common suggestion. This is a satisfactory method in that it is random and does not allow personal bias in selection, no matter how tight, small, or hard a joint is to measure. However, this procedure leads to large weighting factors and a "blind zone" for those joints whose normal vector is perpendicular to the sampling line. These large weighting factors and the blind zone can be reduced by using two mutually perpendicular line segments. Perhaps the easiest method (in that second-stage weighting factors need not be considered at all) is to draw a circle on the outcrop with a piece of chalk and a string and measure each joint passing through it. This method is strongly suggested because the weighting factor is isotropic and because the "outcrop dimension" necessary for determining first stage weights is simply the diameter of the circle.

Estimates of population parameters based on the data collected

periodicity in outcropping  
of joints

outcrop

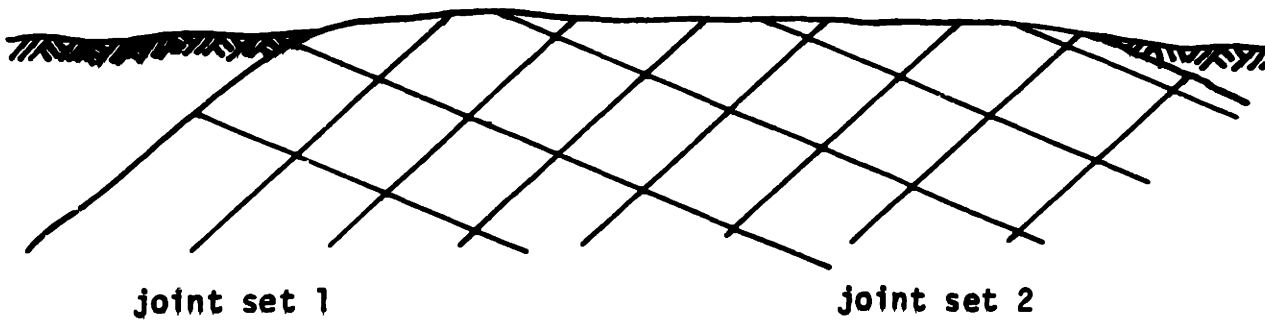
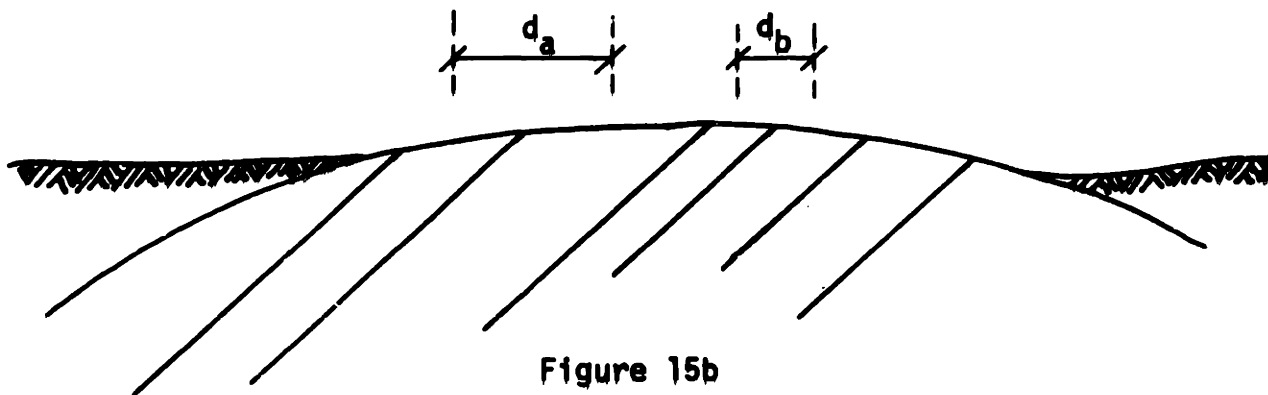


Figure 15a



Differences in individual joint spacings lead to differences in the probability that a randomly located point is closer to one joint than another (e.g.,  $d_a \geq d_b$ )



and the variance in these estimates can be evaluated using the methods of classical sampling theory discussed in the appendix. Tests for homogeneity in the joint population across the site will be discussed in Section 7.2.1.

#### 7.1.4 Intensity of Jointing

Intensity is defined as some measure of the number of frequency of joints in a unit of rock mass. Examples are the number of joints per unit volume, the total area of joint surface per unit volume, and the "fracture frequency." The latter, or its reciprocal (i.e., mean spacing), is perhaps the most common. Fracture frequency is easy to comprehend and easy to treat in design. Furthermore, empirical correlations relating fracture frequency to the engineering properties of rock masses have been developed and are widely available (e.g., Deere, et al, 1966).

Sampling plans for spacing data must conform to the four rules of probability sampling, just as any other sampling plan must if precisions of the estimates are to be evaluated: If all spacings do not have an equal probability of being sampled, appropriate weights must be applied. For the same practical reasons that cluster sampling was preferred for collecting joint orientation data, so it is also preferred for spacing data. There is no reason that spacings should not be sampled on the same outcrops as orientation, and analyzed in the same ways; the only differences being

that spacing is usually measured separately on each joint set and the probability of a certain spacing being sampled does not depend on the orientation of the outcrop. An acceptable plan for measuring spacing data is the following:

1. Observe the average direction of the joint traces on the outcrop.
2. Randomly locate a line segment on the outcrop, oriented perpendicularly to this direction, which has a length much greater than the mean spacing so that boundary effects may be neglected.
3. Measure the average spacing between pairs of adjoining joints within the set being considered, which intersect the line of 2.
4. Correct the data for the angle between the average pole direction and the measurement line.
5. Repeat for each joint set.

Almost always, the only description of the fracture-frequency distribution of importance to the engineer is the average -- the state-of-the-art of incorporating intensity data in design decisions is not well developed.

An interesting problem in sampling fracture-frequency data, however, does arise when the discrimination of joint sets is not obvious by inspection. From the distribution of orientation measurements we can ascertain the relative frequencies of the several joint sets existing at a site, and also their mean orientations.

Let  $E_j(\cdot)$  = the expectation operator over  $j$

$\beta_{ij}$  = the angle between line  $L$  and the normal to the  $j$ th. member of joint set  $i$

$q_i$  = the probability that a randomly selected joint belongs to set  $i$

$l$  = the length of the sampling line  $L$

$d_{ij}$  = the spacing of the  $j$  th. member of joint set  $i$

If we assume  $d_{ij}$  and  $\beta_{ij}$  to be mutually independent, the expected number of joints of set  $i$  intersecting  $L$  is

$$E(n_i) = l E_j(\cos \beta_{ij}) E_j(1/d_{ij}) \quad 13$$

However, if  $n$  is the total (observed) number of joints of all sets intersecting  $L$ ,

$$E(n_i) = n \frac{q_i E_j(\cos \beta_{ij})}{\sum_i q_i E_j(\cos \beta_{ij})} \quad 14$$

Combining 13 and 14

$$E(1/d_{ij}) = \frac{n q_i}{l \sum_i q_i E_j(\cos \beta_{ij})} \quad 15$$

The fracture frequency along a line  $A$ ,  $ff_A$ , is the number of joints intersected per unit length of  $A$ . If  $\gamma_{ij}$  is the angle which  $A$  makes with the  $j$  th member of joint set  $i$ , and  $l_a$  is the length of  $A$ ,

$$E(ff_A) = E_j(\cos \gamma_{ij}) E_j(1/d_{ij}) \quad 16$$

Taking  $A = L$ , and combining equations 15 and 16, the expected fracture frequency along line  $L$  due to joint set  $i$  is,

$$E(\cos \beta_{ij}) = \frac{n q_i E_j(\cos \beta_{ij})}{\sum q_i E_j(\cos \beta_{ij})} \quad 17$$

Since  $E(\cos \beta_{ij})$  may be evaluated from the orientation distribution we can solve for  $E(\cos \beta_{ij})$ . One should note, however, that  $E(\cos \beta_{ij})$  does not equal  $\cos E(\beta_{ij})$ , and therefore may not be evaluated simply by taking the cosine of the average orientations.  $E(\cos \beta_{ij})$  must be evaluated by summing  $\cos \beta_{ij}$  for all the sampled joints in set  $i$  and dividing by the number of measurements.

## Section 7.2 Inferring the Nature of Joint Populations

We have thus far only considered plans for sampling joint populations; we now turn our attention to drawing inferences from the data collected. In the first half of this section we consider statistical techniques for drawing inferences from raw data. Certain of these techniques are approximate, but nevertheless appropriate in the light of the state-of-the-art of rock mechanics design. In the latter half we discuss distribution models for joint populations. Despite that analytical models are not now widely applicable to design, the increasing tendency to apply them (as reflected in the literature) and the growing sophistication of rock mechanics justify their discussion.

Throughout our discussion we assume orientation distributions to be the same at every point in space (i.e., "stationarity"). We

may conclude that joint populations in two regions of a site are different, but within one population we assume there are no spatial trends. This is not precisely true, as we discussed with regard to nested analysis of variance, but it simplified analysis. We further assume that orientations are not autocorrelated. That is, the orientation of a given joint is independent of the orientation of adjacent ones.

### 7.2.1 Methods Not Using Distribution Models for Joint Populations

For most design problems the engineer requires information on four properties of jointing

- 1) The central orientation of each joint set
- 2) The variance of each joint set
- 3) The average intensity of each joint set
- 4) Whether or not each joint set is homogeneous across the site, and the relationship between variance and volume of rock.

The usual first step after collecting data is to plot it on some spherical projection which allows it to be displayed in two dimensions. Throughout this work we shall use equal area projections (e.g., see Vistelius, 1966) because in such projections equal areas of the spherical surface plot as equal areas on the net, making correction factors unnecessary (Figure 16). Other spherical projections are discussed in the literature (e.g., see Phillips, 1972) as well as non-spherical projections (e.g., Pincus, 1961; Robertson

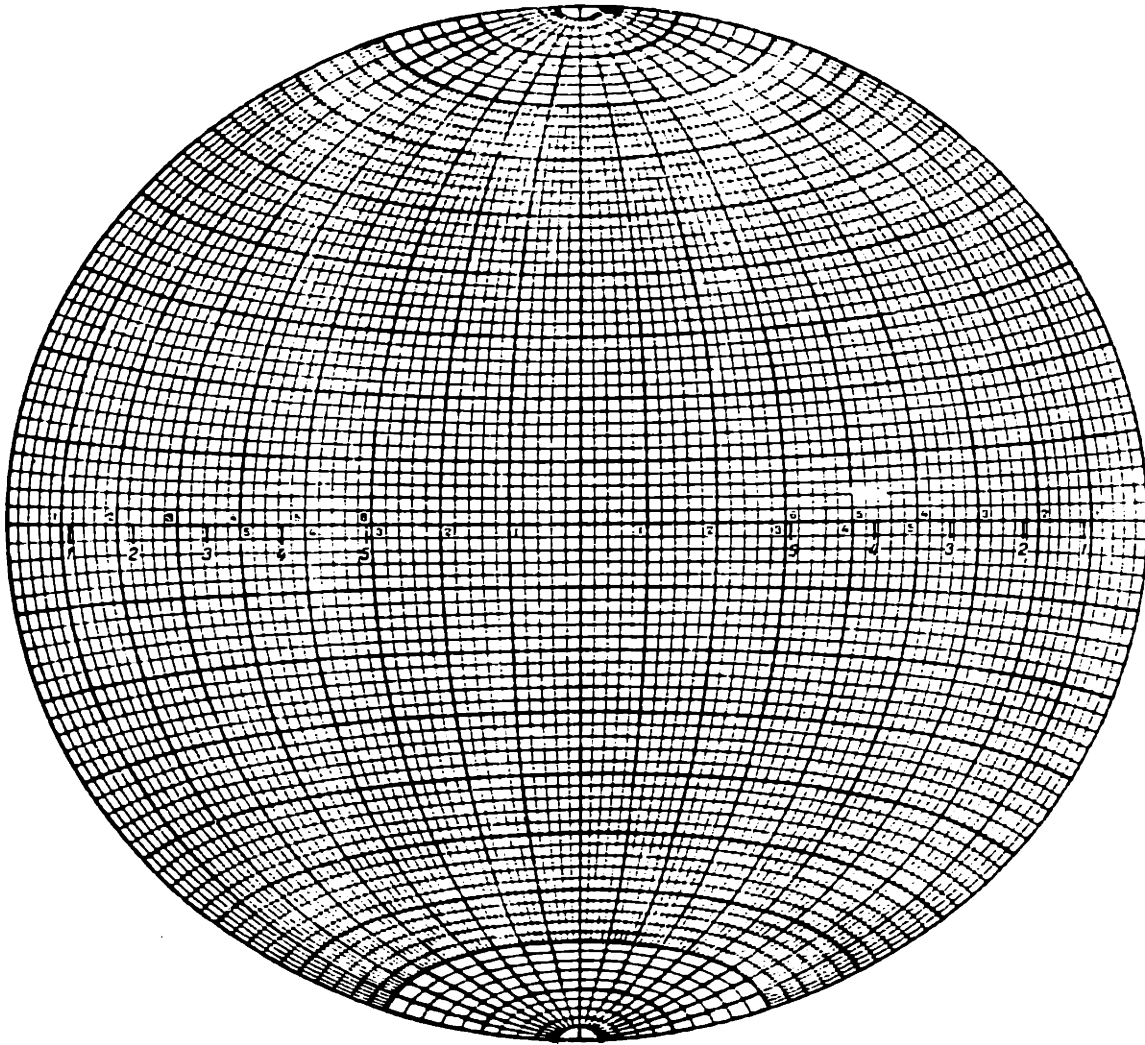


Figure 16: Equal-area net

and Piteau, 1970).

Poles to sampled joints might appear in projection as shown in Figure 17. One begins his analysis by measuring densities of poles per unit area and constructing contours.

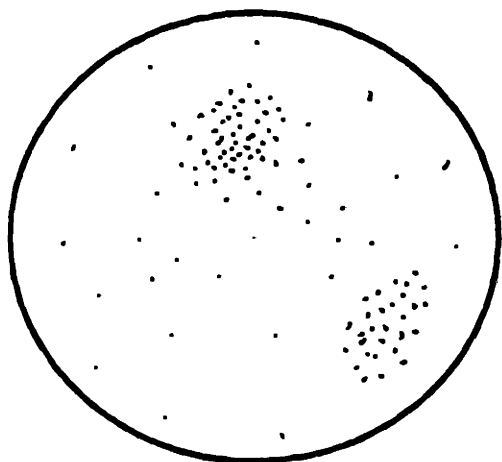


Fig. 17

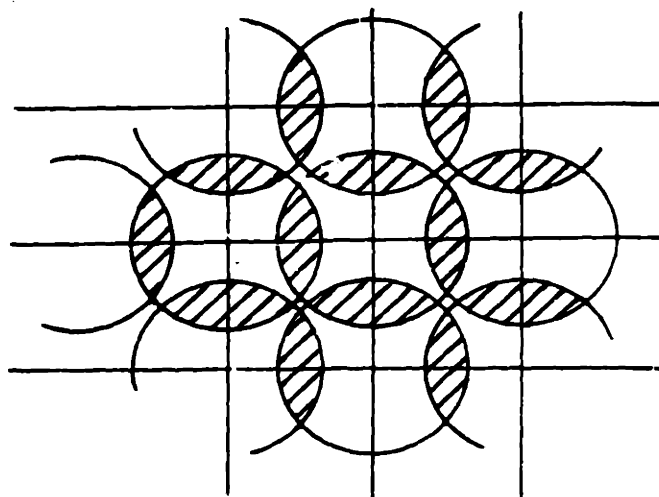


Fig. 18

The common method of measuring density uses the "unit counting circle" which is a circular template with an area equal to 1% of the area of the projected net. A mesh of density measurements is generated by successively placing the center of the circle at the intersections of a square grid overlay and counting the number of poles within the circle. Normalizing this number for the total number of poles gives a grid of the percent of total poles per 1% net area. These percentages are used as density measurements and contoured, usually in 5% increments. This process is normally performed by hand, but computer programs are available (Warner, 1969).

This contoured plot is in a way analogous to the histogram; however, since some points are counted more than once (Figure 18) the volume under the contours is greater than 1.0 (to change this one must normalize by the number of counted poles). The most attractive aspect of the unit counting circle method is that engineering geologists are generally familiar with it; there is no innate reason for this definition of density.

An alternate method of determining density, and one which we prefer, is to divide the polar plot into a number of quadrates of equal area and count the poles in each. This is accomplished by placing a transparent grid like that shown in Figure 19 over the plot, and the result is strictly analogous to the univariate histogram. Actually, the only fundamental difference between these methods is that the unit counting circle counts some poles twice, whereas the second method counts each point only once. The second method has advantages when statistical operations are applied, and is easier to use, but contours of relative density determined by both methods will be approximately the same.

#### Existence of a Preferred Orientation

Before further statistical analysis is performed one must be satisfied that preferred orientations do exist in the total joint population. Often preferred orientation may be inferred by inspection, but if not (e.g., see McMahon, 1968), statistical tests are required. Although we can never definitely conclude that a set of



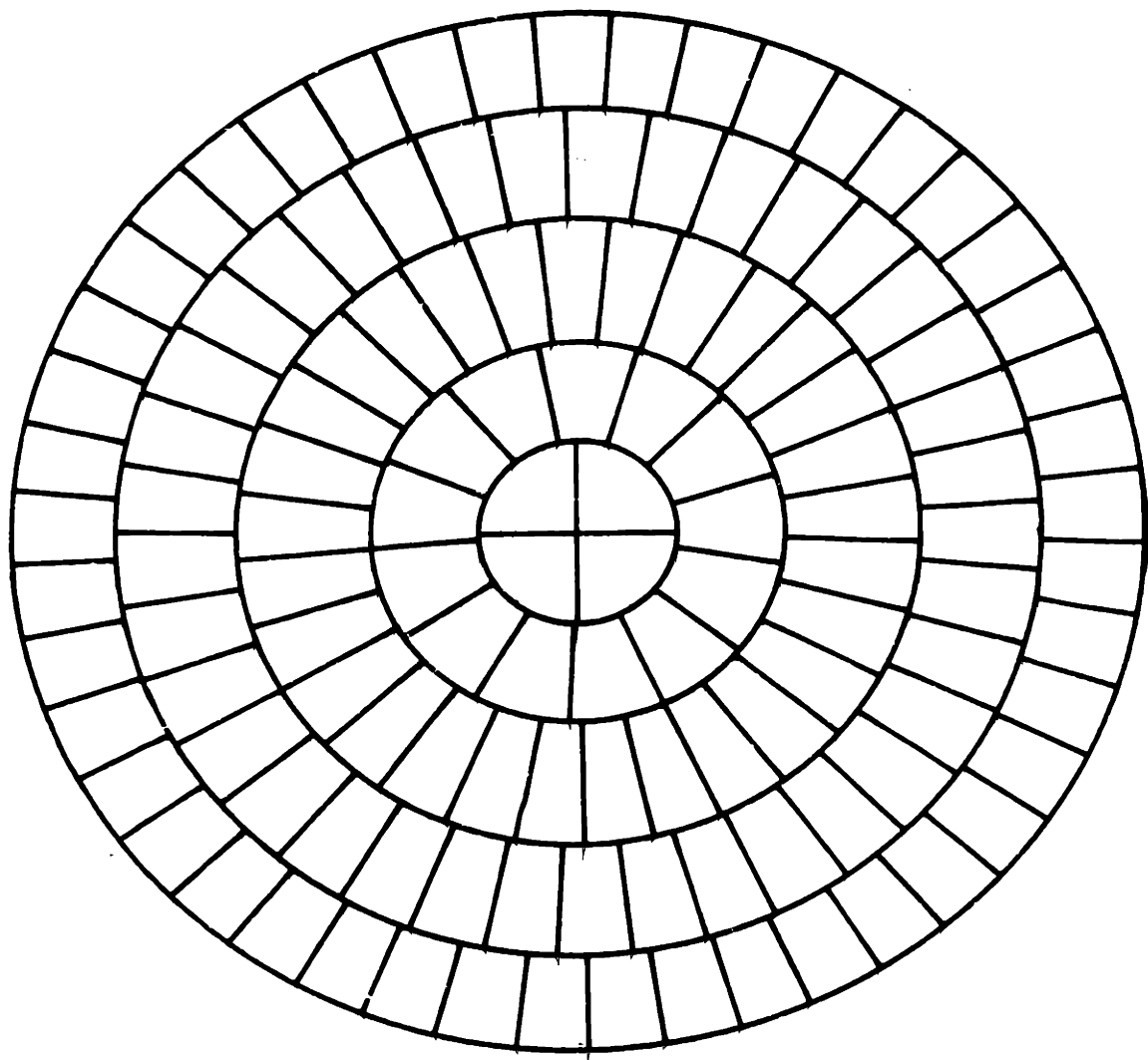


Figure 19

Overlay of equal area quadrates

observations was taken from a random population, we can see if it behaves as such when subject to certain statistical tests. We do this by comparing sample statistics with their theoretical distribution for random populations.

Were orientation random, the distribution of poles to sampled joints (on an equal-area projection) would also be random, and hence would have the following properties:

- 1) Any pole would have an equal probability of occurring at any point on the projection.
- 2) Each small element of area would have the same probability of containing a pole as any other small element of the same size.
- 3) The position of any pole would be independent of the position of any other pole.

If we denote the probability of a pole being located in a given unit area by  $p$ , the probability of exactly  $k$  poles being located in a unit area is,

$$p_k(k_0) = \binom{n}{k_0} p^{k_0} (1-p)^{n-k_0} \quad 18$$

where  $n$  is the total number of poles, and  $\binom{n}{k_0}$  is the number of combinations of  $n$  things taken  $k_0$  at a time. This is simply the expression of the binomial distribution. If we divide the unit area into equal subareas small enough that the probability of two poles lying in the same subarea is negligible, the expected number of poles per subarea,  $np$ , is small and the binomial distribution of 18 may be approximated by the Poisson distribution as  $n \rightarrow \infty$ ,

$$Pr(k_0) = \frac{e^{-\lambda} \lambda^{k_0}}{k_0!} \quad 19$$

where  $\lambda = np$ . The advantage of the Poisson approximation is only computational: when  $n$  is large, the number of computations associated with the binomial distribution is large. We will discuss two sets of statistical tests for randomness based on the Poisson distribution: nearest neighbor tests, and direct comparison tests.

#### Nearest Neighbor Methods:

Nearest neighbor techniques treat the distribution of spacing between points and their closest neighboring point, and have been extensively developed in the study of plant ecology. The advantage of the nearest neighbor techniques is that they do not depend on artificially dividing the data into quadrates as do most other methods. This eliminates the effects of where boundaries are placed and of how many quadrates are used.

For a Poisson distribution of points the pdf of distance from an arbitrary point  $z_0$  to its nearest neighbor,  $r_1$ , is (Skellum, 1952; Kendall, 1963) (Figure 20a),

$$f_r(r_1) = 2\pi\lambda r_1 e^{-\lambda\pi r_1^2} \quad 20$$

Since  $z_0$  is an arbitrary point, the distribution of first neighbor spacing for any set of points can be determined empirically by measuring the distance from each point to its closest neighbor and the result compared with (20) using a goodness-of-fit test.

Expected number of reflexives at the  $n^{\text{th}}$  level, given that points are from a random distribution in space

<u>Level of nearest neighbor</u>	<u>Expected number of reflexives</u>
1	.6215
2	.3863
3	.2401
4	.1492
5	.0927
6	.0576
7	.0358
8	.0228
9	.0138
10	.0086
11	.0053
12	.0033

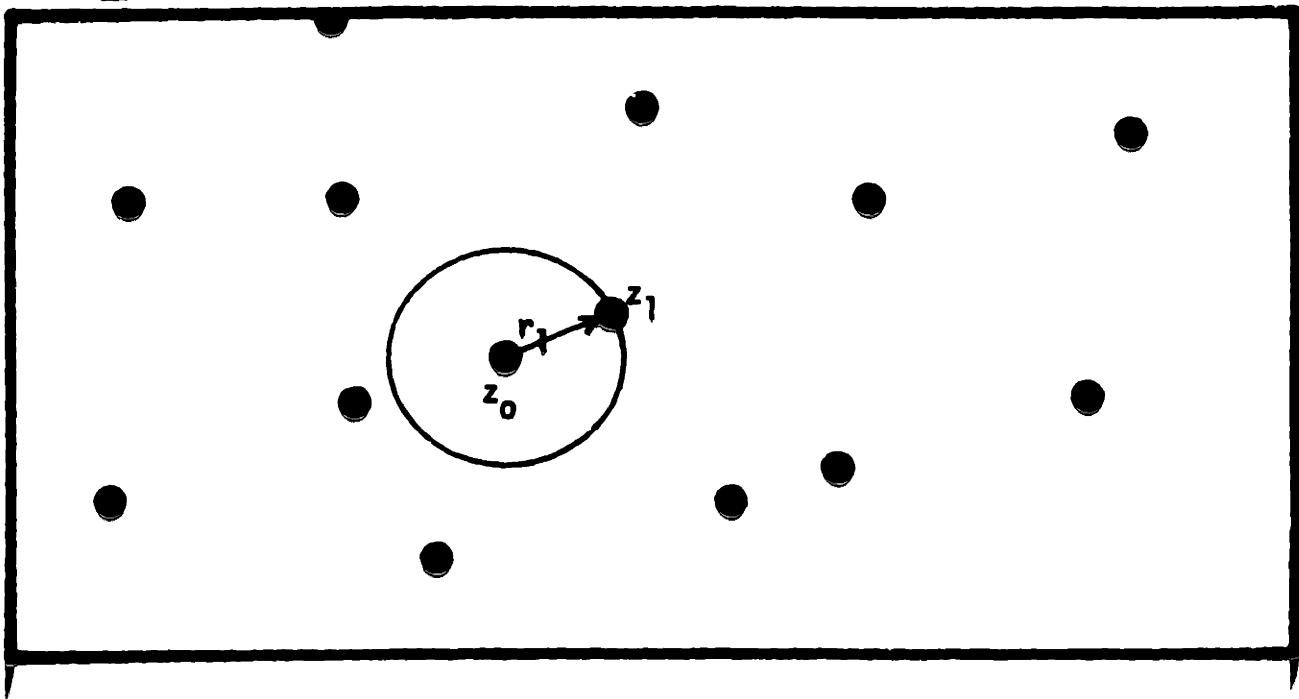


Figure 20a

Nearest neighbor spacing of point  $z_0$  is  $r_1$

The disadvantage of this method is that measuring the distances to the nearest neighbor of several points and calculating the function  $f_r(r_1)$  is time consuming. Testing for frequency of "reflexivity", which is the property that two points are mutual closest neighbors, saves time (Figure 20b).<sup>\*</sup> If the observed number of reflexives is significantly greater than the expected number, one concludes that a preferred orientation exists. Significantly too few reflexives indicate points which are too evenly spaced to be random and has no meaning for pole diagrams: we are concerned with a so-called "one-sided" test.

If  $z_0$  and  $z_1$  are points in  $k$ -space,  $(z_0, z_1)$  are said to be "reflexive" at the  $n$ th level if  $z_1$  is the  $n$ th nearest neighbor of  $z_0$ , and  $z_0$  is the  $n$ th nearest neighbor of  $z_1$ . Clark (1956) has shown that the probability of two points being reflexive, given that they are samples from an, in fact, random spatial distribution, is

$$\text{Pr}(\text{reflexivity}) = \int_0^{\infty} \frac{k}{(n-1)!} \left[ \frac{\lambda \pi^{k/2}}{\Gamma(k/2 + 1)} \right] r^{(kn-1)} e^{-2r^k \lambda L} dr \quad 21$$

where,

$$L = \left[ \frac{\pi^{k/2}}{\Gamma(k/2 + 1)} - \frac{\pi^{(k-1)/2}}{\Gamma(k/2)} \int_{1/2}^1 (1-x^2)^{(k-1)/2} dx \right] \quad 22$$

---

<sup>\*</sup> It also allows testing of higher order randomness, were that important. For example, if poles are not random at first neighbor spacing, they might still be random at 2nd neighbor spacing (i.e., the points might be clumped).

and  $\Gamma$  is the Gamma function:

$$\Gamma(x) = \int_0^{\infty} e^{-u} u^{x-1} du \quad 23$$

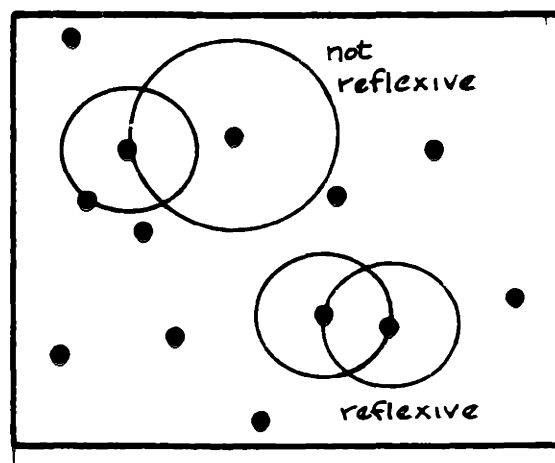
Solutions to this equation for  $k = 2$  are given in Table 1.

Since the probability of any pair of points being nearest neighbor reflexive is constant, given that the points are from a random spatial distribution (e.g.,  $p = 0.6215$  in two dimensions), the actual number of reflexives in  $N$  pairs follows the binomial distribution. For  $Np$  large, this can be approximated by a normal distribution with mean  $Np$  and variance  $Np(1-p)$ . A test for significant departure from the expected number,  $Np$ , can be made by comparing the discrepancy

$$\Delta = \left\{ \begin{array}{l} \text{observed number} \\ \text{of reflexives} \end{array} \right\} - Np \quad 24$$

with tables of the normal distribution (e.g., a discrepancy greater than  $2(Np(1-p))^{1/2}$  would only occur about 2% of the time, were the points actually random).

Figure 20b



### Direct Comparison with Poisson Distribution

A second test for randomness involves directly comparing observations with the Poisson distribution by a goodness-of-fit test, rather than treating a derived property such as nearest neighbor spacing. This method is easy, since it uses data directly from the histogram constructed to contour density, but its results depend on the exact location of arbitrarily placed partitions.

The expected number of poles per quadrate equals the average density times the quadrate area

$$E(x) = \lambda (\text{quadrate area}) \quad 25$$

This can be compared with the observed number in each quadrate using Fisher's index of dispersion (Fisher, et al, 1922),

$$\sum_i \frac{(x_i - \bar{x})^2}{\bar{x}} \quad 26$$

which is distributed as  $\chi^2$  with  $k-2$  degrees of freedom ( $k-1$ -number of parameters estimated from data).

To use the test, the value of the index of dispersion is calculated from data and compared with tabulations of  $\chi^2$ . Tabulated values of  $\chi^2$  for the  $\alpha$  % confidence level are ones which would be equalled or exceeded with only probability  $\alpha$  if the points were indeed randomly distributed.

A difficulty in using index of dispersion is that the statistic is only approximately distributed as  $\chi^2$ , and the approximation is in question for quadrate frequencies less than about 3 (Cochran, 1936; Neyman and Pearson, 1931; Shanary, 1936). Quadrates should be

grouped if their frequencies are less than or equal to 3. Authors in geologically related fields (e.g., Hazen, 1967; Vistelius, 1966; Chayes, 1945) neglect to mention this limitation in their discussions.

We could alternately use Fisher's index of dispersion by comparing the observed frequency of quadrates containing a given number of poles with the frequency predicted by the Poisson distribution.

Neither of these direct comparison tests take into account the spatial distribution of the quadrate numbers, and both are sensitive to the exact boundaries chosen for quadrates. Artificial point distributions can be constructed which satisfy both of these tests, yet which are definitely not random (Figure 22). This is why the non-dependence of the nearest neighbor techniques on quadrate partitioning is so favorable.

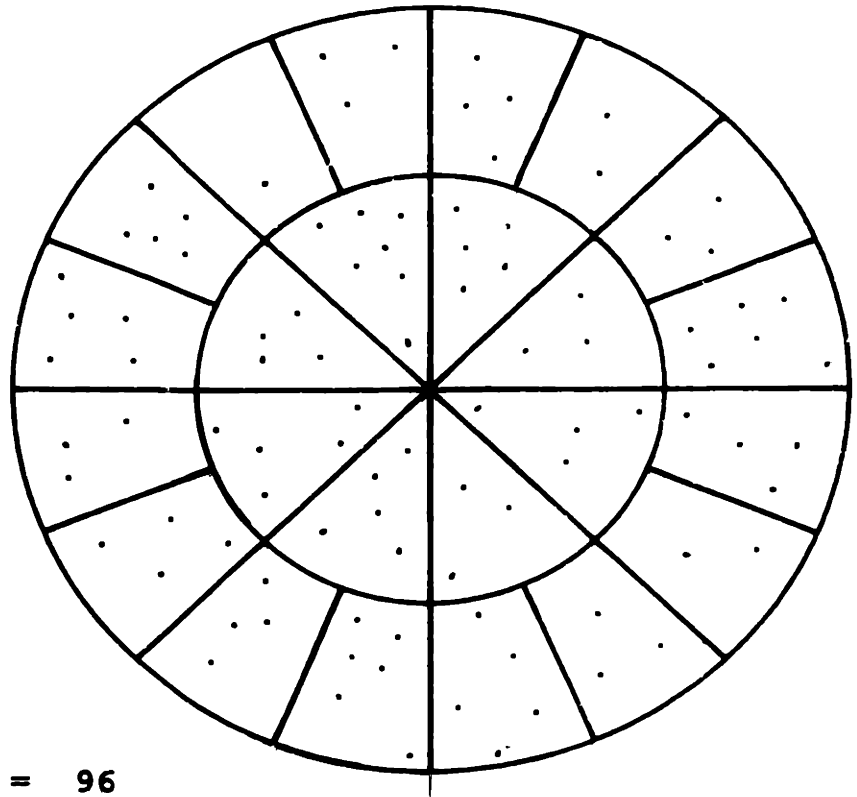
"Student" (1906) has proposed a method based on correlation between adjacent quadrates which leads to a conclusion of non-randomness for the example of Figure 22, but we will not discuss it here.

Clapham (1936) has suggested using the ratio of variance of the number of points per quadrate to the mean as a test for random distribution. If the points are Poisson distributed, this ratio should equal unity. If the sample variance-to-mean ratio is significantly different from 1, then the hypothesis of random distribution can be rejected. Unfortunately, this ratio is extremely sensitive to location and size of quadrates, and Miller and Kahn (1962) have shown that the ratio can fluctuate greatly for the same



## Index-of-Dispersion test for randomness

Question: Are the 96 poles shown on the polar plot below randomly distributed?



1--  $k = 24$

2--  $n_i = 96$

3--  $\frac{(O-P)^2}{p^2} = 13.5$

4--  $\chi^2_{(23)} (\alpha=5\%) = 35.172 \geq 13.5$

5-- Conclude that points are randomly distributed

Figure 22: An example of a point distribution which even by inspection is non-random, yet which does produce a significant index of dispersion as an indication of non-randomness.

1	2	3	2
2	4	5	3
3	6	7	3
2	4	5	3
0	1	2	1

$n$	$k_p(n)$	obs'd
0	0.39	4
1		
2	4.4	5
3	4.4	4
4	5.3	3
5	3.4	
6		4
7		
8		

$$\sum \frac{(\text{obs'd.} - \text{pred.})^2}{\text{pred.}} = 0.227$$

$$\chi^2_{(4)}(95\%) = 1.15$$

$\therefore$  Conclude: random

set of data by changing quadrature size. Other tests should be used in preference to the variance-mean ratio.

### Means and Variances of Joint Distributions

If preferred orientation is concluded to exist (i.e., if the pole diagram is concluded non-random) we face the problem of determining the mean and variance of joint sets.

Estimates of mean orientation and variance, and the precision of these estimates, may be determined using classical sampling theory as discussed in the appendix. These values depend on the sampling plan, and must be based on weighted samples if appropriate. However, when computations are performed by hand and approximate results are satisfactory, graphical techniques for determining central orientation may be desired.

If the distribution of orientations within a joint set is symmetrical through its center, mean and modal directions of the distributions will be the same. The latter can easily be determined by inspection, since it is the direction of maximum density. For normal use, inaccuracy introduced through determining the mode by inspection is negligible. Even if the distribution is not symmetric (and hence the mean and the mode not the same), the modal direction may still be of more interest than the mean since it is the orientation of the largest single group of joints.

If the mean of a distribution is desired, it can be obtained approximately by a graphical procedure. Considering a univariate

distribution, the mean is the value of the variate about which the integral of  $f_x(x_i) d(x_i)$  equals zero (Figure 23), and hence is analogous to center of mass.

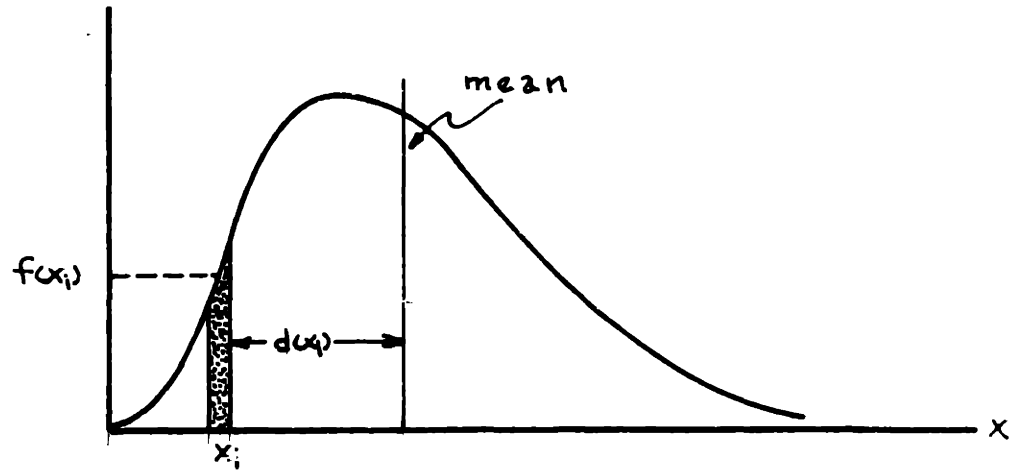
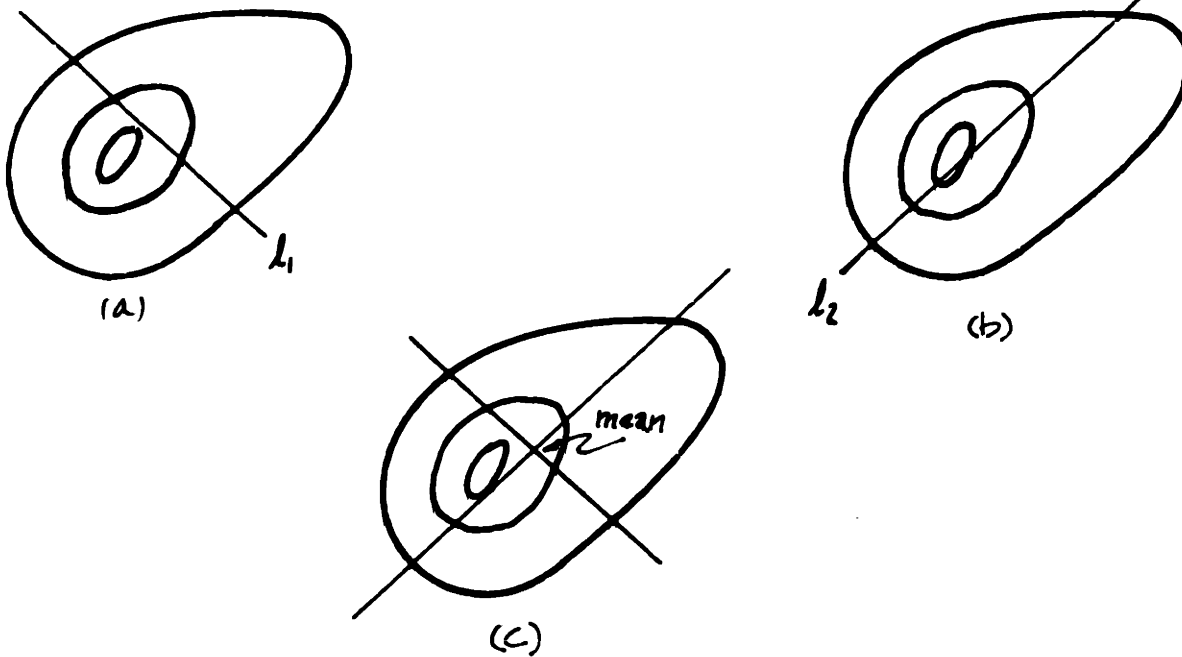


Fig. 23

In Figure 24 the line  $l_1$  is a line such that the integral of the frequency times distance from  $l_1$  equals zero. Since any line through the center of mass has this property (and no other line), line  $l_1$  must pass through the mean. By finding another line,  $l_2$ , having the same property, the mean can be determined from their intersection. The procedure for doing this is a trial and error one as follows:

Fig. 24



- 1 - On the mesh of density values obtained by overlaying a grid and counting poles place a line in a convenient direction which passes through an initially guessed location of the mean.

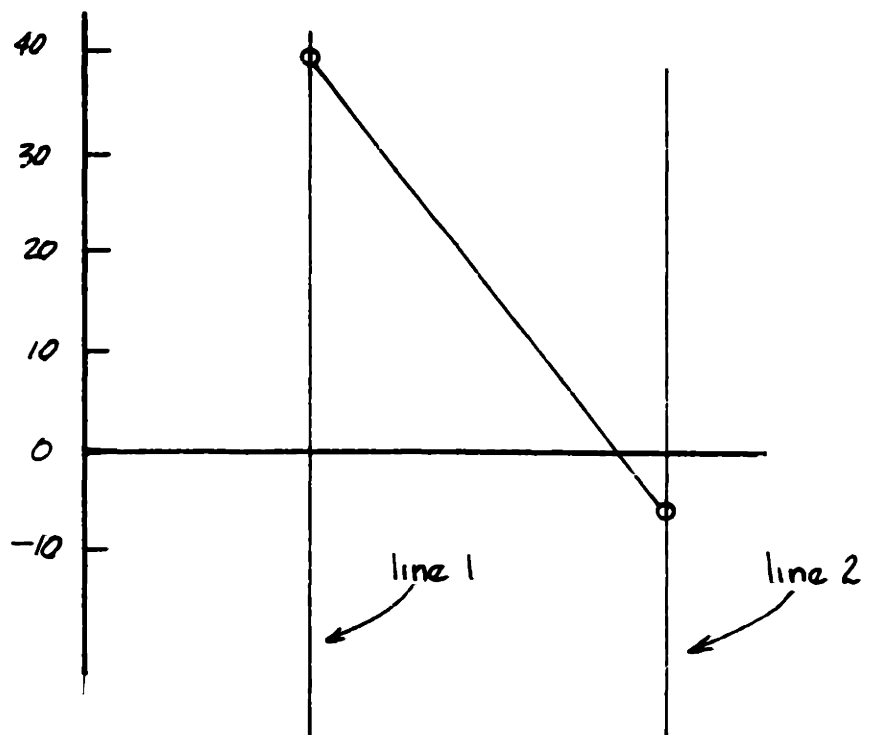
1	3	2	1	0
2	9	7	4	1
3	15	11	9	2
5	15	13	11	4
3	13	9	6	1
14	55	42	31	8
	#1	#2		

- 2 - Sum the columns of density measures parallel to the line and find the sum of their magnitudes  $\times$  distance (i.e., their moments).

line 1	-1.5	x	14	=	-21
	-0.5	x	55	=	-27.5
	0.5	x	42	=	21
	1.5	x	31	=	46.5
	2.5	x	8	=	20
					<hr/>
					+ 37.0

3 - Move the line in the direction which will lessen this sum, and continue the process until the line for which sum = 0 is found. Since the sum is a linear function of the location of the line, the desired location can be found by interpolation.

line 2	-1.8	x	14	=	-25.2
	-0.8	x	55	=	-44.0
	0.2	x	42	=	8.4
	1.8	x	31	=	37.2
	1.2	x	8	=	17.6
					<hr/>
					-6.0



4 - Perform a similar operation in a second convenient direction.

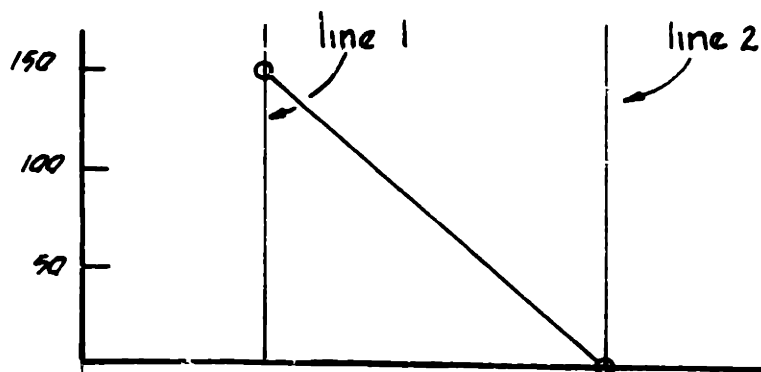
1	3	2	1	0	$\Sigma = 7$	
2	9	7	4	1	23	line 2
3	15	11	9	2	40	line 1
5	15	13	11	4	48	
3	13	9	6	1	32	

line 1

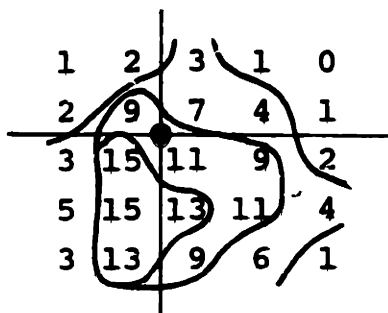
$$\begin{array}{rcl}
 -1.5 \times 7 & = & -10.5 \\
 -0.5 \times 23 & = & -11.5 \\
 0.5 \times 40 & = & 20 \\
 1.0 \times 4 & = & 72 \\
 1.5 \times 32 & = & 48 \\
 \hline
 & & 149
 \end{array}$$

line 2

$$\begin{array}{rcl}
 -2.5 \times 7 & = & -17.5 \\
 -1.5 \times 23 & = & -34.5 \\
 -0.5 \times 40 & = & -20 \\
 0.5 \times 48 & = & 24 \\
 1.5 \times 32 & = & 48 \\
 \hline
 & & 0
 \end{array}$$



5 - The intersection of these two lines in the mean orientation of the distribution.



The mean may also be computationally determined by normalizing the vector sum of the sample measurements. If  $\underline{R} = (R_1, R_2, R_3)$  denotes the mean orientation, and  $(l_i, m_i, n_i)$  denotes the direction cosines of the  $i$ th sample element, then

$$\underline{R} = \left( \frac{\sum l_i}{N}, \frac{\sum m_i}{N}, \frac{\sum n_i}{N} \right) \quad 27$$

The difference between this direction and the mean chosen by inspection is frequently small, and most problems do not merit the computational effort associated with evaluating the vector mean.

The variance in the orientation distribution is more involved. The variance of a population of orientation vectors about the mean orientation is (Kiraly, 1969)

$$S_m^2 = \frac{1}{N} (1 - |\cos \theta_i|)^2 \quad 28$$

where  $\theta_i$  is the angle between the  $i$ th pole and the mean vector. The angle in degrees corresponding to the standard deviation  $S_m$  is

$$\theta_{S_m} = 2 \arccos (1 - S_m) \quad 29$$

Although the combination of mean orientation and variance does not completely describe a population of joint orientations, it is often sufficient for design. For example, Scheidegger (1960) has attempted to use such descriptions, combined with spacing data, to estimate the permeability tensor in jointed media.



### Tests for Homogeneity of Joint Populations

Were we to take several samples from the same population, the estimates of mean and variance made from each would not be identical. So an engineer in taking samples from each of two or more locations at a site will not obtain identical estimates even if the joint distribution is everywhere the same. He must use statistical tests to infer similarity of the parent populations from which his samples have been taken.

The property of two or more populations being the same is called "homogeneity." The classical way of testing for homogeneity, as with testing for randomness, is to derive the distribution of certain sample statistics assuming the populations to be the same, then evaluating the probability of a value of that statistic as large or as small (as may be the case) as the one calculated from samples. Traditionally, tests for the homogeneity of means are based on statistics distributed as "Student's"  $t$  for one-dimensional variables, or Hotelling's  $T^2$  for multidimensional variables. When the number of observations is large these methods require considerable computational effort.

Graphical methods can be used to test homogeneity, and require less effort than the tests above; however, they generally require dividing the data set into quadrates and have the disadvantage that their results depend on the exact partitioning of quadrates.\*

---

\*The overlay of Figure 19 is an excellent vehicle for partitioning orientation data. If this overlay is used in constructing contours of density, the data is put directly in a useable form for the methods discussed here.

The first technique considers the sum of the squared differences between the observed number of poles in corresponding quadrates on two diagrams being compared. We will derive a distribution of this statistic assuming homogeneity, then compare values obtained from samples with the theoretical distribution.

Let the total number of joints sampled at locations  $i$  and  $j$  be  $n_{.i}$  and  $n_{.j}$ , respectively, and the number of poles falling in the  $r$ th quadrate of each,  $n_{ri}$  and  $n_{rj}$ . Assuming homogeneity, the expected proportion of poles falling in corresponding quadrates should be the same, or

$$E \left[ \frac{n_{ri}}{n_{.i}} - \frac{n_{rj}}{n_{.j}} \right] = E(\Delta_r) = 0 \quad \text{for all } r \quad 30$$

Let the probability that a joint sampled at location  $i$  will have a pole falling in quadrate  $r$  be  $p_{ri}$ . By the binomial theorem, the distribution of the number of poles falling in  $r$ ,  $n_{ri}$ , must be

$$P_{n_{ri}}(n_o) = \binom{n_{.i}}{n_o} (p_{ri})^{n_o} (1-p_{ri})^{n_{.i}-n_o} \quad 31$$

which for  $n_{.i}$  large and  $p_{ri}$  not very small may be approximated

by a normal distribution with mean  $n_{.i} p_{ri}$  and variance  $(n_{.i} p_{ri} (1-p_{ri}))^{1/2}$ . Since both  $n_{ri}$  and  $n_{rj}$

are approximately normally distributed, the quantity  $\Delta_r$  is also approximately normally distributed, being the difference of two normal r.v.'s. The variance of  $\Delta_r$  may be estimated by

$$\text{Var}(\Delta_r) = \hat{\sigma}^2(\Delta_r) = \sum_{r=1}^k \left[ \frac{(\Delta_r - E[\Delta_r])^2}{k} \right] = \sum_{r=1}^k \Delta_r^2 / k \quad 32$$

where  $k$  is the number of quadrates into which the pole diagram is divided. So the quantity  $\Delta_r/k$  is a standard normal variate (i.e., a normal r.v. with mean zero and variance 1).

The sum of  $k$  independent squared normal variates is distributed as  $\chi^2$  with  $k$  degrees of freedom. Because the  $\Delta_r$  are constrained by  $\sum_r \Delta_r = 0$ , and one parameter is estimated, there is a loss of 2 d.o.f., and the statistic

$$\sum_{r=1}^k \left(\frac{\Delta_r}{\sigma}\right)^2 \quad 33$$

is distributed as  $\chi^2$  with  $k-2$  d.o.f.

A test for homogeneity can be made by calculating the quantity 33 and comparing its magnitude with tabulated values of  $\chi^2_{(k-2)}$ .

If the calculated value of 33 is significantly too large, one concludes that the two joint distributions are not actually the same.

The second test is also based on quadrate frequencies, but has the advantage that it may be extended to the simultaneous comparison of more than two pole diagrams. One can show (Cramer, 1955) that an index of dispersion test may be applied even when predicted frequencies are substituted by their maximum likelihood estimates from sample data (although this causes a loss of a d.o.f. per estimate).

Therefore, we can obtain estimates of predicted quadrate frequencies, calculate the index of dispersion, and compare it to tabulated values of  $\chi^2$ .

The maximum likelihood estimate of the predicted number of poles in quadrate  $r$  is,

$$n_{r.} = \frac{\sum_j n_{rj}}{n} \quad 34$$

where  $n = \sum_i n_{.i}$  . Rao (1965) shows that the quantity

$$\sum_{i=1}^m \sum_{r=1}^k \frac{\left[ n_{ri} - \frac{n_{r.} n_{.i}}{n} \right]^2}{\frac{n_{r.} n_{.i}}{n}} \quad 35$$

which is simply the index of dispersion, is distributed as  $\chi^2$  with  $(m-1)(k-1)$  d.o.f. For values of 35 which are significantly too high one concludes that homogeneity does not exist. The same comments made with respect to quadrature frequency in testing randomness by index of dispersion (p.415) apply here as well.

When treating joint populations one should be mindful that different joint sets are often formed at different points in time, and by different stress systems; which means that one set may be homogeneous across the site while others are not. Joint populations should always be tested set by set.

### Number of Joints to be Sampled

We can approach the problem of the number of joints to be sampled in two ways: variance in the estimates of population parameters or empirically determined heuristic rules.

The standard error of the mean for random samples is (Appendix)

$$\sigma_{\bar{x}} = \sqrt{V(\bar{x})} = s / \sqrt{n} \quad 36$$

where  $s$  is the standard deviation of the population and  $n$  is the number of samples. For a typical joint set  $s$  might be on the order of  $10^\circ$ , so a sample size of  $n=100$  would lead to a standard error of the mean of about  $1^\circ$ . A standard error of  $1^\circ$  would mean that 99% of the time the estimate of the mean orientation would be within  $3^\circ$  of the actual orientation -- about the same magnitude as random measurement error and certainly precise enough for most applications. A sample of 200 would only reduce the standard error to about  $0.7^\circ$ .

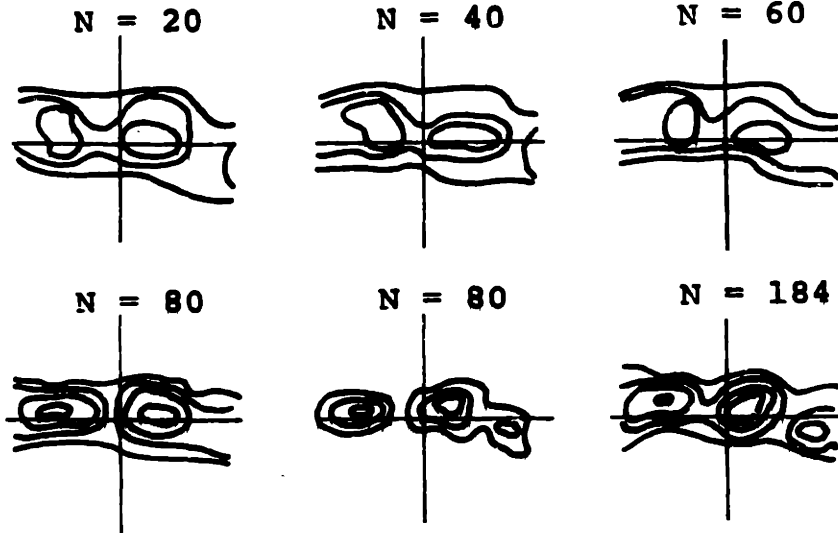
Two empirical studies reported in the literature were performed by Pincus (1961) and Larsson (1952). Both investigated changes in the sample characteristics as the number of measurements increased. Pincus (Figure 25) concluded that the distribution did not change appreciably above  $n=60$ , while Larsson (Figure 26) concluded that  $n \geq 100$  was necessary.

From these considerations a sample size of 100 to 150 should be taken for each diagram. This conclusion corresponds to sizes normally recommended in the literature (e.g., Müller, 1933, recommends 150 to 200, and reports that Stiny recommended 80 to 100).

### 7.2.2 Mathematical Models for Orientation Distributions

One observes an emerging trend in the engineering literature toward the use of mathematical functions to model joint orientation distributions. Such models serve as smooth, continuous summaries of actual observations and will probably enjoy increased use as the analytical capabilities of rock mechanics increase. There are several

## Location 1



## Location 2

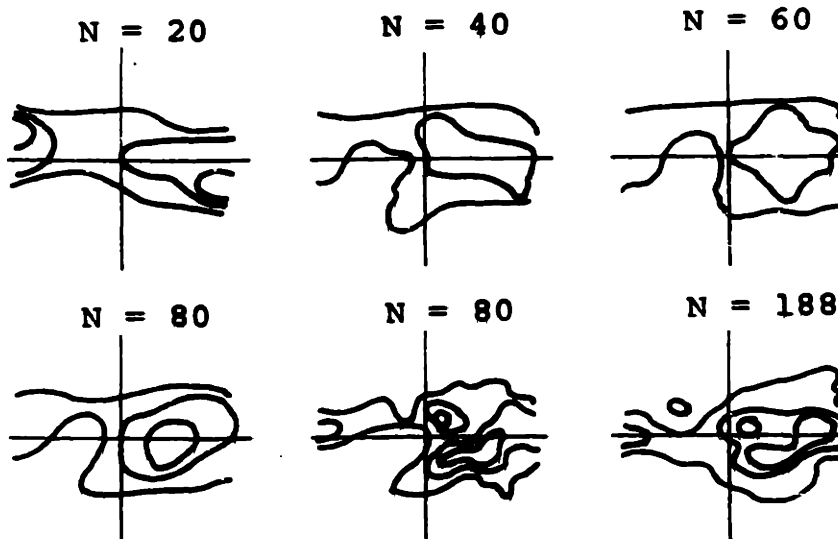


Figure 25

Changes in graphical orientation distribution as sample size is increased (after Pincus, 1961):  
 Contours indicate relative density of poles,  
 N indicates number of joints measured

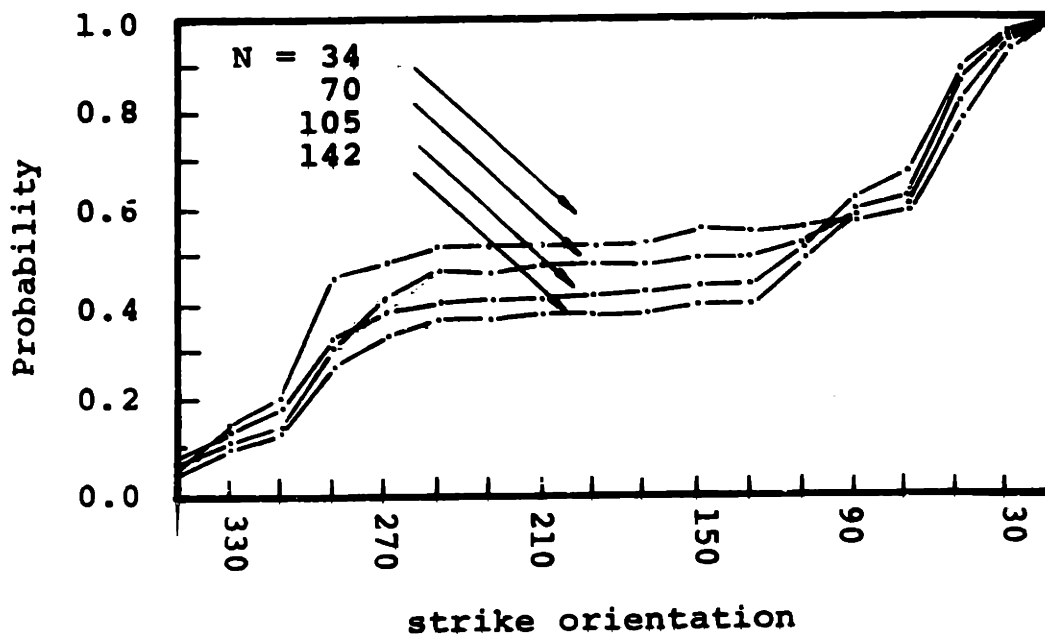
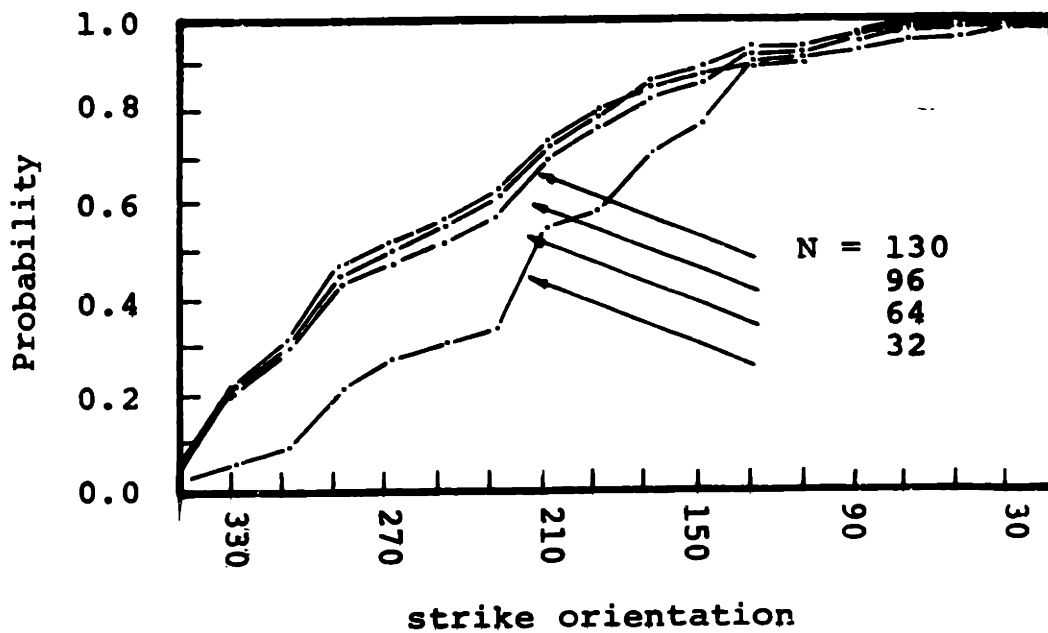


Figure 26

Changes in the cdf of orientation data as sample size increases (after Larsson, 1952)

problems associated with the use of analytical models for joint orientation distributions, however, and at present one finds work centered about the use of topologically planar distributions which are not actually appropriate to the problem. In this last section we will discuss available models, their properties, and some statistical tests associated with them. We will indicate some of the problems one should be wary of, and consider the use of Bayesian methods in inferences associated with mathematical models of joint orientation.

### Spherical Normal Distributions

The orientation distribution of planes in space can be represented by the distribution of their unit normal vectors; thus the problem of analytically characterizing a distribution can be approached by relating infinitesimal angular intervals to vector density. Since the poles of these vectors lie on a unit sphere about the origin, an analytical model may alternately be thought of as relating location on the unit sphere to pole concentration.

Location on the unit sphere is uniquely defined by two independent coordinates: in spherical notation,  $(\theta, \phi, \rho)$ , the radius  $\rho$  is constrained to equal 1.0.

In the simplest case we can assume that density is symmetrically distributed about the average orientation. Since we can assume with no loss of generality that the mean orientation is directed along the positive z axis, this reduces the unique description of location to one independent variable -- angular distance from the mean. A



symmetric spherical distribution, simple and moderately well known (Fisher, 1953; Watson, 1966), is the spherical normal (or "Fisher") distribution.\*

$$f(\theta|k) = \frac{k}{4\pi \sinh k} e^{k \cos \theta} \quad 37$$

where  $k$  is a measure of precision (akin to the reciprocal of variance). To clarify the meaning of the parameter  $k$ , Figure 27 relates  $k$  to the 1/2-angle of the cone enclosing  $\zeta\%$  of the vectors in a population distributed as equation 37.  $k=0$  for a uniform distribution, and increases without bound as the distribution becomes less disperse.

The spherical normal distribution is precisely analogous to the normal distribution, which decays as,

$$\exp(-1/2 (x-\bar{x}) / \sigma^2) \quad 38$$

as the discrepancy between the variate and its mean increases. Although several authors have used bivariate normal distributions to model strike and dip data (e.g., Robinson and Picteu, 1971; Pincus, 1961; McMahon, 1971), this distribution does not satisfy the constraints innate to orientation measurement. Orientation data are not defined on a plane, they are defined on a sphere. Now, often data which are defined on a sphere can be approximated by

---

\*This is analogous to the normal distribution of planar topology, and is a generalization of the frequently discussed vonMises circular normal distribution (e.g., see Gumbel, et al, 1953).

topologically planar distributions -- as a demographer might represent population density in a region or a surveyor might represent errors of angular measure -- but only if in the region of data the sphere can be approximated by a plane. Were the orientations of joints within a joint set highly localized (i.e., had very low variance), we could also make this approximation; but they almost always are not.

The primary consequence of spherical topology is that the distribution must be normalized over a unit sphere instead of an infinite plane (i.e.,  $\theta$  is defined over the interval  $-\pi < \theta < \pi$ ). This results in the constant  $k/4\pi \sinh k$  rather than the constant  $(1/2\pi)^{1/2} \sigma^{-1}$  characteristic of the normal distribution.

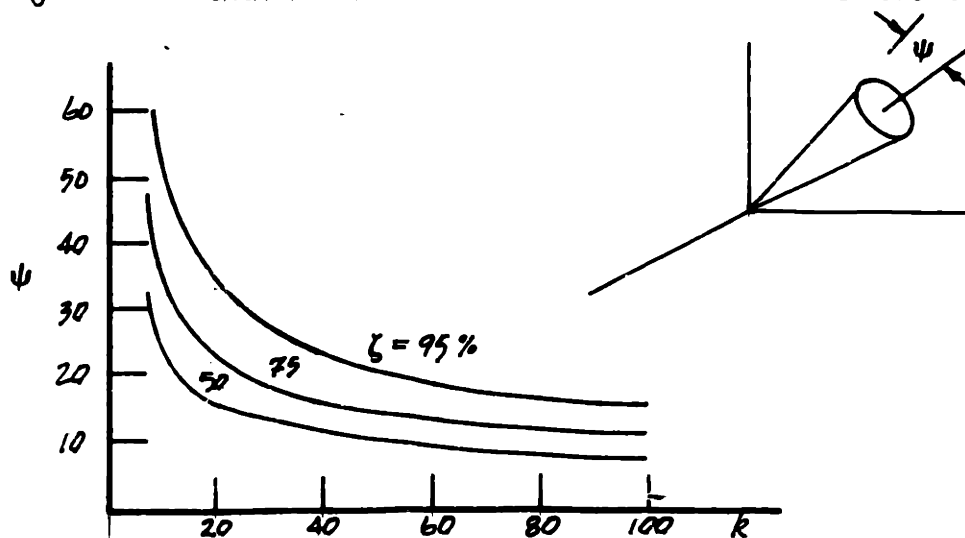


Fig. 27

The spherical normal model can be fit to a set of data by estimating the parameters, mean orientation and  $k$ . Given a set of data, the maximum likelihood estimate of the mean direction is the normalized vector sum of the sample elements (Fisher, 1922). If the

direction cosines of the  $i$  th observation are  $(z_{i1}, z_{i2}, z_{i3})$ , the direction cosines of the maximum likelihood estimate of the mean are,

$$\begin{aligned} R_1 &= \sum_i z_{i1} / R \\ R_2 &= \sum_i z_{i2} / R \\ R_3 &= \sum_i z_{i3} / R \end{aligned} \quad 39$$

where  $R$  is the magnitude of the sample vector sum

$$R = \sqrt{(\sum z_{i1}^2 + \sum z_{i2}^2 + \sum z_{i3}^2)} \quad 40$$

The maximum likelihood estimate of  $k$  is the solution to (Fisher, 1953)

$$\coth k - 1/k = R/N \quad 41$$

where  $N$  is the number of observations. For  $k > 3$  this is approximately (Watson, 1966)

$$k = (N-1)/(N-R) \quad 42$$

**Hemispherical Normal Distribution:**

The spherical normal distribution fits a density distribution function to the entire sphere. Joint orientation vectors, however, are defined in only one hemisphere. Assuming again that density decays as  $\exp(k \cos \theta)$ , but renormalizing the distribution such that the volume under the distribution over a hemisphere is 1.0

Arnold (1941) has shown

$$f(\theta|k) = \frac{k}{4\pi(e^k-1)} e^{k \cos \theta} \quad 43$$

The same maximum likelihood estimates apply for the mean orientation, but now the maximum likelihood estimate of  $k$  satisfies (Arnold, 1941)

$$R/N = \frac{1}{2} \left( \frac{ke^k}{e^k-1} - 1 \right) \quad 44$$

which for  $k > 5$  is approximately

$$k \cong N / (N - R) \quad 45$$

Values of the probability integral for spherical and hemispherical normal distributions are given in Tables 2 and 3. Unfortunately, almost no work on the hemispherical normal distribution exists in the literature.

#### Non-Axi-Symmetric Distributions:

Due to the greater complexity of non-axi-symmetric distribution functions, and the observation that many natural orientation populations (although not necessarily joints) are adequately modeled by the symmetric distribution function, exceedingly little work has been allocated to the study of the non-axi-symmetric case. Although many joint orientation distributions display symmetry about the mean, many do not. Before analytical methods can be used widely in modeling joint populations, an anisotropic model must be developed. Of

Table 2: Volume under the spherical normal distribution (adapted from Arnold, 1941)

Cos $\theta$ :	k = 0.0	k = 0.5	k = 1.0	k = 1.5	k = 2.0	k = 2.5	k = 3.0
1.0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.9	0.0500	0.0772	0.1101	0.1466	0.1847	0.2227	0.2598
0.8	0.1000	0.1505	0.2096	0.2728	0.3358	0.3961	0.4523
0.7	0.1500	0.2204	0.2997	0.3814	0.4596	0.5312	0.5949
0.6	0.2000	0.2868	0.3813	0.4748	0.5609	0.6364	0.7005
0.5	0.2500	0.3499	0.4551	0.5553	0.6439	0.7183	0.7788
0.4	0.3000	0.4100	0.5218	0.6245	0.7118	0.7821	0.8368
0.3	0.3500	0.4672	0.5822	0.6841	0.7675	0.8318	0.8797
0.2	0.4000	0.5215	0.6369	0.7354	0.8183	0.8705	0.9115
0.1	0.4500	0.5733	0.6863	0.7796	0.8503	0.9007	0.9351
0.0	0.5000	0.6225	0.7311	0.8176	0.8808	0.9241	0.9526
- 0.1	0.5500	0.6693	0.7715	0.8503	0.9058	0.9424	0.9655
- 0.2	0.6000	0.7138	0.8082	0.8784	0.9262	0.9567	0.9751
- 0.3	0.6500	0.7561	0.8413	0.9027	0.9430	0.9677	0.9822
- 0.4	0.7000	0.7964	0.8713	0.9235	0.9567	0.9674	0.9875
- 0.5	0.7500	0.8347	0.8985	0.9415	0.9679	0.9631	0.9913
- 0.6	0.8000	0.8711	0.9230	0.9569	0.9771	0.9883	0.9942
- 0.7	0.8500	0.9058	0.9452	0.9702	0.9847	0.9924	0.9964
- 0.8	0.9000	0.9388	0.9653	0.9817	0.9908	0.9956	0.9980
- 0.9	0.9500	0.9702	0.9835	0.9915	0.9959	0.9981	0.9991
- 1.0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table 3: Volume under the hemispherical  
normal distribution (adapted from  
Arnold, 1941)

Cos $\theta$ :	k = 0.0	k = 0.5	k = 1.0	k = 1.5	k = 2.0	k = 2.5	k = 3.0
1.0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.9	0.1000	0.1240	0.1505	0.1793	0.2096	0.2406	0.2408
0.8	0.2000	0.2419	0.2868	0.3336	0.3813	0.4286	0.4748
0.7	0.3000	0.3540	0.4100	0.4665	0.5218	0.5748	0.6245
0.6	0.4000	0.4607	0.5215	0.5808	0.6369	0.6886	0.7354
0.5	0.5000	0.5622	0.6225	0.6798	0.7311	0.7773	0.8176
0.4	0.6000	0.6387	0.7138	0.7639	0.8082	0.8463	0.8784
0.3	0.7000	0.7505	0.7964	0.8368	0.8713	0.9001	0.9235
0.2	0.8000	0.8379	0.8711	0.8995	0.9230	0.9420	0.9569
0.1	0.9000	0.9210	0.9388	0.9535	0.9653	0.9746	0.9817
0.0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

$$\text{Probability integral for hemisphere} = \frac{e^k - e^{k \cos \theta}}{e^k - 1}$$

course, the complexity of developing this model is greater than for the symmetric case because two parameters are involved.

If we define  $\theta_x$  as the angle between the z-axis and the projection of a unit vector on the x,z plane, and  $\theta_y$  as the angle between the z-axis and the projection of the y,z plane (Figure 28), density in a non-axi-symmetric spherical normal case might decay as

$$\exp \{ k_x \cos \theta_x + k_y \cos \theta_y \} \quad 46$$

where  $k_x$  and  $k_y$  are the precisions (Figure 29). A non-axi-symmetric spherical normal density function then would be

$$f(\theta_x, \theta_y | k_x, k_y) = (\text{Constant}) \exp \{ k_x \cos \theta_x + k_y \cos \theta_y \} \quad 47$$

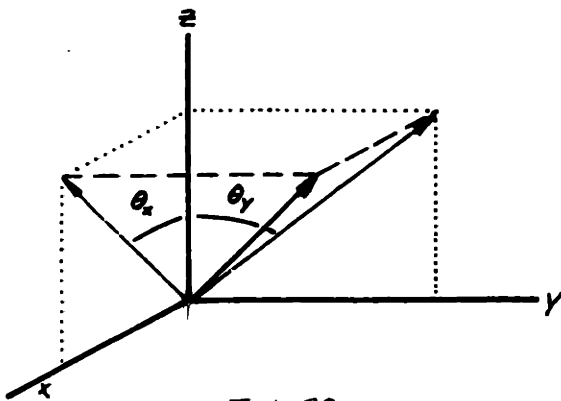


Fig. 28

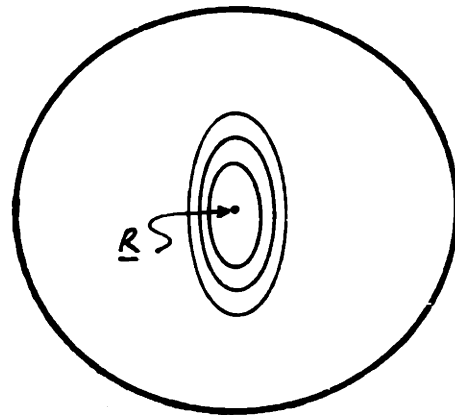


Fig. 29

where the constant normalizes the integral over the hemisphere (i.e., makes the volume under the function 1.0). This expression assumes independence between directions x and y, which is acceptable only if the engineer aligns these axes with the major axes of the observed ellipse. Otherwise, some additional term to account for dependence

must be introduced into the exponent. At this time (1972) little, if any, work has been done using non-axi-symmetric models, and their properties are not known.

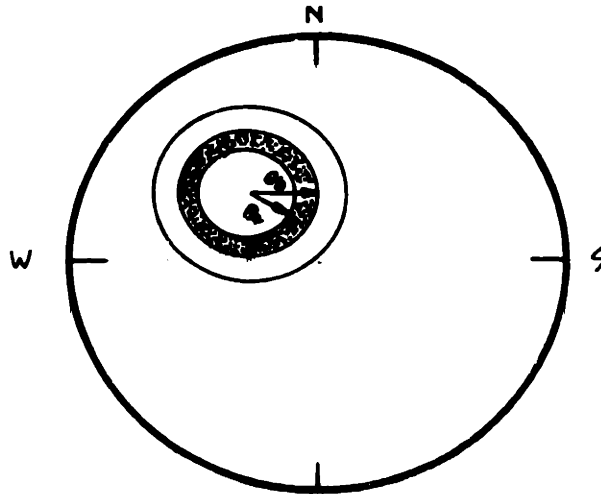


Fig. 30

### How Well Does a Model Fit?

Given a set of data and an analytical model whose parameters have been estimated from the data, we can evaluate the "goodness of fit" of the model using the  $\chi^2$  distribution. From this we can infer whether or not the model is a good representation of the sampled population.

Dividing the distribution into elements (e.g., as shown in Figure 30) we can count the number of poles falling in each, compute the predicted number,  $P_i$ , from the model, and evaluate Fisher's index of dispersion

$$\sum_i \frac{(O_i - P_i)^2}{P_i}$$



which is distributed as  $\chi^2$  with  $k-3$  d.o.f. ( $k-3 = k-1-2$  estimated parameters) when the data were sampled from a population described by the model. For values of this statistic which are significantly too large we conclude that the sampled population is not adequately described by the model. The level of improbability (or "significance") at which we draw this conclusion is solely convention (e.g., one "customarily" uses 5%). In Figure 30, the predicted number of poles between  $\theta_x$  and  $\theta_y$  (the ring-shaped area of the spherical surface shown) is

$$\left\{ \begin{array}{l} \text{predicted number of} \\ \text{poles} \end{array} \right\} = N \int_a^b f(\theta|k) d\theta \quad 49$$

which can be determined from Table 3, and the observed frequency is the number of poles actually observed within  $(\theta_a, \theta_b)$ .

### Comment

Although the statistical theory of the symmetric spherical normal distribution has been developed by such people as Fisher and Watson, the corresponding theory of hemispherical and non-hemispherical distributions is barely developed at all; most of their properties are unknown, and theoretical work on the mathematical basis of the latter distributions is required before spherical models can be widely applied to joint populations.

The spherical distributions discussed are appropriate for modeling joint orientation populations as long as the "goodness-of-fit" of the model is evaluated and normal precautions are taken.

Models fitted to empirical data are weighted most heavily by how well they model the central portion of the distribution where the majority of the observations are -- this is what the  $\chi^2$  test most heavily depends on. Therefore, the fit of a model at the extremes of an observation set may not be good at all, even if the  $\chi^2$  test did not show significant disparity over the whole distribution.

The purpose of modeling joint populations is to summarize the observations with a smooth, reasonable distribution which will make subsequent treatment easier. We have no random process model for the mechanism of formation of joint systems and therefore there is no theoretically correct distribution that orientation data should have. So, the extrapolation of distribution models outside of the observation set is a questionable undertaking.

#### Properties of the Hemispherical Normal Distribution:

From a sample of orientation measurements an estimate of the mean orientation can be made by equation 39. The precision of this estimate is increased by increasing the sample size. Watson (1966) shows, as we also do in the next section, that the precision of the estimate is  $qk$  where  $q$  is the sample size and  $k$  is the precision of the entire population. Watson assumes that the estimate is distributed in a hemispherical normal distribution and computes confidence limits from tables of the volume under the hemispherical normal curve (e.g., see Table 3, or Arnold, 1941 for more detailed tables). For example, for  $qk = 1000$ , 95% of the volume of the

hemispherical normal distribution is contained with  $2^\circ$  of the mean, so the 5% confidence limit on the true mean is  $2^\circ$ .\*

When we sample the same population at several locations we obtain samples, each of whose estimate of the mean is different. These estimates can be combined into a total estimate by summation of variances (Watson, 1966). If  $k$  is the precision of any sample and  $k_{bm}$  is the precision of the distribution of the sampled means, then the variance of the mean estimated from  $b$  samples is

$$1/k_R = 1/qk + 1/bk_{bm} \quad 50$$

The question of homogeneity in the joint distribution across the site is an important one in joint surveys and we would like to use the analytical model as an aid in deciding whether or not two samples could have come from the same population.

Let  $R_1$  and  $R_2$  be the vector sums of two samples and  $N_1$  and  $N_2$  the respective sample sizes. Given that  $k$  is large (say 5 to 8 or higher), as is almost always the case with joint populations, Watson (1956) has shown that the statistic,

$$\frac{R_1 + R_2 - R}{(N_1 - R_1) + (N_2 - R_2)} \quad \frac{2(LN-2)}{2} \quad 52$$

---

\*When  $qk$  is large ( $\geq \sim 7$ ) the  $\alpha$  % confidence limit can be found from

where  $N = N_1 + N_2$  and  $R$  is the total vector sum, is distributed as the  $F$  distribution with 2 and  $2(n-1)$  degrees of freedom. Values of 52 far from unity strongly suggest that the means are not homogeneous. To test for homogeneity at any significance level the value of 52 is computed and compared with tabulated values of  $F$  for the appropriate degrees-of-freedom and the desired significance level. For example, the tabulated value of  $F$  at the 5% level is the value that would be exceeded 5% of the time if the means really were homogeneous. This test may be extended to  $m$  populations by the statistic

$$\frac{R_1 + R_2 + \dots + R_m - R}{\sum_i (N_i - R_i)} \quad \frac{2(N-m)}{2(m-1)}$$

which is distributed as  $F$  with  $2(m-1)$  and  $2(N-m)$  degrees-of-freedom.

The value  $N$  equals the total number of observations  $\sum N_i$ ,

$R$  is the length of the resultant of all the vectors, and  $R_1, \dots, R_m$  are the resultants of the vectors in each population. (Watson, 1966).

The ordinary methods for comparison of variances may be used to test whether the variances of two or more samples differ from one another (Watson, 1956). The variance ratio

$$\max_i k / \min_j k$$

53

is distributed as  $F$  with  $N_i - 1$  and  $N_j - 1$  degrees-of-freedom.

Watson (1966) reports that the tests discussed here are robust against small deviations from the Fisher form, which means that

the distributions of the statistics tested are not sensitive to small departures from the assumed distribution model.

### Bayesian Inference

We have adopted in other parts of this thesis a Bayesian philosophy to probability because this facilitated decision making. In this chapter we have used methods of classical statistical inference exclusively.

Bayesian statistical inference differs from classical inference in that it defines the probability distribution of population parameters directly rather than assigning confidence limits to these parameters (see Chapter 1). In the next few pages we shall briefly explore a Bayesian approach to fitting the hemispherical normal model and drawing inferences from it, followed by a discussion of the usefulness of such an approach.

#### Distribution of Parameters $f(\underline{R},k)$ :

From the  $q$  unit vectors  $\underline{z}_i$  ( $i=1, \dots, q$ ) which constitute the sample data, we would like to determine the parameters  $\underline{R}$  and  $k$  of a corresponding analytical model. Since the data are only a sample of the population, true values of  $\underline{R}$  and  $k$  cannot be known exactly; instead the result is a joint pdf of  $\underline{R}$  and  $k$ ,  $f(\underline{R},k)$ .

If there is no prior information on  $\underline{R}$  and  $k$  (including intuitive judgement), as is usually the case, we will assume all values to be equally likely. Note that since  $k$  is proportional to the

reciprocal of the variance, this is not the same as assuming all values of the latter to be equally probable, and the two assumptions result in different posterior distributions (Chapter 1 contains a discussion of this point). From Bayes' theorem the posterior distribution of  $\underline{R}$  and  $k$  is,

$$f'(\underline{R}, k | \{\underline{z}_i\}) = \frac{f(\underline{R}, k) L(\{\underline{z}_i\} | \underline{R}, k)}{\int_{\underline{R}} \int_k f(\underline{R}, k) L(\{\underline{z}_i\} | \underline{R}, k) d\underline{R} dk} \quad 54$$

$$= \frac{L(\{\underline{z}_i\} | \underline{R}, k)}{\int_{\underline{R}} \int_k L(\{\underline{z}_i\} | \underline{R}, k) d\underline{R} dk} \quad 55$$

where  $\underline{z}_i$  is the set of observed orientations,  $i=1, \dots, q$  and,

$$\int d\underline{R} = \iiint dR_1 dR_2 dR_3 \quad 56$$

denotes integration over the vector  $\underline{R}$  (an underscore indicates a vector quantity).

The likelihood  $L(\{\underline{z}_i\} | \underline{R}, k)$  is the conditional probability of observing the sample data given the distribution parameters  $\underline{R}$  and  $k$ . For orientation data this is

$$\begin{aligned} L(\{\underline{z}_i\} | \underline{R}, k) &= \prod_{i=1}^q \frac{k}{4\pi(e^k - 1)} e^{k \cos \theta} \\ &= \left( \frac{k}{4\pi(e^k - 1)} \right)^q e^{k \sum_i \underline{R} \cdot \underline{z}_i} \end{aligned} \quad 57$$

where  $\underline{R} \cdot \underline{z}_i$  is the scalar product

$$\begin{aligned} \underline{R} \cdot \underline{z}_i &= |\underline{R}| |\underline{z}_i| \cos \theta_i \\ &= \cos \theta_i \end{aligned} \quad 58$$

Substituting the likelihood into equation 55 gives

$$f'(\underline{R}, k | \{z_i\}) = \frac{\left(\frac{k}{4\pi(e^k-1)}\right)^3 e^{k \sum \underline{R} \cdot \underline{z}_i}}{\int \int \left(\frac{k}{4\pi(e^k-1)}\right)^3 e^{k \sum \underline{R} \cdot \underline{z}_i} d\underline{R} dk} \quad 59$$

where the denominator is just a normalizing constant which insures that the integral of the posterior pdf over all parameters equals unity. This is a complicated equation to solve, but once the form of the solution is known its application becomes trivial.

Assume that by some process the precision  $k$  of the orientation distribution were known a priori. We shall show that with  $k$  given, the posterior distribution of  $\underline{R}$  is hemispherical normal with mean

$$\begin{aligned} \bar{R}_1 &= \sum z_{i1} / R \\ \bar{R}_2 &= \sum z_{i2} / R \\ \bar{R}_3 &= \sum z_{i3} / R \end{aligned} \quad 60$$

and precision  $kq$ .

From equation 59

$$f(\underline{R} | k, \{z_i\}) \propto \exp(k \sum \underline{R} \cdot \underline{z}_i) \quad 61$$

which Fisher (1953) has shown to be a minimum when 39 is satisfied.

Expanding the exponent in 61,

$$\begin{aligned}
 f'(\underline{R} | k, \{z_i\}) &\propto \exp (k (R_1 \sum z_{i1} + R_2 \sum z_{i2} + R_3 \sum z_{i3})) \\
 &\propto \exp (kq (R_1 \bar{R}_1 + R_2 \bar{R}_2 + R_3 \bar{R}_3)) \\
 &\propto \exp (kq (\underline{R} \cdot \bar{\underline{R}}))
 \end{aligned}
 \tag{62}$$

where  $\bar{\underline{R}} = \{ \bar{R}_1, \bar{R}_2, \bar{R}_3 \}$

and the proportionality factor is just that which normalizes the volume

$$f'(\underline{R} | k, \{z_i\}) = \frac{kq}{4\pi (e^{kq} - 1)} e^{kq \underline{R} \cdot \bar{\underline{R}}}
 \tag{63}$$

Since  $\underline{R} \cdot \bar{\underline{R}} = \cos \theta_r$ , where  $\theta_r$  is the angle between  $\underline{R}$  and the average  $\bar{\underline{R}}$ , equation 63 is a hemispherical normal distribution with precision  $kq$ .

The solution is much more difficult when  $k$  is also unknown, because  $k$  and  $\underline{R}$  are dependent variables. The conditional distribution of  $\underline{R}$ , as we have shown, is defined on the unit hemisphere, while  $k$  is defined on the line  $(0, \infty)$ . Until further theoretical work is done which results in a solution to the  $k$ -unknown problem, we must, as a first approximation, use the maximum likelihood estimate of  $k$  in evaluating the posterior distribution of  $\underline{R}$ .

### Homogeneity

The probability that two populations are the same (i.e., homogeneous) can be evaluated by Bayes' theorem. Define the hypotheses,



$H_0$ : the two populations are homogeneous

$H_1$ : the two populations are not homogeneous

If  $p(H_0)$  is the engineer's subjective prior probability of homogeneity, the posterior probability of homogeneity given for a sample from each population,  $\underline{z}_i$  and  $\underline{z}_j$  respectively, is

$$p'(H_0) = \frac{p(H_0) L(\{z_i\}, \{z_j\} | H_0)}{p(H_0) L(\{z_i\}, \{z_j\} | H_0) + (1-p(H_0)) L(\{z_i\}, \{z_j\} | H_1)} \quad 64$$

Let

$\xi(\underline{R}, k)$  = pdf of  $\underline{R}, k$  inferred from the sample of population 1

$\xi(\underline{R}, k)$  = df of  $\underline{R}, k$  inferred from the sample of population 2

$\xi(\underline{R}, k)$  = pdf of  $\underline{R}, k$  inferred from the combined sample

Then

$$L(\{z_i\}, \{z_j\} | H_0) = \prod_i \int \int f(z_i | R, k) \xi(R, k) dR dk \\ \times \prod_j \int \int f(z_j | R, k) \xi(R, k) dR dk$$

and

$$L(\{z_i\}, \{z_j\} | H_1) = \prod_i \int \int f(z_i | B, k) \xi_1(B, k) dB dk \\ \times \prod_j \int \int f(z_j | B, k) \xi_2(B, k) dB dk$$

66

which, when substituted into 64 require considerable computation to solve. Following the discussion of homogeneity between target and sampled populations we will present an example.

#### Target and Sampled Population Homogeneity:

As information becomes available from deep borings, construction openings, and similar sources, the new data should be compared with joint survey data to validate assumptions that the target and sampled populations are similar; the need for continuous evaluation of design assumptions need not be emphasized here. If the new data are consistent with the survey data, they can be incorporated into the total data set by modifying the parameter distribution accordingly.

The incoming data and the assumed joint population are compared for consistency (considered below). If the new data are consistent with the survey data, then advantage can be taken of the new information by incorporating it into the total data set, modifying the hypothesized joint population accordingly. Both the decision

on consistency and the subsequent modifications are facilitated by using Bayesian techniques.

If the number of joints in the new survey is large, the homogeneity of the target and sampled population can be evaluated by methods discussed already; however, if the new number is limited, as is often the case, Bayesian methods can be used which take account of the engineer's subjective probability that the two populations are the same.

On the basis of professional judgement the engineer has an idea of whether or not the target and sampled populations are similar. This is a subjective prior probability which can be quantified. Let the prior probability that the target and sampled populations are homogeneous be  $p$ , and hence the prior probability that the two populations are not homogeneous is  $(1-p)$ . From Bayes' Theorem the posterior probability of homogeneity is,

$$p' = \frac{p \prod_i L(\underline{z}_i | H_0)}{p \prod_i L(\underline{z}_i | H_0) + (1-p) \prod_i L(\underline{z}_i | H_1)} \quad 67$$

The likelihood of  $\underline{z}_i$ , given the two populations are the same, is,

$$L(\underline{z}_i | H_0) = \int \int f(\underline{z}_i | \underline{R}, k) \xi(\underline{R}, k) d\underline{R} dk \quad 68$$

where  $f(\underline{R}, k)$  is the density distribution of the parameters  $\underline{R}$  and  $k$  inferred from the original survey.

The likelihood of  $\underline{z}_i$  given the two populations are different is

more involved, for we require a prior probability distribution of  $\underline{R}$  and  $k$  in the new population, conditioned on them not equaling those of the original sampled population. That is, if we know that the parameters of the new population do not equal those of the original one, what might they equal? This distribution, which we will denote  $g(\underline{R},k)$  must be inferred subjectively. With little prior formation,  $g(\underline{R},k)$  may be assumed uniform except in the vicinity of  $\xi(\underline{R},k)$  (one might try to assess his subjective feelings more precisely, but he is usually plagued by almost complete lack of intuition about jointing at depth). In the vicinity of  $\xi(\underline{R},k)$ , the value of  $g(\underline{R},k)$  must diminish. For example, if the most likely values of the sampled population are  $(\underline{R}_0, k_0)$  and the target population is assumed to be different from the sampled population, the least likely values we would expect the target population to have are  $(\underline{R}_0, k_0)$ . Therefore, our assumed  $g(\underline{R},k)$  is a flat function with a "dimple" near  $\xi(\underline{R},k)$ . Figure 31 shows a one dimensional analogue.

Treating a function with the shape of  $g(\underline{R},k)$  will be difficult, so an assumption is made to simplify the derivation. We will neglect the distribution  $\xi(\underline{R},k)$  and assume that the sampled population is described only by the maximum likelihoods of the parameters. This is a reasonably good assumption when the number of observations in the original survey is large (e.g.,  $q=100$ ) since the precision in the posterior parameter distribution increases as  $qk$ . Under this assumption the distribution of  $g(\underline{R},k)$  is uniform

everywhere except for a valley of zero volume at  $(\underline{R}_0, k_0)$ . The function  $g(\underline{R}, k)$  may be integrated neglecting the valley at  $(\underline{R}_0, k_0)$  since the valley has zero volume and does not change the integral.

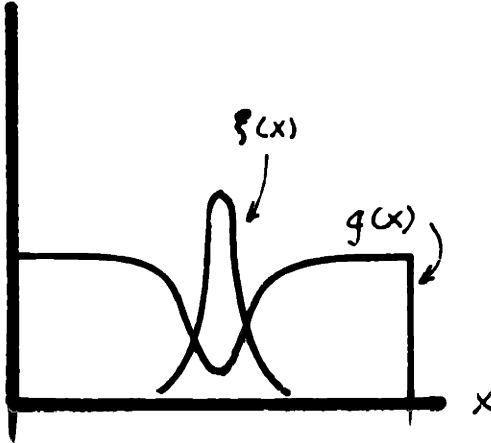


Fig. 31

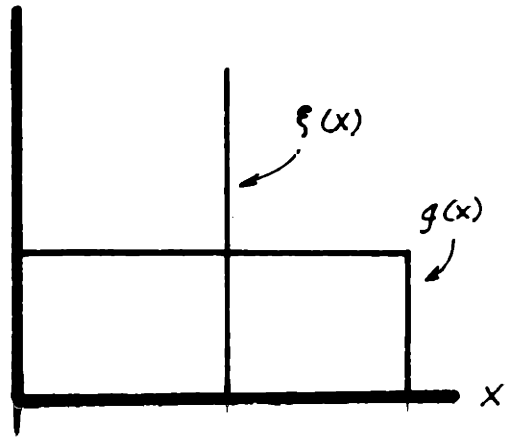


Fig. 32

The likelihood of  $\underline{z}_i$  given heterogeneity, then, is

$$L(\underline{z}_i | H_1) = \int \int f(\underline{z}_i | \underline{R}, k) g(\underline{R}, k) d\underline{R} dk \tag{69}$$

and the posterior probability of homogeneity from equation 67 becomes,

$$p' = \frac{p \prod_i f(\underline{z}_i | \underline{R}_0, k_0)}{p \prod_i f(\underline{z}_i | \underline{R}_0, k_0) + (1-p) \int \int f(\underline{z}_i | \underline{R}, k) g(\underline{R}, k) d\underline{R} dk}$$

Example:

From surveys prior to construction, parameters of the joint orientation distribution are estimated to be ( $\underline{R}_0 = N, 70^\circ$ ;  $k_0 = 10$ ). Since sample sizes were large, these estimates are precise, and are assumed to equal the true parameter values.

The first orientation data to be collected from construction openings consist of the 20 strike and dip measurements shown in Figure 33. Since design decisions must be made earlier than additional data are expected to arrive, the engineer would like to modify his original assessment of the probability of target-sampled population homogeneity in light of these 20 observations.

Before seeing the new data the probability of homogeneity was assessed as being 0.7. Since the engineer has no intuitive feeling about the most probable orientation at depth, he considers every orientation equally probable. In other words, his assessment of  $g(\underline{R}, k)$  is uniformly distributed over the unit hemisphere. He does have an intuitive feeling for the possible dispersion of orientation, however. He deems the probability that  $k > 100$  for the target population to be negligible, and so assigns a uniform value to  $g(\underline{R}, k)$  over the interval  $k = (0, 100)$ . So,

$$p(H_0) = 0.7$$

$$g(\underline{R}, k) = (1/2\pi)(1/100) = 1.59 \times 10^{-3}$$

The likelihood of the observed data, given  $H_0$ , is,

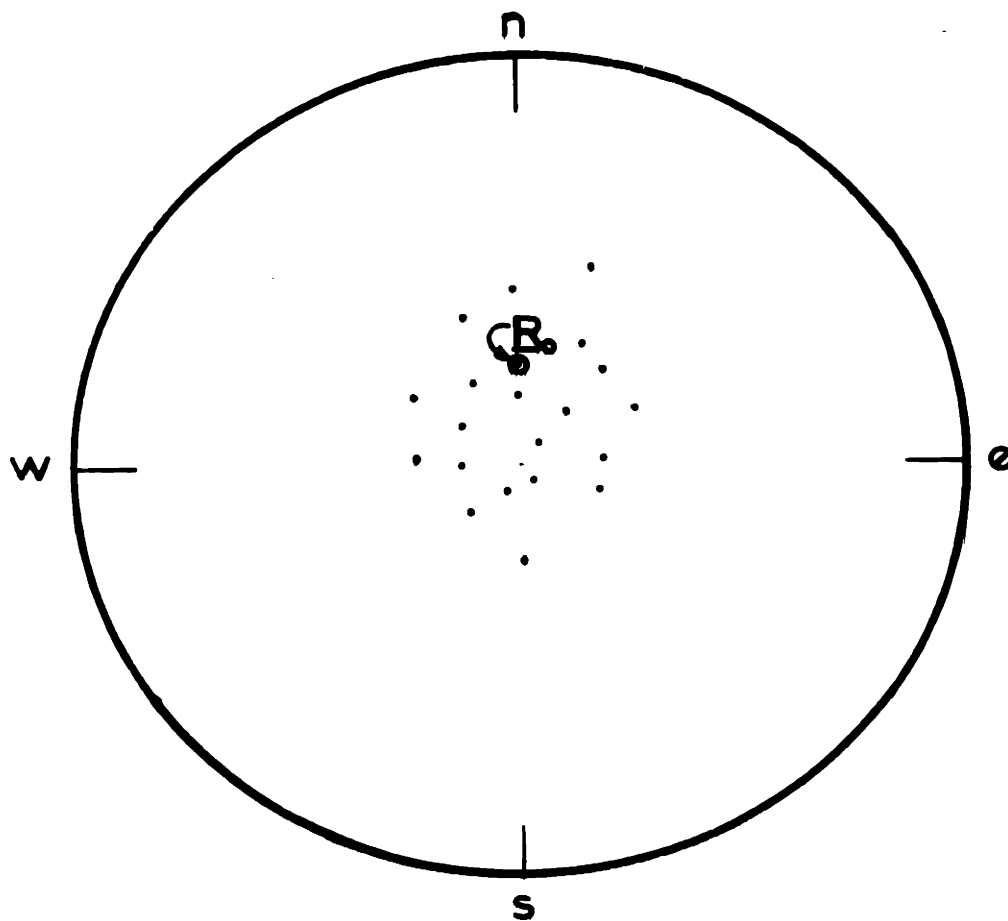


Figure 33

Initial joint orientation  
data from target population

$$L(\{\underline{z}_i\} | H_0) = \prod_i \frac{k_0}{4\pi(e^{k_0}-1)} \cdot e^{k_0 \cos \theta_i}$$

$$= 5.8 \times 10^{-9}$$

where  $\theta_i$  is the angle between  $\underline{z}_i$  and  $\underline{R}_0$ . The likelihood of the observed data, given  $H_1$ , is,

$$L(\{\underline{z}_i\} | H_1) = \int_k \int_{\mathbb{R}} g(\mathbb{R}, k) \prod_i \frac{k}{4\pi(e^k-1)} e^{k \cos \theta_i} d\mathbb{R} dk$$

This equation is difficult to solve analytically, but numerical techniques can be used. Here, the inner integral (i.e., over the unit hemisphere) is solved by spherical Gaussian quadrature which results in an expression having only  $k$  as a parameter. If we assume  $k > 3$  (i.e.,  $k=(3,100)$ , which seems a good assumption, since few joint populations are more dispersed than  $k=3$  -- a distribution with  $k=3$  has the property that more than 50% of the poles make angles of over  $40^\circ$  with the mean), the integration over  $k$  can be evaluated analytically. Based on the data of Figure 33,

$$L(\{\underline{z}_i\} | H_1) = 7.9 \times 10^{-9}$$

So



$$\begin{aligned}
 L \quad p'(H_0) &= \frac{p(H_0) L(\{z_i\} | H_0)}{p(H_0) L(\{z_i\} | H_0) + (1 - p(H_0)) L(\{z_i\} | H_1)} \\
 &= \frac{0.7(5.8) \times 10^{-9}}{0.7(5.8) \times 10^{-9} + 0.3(7.9) \times 10^{-9}} \\
 &= 0.63
 \end{aligned}$$

Therefore, the new data lessened the engineer's degree-of-belief in the homogeneity of target and sampled populations from 0.7 to 0.63.

#### Modification of $f(\underline{R}, k)$ for New Information

With a decision that the new information is from the same parent population as the joint material, modification of the initial pdf of  $(\underline{R}, k)$  to account for the new data can be made using Bayes' Theorem.

Let the distribution of  $(\underline{R}, k)$  from the joint survey be  $\xi(\underline{R}, k)$  -- the prior pdf. The posterior pdf, which is the distribution of  $(\underline{R}, k)$  modified by the new information, from Bayes' Theorem is

$$g^i(\underline{R}, k) = \frac{g(\underline{R}, k) L(\{z_i\} | \underline{R}, k)}{\int \int g(\underline{R}, k) L(\{z_i\} | \underline{R}, k) d\underline{R} dk} \quad 71$$

### Comments on Bayesian Inference

Using Bayesian methods to fit distribution models and test homogeneity requires considerable computation; particularly compared to the ease with which classical methods can be used. Since one usually has little intuitive feel for the parameters describing joint distributions prior to observing sample data, the main advantage of Bayesian methods is that they result in density functions of parameters and hypotheses rather than confidence limits. For two reasons the additional effort of using Bayesian methods is not justified. First, the precision of the maximum likelihood estimators of population parameters is very high, since sample sizes are commonly in the range of 100 to 200; therefore, assuming the estimates to be the true population parameters results in little error. Second, the design methods within which the joint information is used are not refined, and introduce error which overshadows errors of inference.

Were the model we fit to orientation data based on some random process model for the origin of jointing, Bayesian inference would perhaps be more justified, since the fitted model would be theoretically correct. The models we actually do fit, however, are chosen only

for convenience and so are themselves approximations.

Only in the case of testing homogeneity between the population originally surveyed and that found in construction openings do Bayesian techniques seem justified. They are of use here because inferring the homogeneity (or lack thereof) between target and sampled population is very much a subjective undertaking. Bayesian techniques allow the inclusion of subjective prior probabilities which, when made by an experienced geologist or engineer, may result in a more realistic inference than that based on classical methods.

### Section 7.3 Summary

Joint surveys are sampling programs for collecting data from which the distributions of orientation and intensity of joint sets may be inferred; they are one problem class within exploration which may justifiably be treated by relative-frequency techniques. To obtain estimates of population parameters for which quantitative precisions can be evaluated, sampling plans must possess four characteristics. One of these is that each element in the sample have the same influence on weighted sample statistics. The population of interest in a joint survey is called the target population and differs from the population actually sampled since not all joints are "available" for measurement. Correspondence between target and sampled population is a matter of professional judgement, and outside statistical analysis.

A sampling plan for joint surveys is suggested, which is based on cluster sampling. This plan results in data which can be tested for preferred orientation and homogeneity by statistical methods. A mesh of equal sized quadrates is suggested for estimating pole densities (instead of the classical "unit counting circle"), since it is easy to use and interfaces easily with statistical tests.

Mathematical functions may be used to model orientation distributions, but since there is no theoretically correct model for joint population, extrapolations outside the data set are questionable. Models must be thought of merely as smooth, reasonable summaries of sample observations. Distribution models should be chosen which satisfy the constraints inherent to orientation data, and one such model is the hemispherical normal distribution. Bivariate normal distributions, defined on a plane, do not satisfy these constraints.

Bayesian techniques for fitting and treating analytical models of joint orientation require considerable computational effort relative to classical methods, and seem justified only for inferring the homogeneity of target and sampled population (on the basis of data from construction openings) in which judgement is highly important.

## APPENDIX TO CHAPTER 7

This appendix contains a brief introduction to common classical sampling plans. It is not intended to be an exhaustive treatment, but rather a summary. Many books have been written on sampling theory and for more complete discussion one should see Cochran (1963), Hansen, et al., (1966), or any of the texts listed in the bibliography.

The purpose of sampling is to obtain estimates of population parameters (e.g., mean, variance) without observing and measuring every element in the sampled population. An estimator is a sample statistic which can be used to estimate true population parameters; it can be chosen on several bases: an unbiased estimator is one whose expected value over many samples equals the parameter it estimates; a minimum variance estimator is one whose distribution about the parameter it estimates (over many samples) has the least variance; and a maximum likelihood estimator is one which maximizes the conditional probability of the observed sample given that the true parameter value equals the estimate. An individual estimator does not necessarily satisfy all three criteria.

Estimates are not exact, and their uncertainty is reflected in the variance of their distribution about the true parameter value they estimate. This variance is, in turn, a function of both the sampling plan and the sampled population. By knowing this variance and making assumptions about the distribution shape, confidence limits on true population parameters can be set. The usual assumption is that estimates are normally distributed. One can show (Kendall

and Stuart, 1966) that if the sampled population is normally distributed, the estimate of the mean must also be normally distributed; and that if the sampled population is not normally distributed, the estimate of the mean approaches normality as sample size increases (Feller, 1957). Empirical evidence suggests (Cochran, 1963; West, 1951) that the normal approximation is good even for small sample sizes. Wilson (1952) discusses tests which can be used when the validity of a normal approximation is suspect.

A sampling plan is a program of action for collecting data from a sampled population. Common plans may be grouped into four types: simple random, systematic, stratified random, and cluster. In deciding between plans, or in designing a specific program once the plan type has been chosen, one attempts either to obtain the highest precision for a fixed sampling cost or the lowest sampling cost for a fixed precision.\* Such maximizing or minimizing plans will be called optimal.

No matter what sampling plan is chosen, the selection of sample elements must be accomplished using some "randomized" procedure. Since arbitrary selection is not random -- we all display bias, consciously or unconsciously -- this requires a mechanism for generating random numbers such as dice, spinners, or (most commonly) random number tables. This requirement is no hardship, for a table of

---

\* In contrast to Bayesian methods which attempt to maximize the overall expected utility from both sampling and design.

random numbers can be easily taped to the underside of a clipboard.

Since the rudiments of classical sampling plans are easily seen in the analysis of simple random sampling, we will begin there, then broaden our discussion to include the three other plans.

### Section 7A.1 Simple Random Sampling

The characteristic property of simple random sampling is that sample elements are chosen individually, at random, from the sampled population, and each element in the sampled population has an equal probability of being sampled. Thus, a simple random sample is self-weighting.

#### Estimators

An unbiased estimator of the true population mean for simple random samples is the sample mean (Table A.1),

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{A1}$$

( $x_i$  are the sample elements) which has variance,

$$\begin{aligned} V(\bar{x}) = \sigma_{\bar{x}}^2 &= \left(\frac{s^2}{n}\right) \left(\frac{N-n}{N}\right) \\ &= \left(\frac{s^2}{n}\right) (1-f) \end{aligned}$$

A2

where  $s^2$  is the true population variance,  $N$  is the size of the sampled population, and  $n$  is the sample size. The ratio  $n/N$  is

TABLE A.1: Formulae for Estimating Population Characteristics

plan	$\bar{x}$	$V(\bar{x})$	$s^2$
random	$\frac{\sum x_i}{n}$	$\frac{N-n}{N} \frac{s^2}{n}$	$\frac{\sum (x_i - \bar{x})^2}{n-1}$
systematic	$\frac{\sum x_i}{n}$	$(\frac{N-1}{N}) s^2 + \frac{k(N-1)}{N} s_w^2$ — or — $\frac{s^2}{n} (\frac{N-1}{N}) (1 + (n-1) f_w)$	depends on population model
stratified random	$\frac{\sum N_h \bar{x}_h}{N}$	$\sum (\frac{N_h}{N})^2 (\frac{s_h^2}{n_h}) (1 - \frac{n_h}{N})$	$\frac{1}{n} \sum s_h^2 + s_{bm}^2$
cluster	$\frac{\sum_{i,j} x_{ij}}{mn}$	$(1-f_1) \frac{s_1^2}{n} + \frac{f_1(1-f_2) s_2^2}{nm}$ assumes equal cluster sizes	$\sum_{i,j}^m \frac{(x_{ij} - \bar{x}_i)^2}{m(n-1)}$ + $\frac{\sum (\bar{x}_i - \bar{\bar{x}})^2}{n-1}$



## Notation for Table A.1:

$x_i$	element of sample
$\bar{x}_i$	sample mean
$\bar{x}_h$	h th stratum sample mean
$x_{ij}$	i th sample element of j th cluster
$\bar{\bar{x}}$	total sample mean for cluster sampling
f	sampling fraction
$f_1$	sampling fraction for clusters in cluster sampling
$f_2$	sampling fraction for elements within clusters
h	stratum number
k	systematic sample spacing
m	number of clusters sampled
N	number of elements in total sampled population
$N_h$	number of elements in h th stratum of sampled population
n	number of elements sampled
$s_{bm}$	standard deviation of distribution of strata means
s	standard deviation of sampled population estimated
$s_h$	standard deviation of stratum h
$s_w$	standard deviation within systematic sample
$s_1$	standard deviation of distribution of cluster means
$s_2$	standard deviations of elements within one cluster
$\rho_w$	correlation between elements of systematic sample

called the "sampling fraction," and  $(1-f)$  is called the "finite population correction" of fpc. For joint populations the fpc is essentially 1.0.

An unbiased estimator of the true population variance is

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{(n-1)}$$

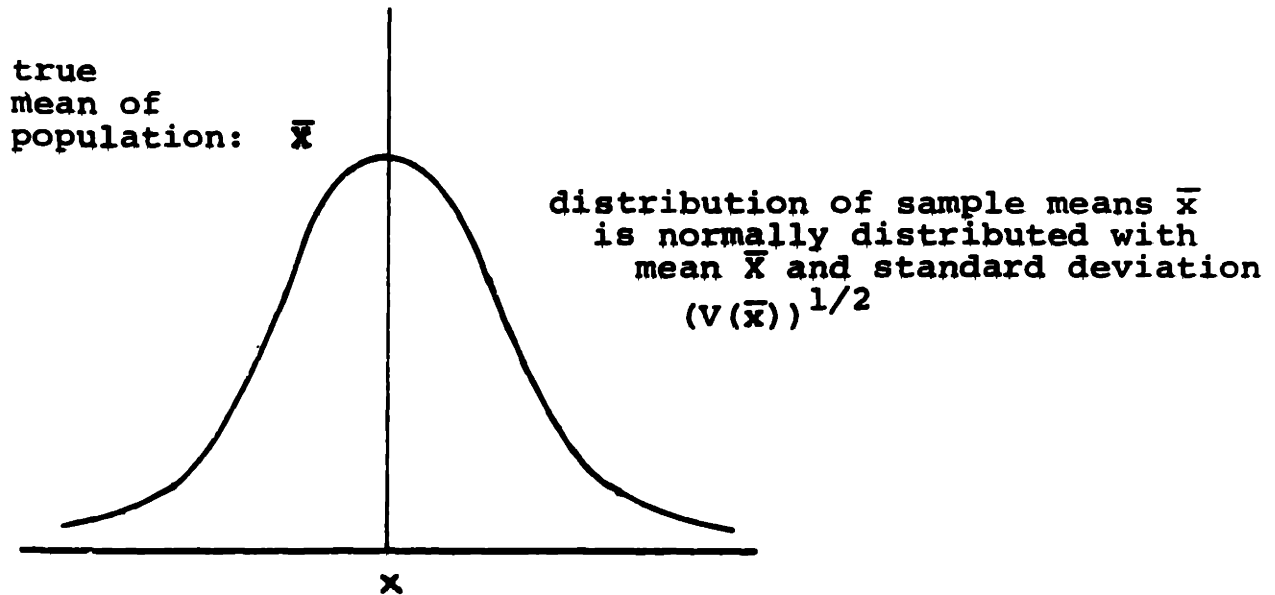
A3

### Example

Consider the problem of sampling a sandstone and shale series for intact strength, from which to estimate the life expectancy of tunneling machine cutter heads. If the borehole locations and the depths of the specimens are taken at random, the probability of any one element within the rock mass being sampled is the same. The probability of a shale specimen being sampled is proportional to the fraction of shale in the formation, but the probability of any individual specimen being sampled, whether it be shale or sandstone, is the same.

If we wish our estimate of the mean to be within  $5 \text{ kg/cm}^2$  of the actual value 95% of the time, how many samples should we take? To answer this question we must have at least an estimate of the actual population variance. Say, that from tests on similar formations we estimate the variance to be 900 (i.e., a standard deviation of  $30 \text{ kg/cm}^2$ ). Since from Figure A.1a we see that 95% of the time the estimate is within  $\pm 1.96 V(\bar{x})$  of the true value,

Figure A.1

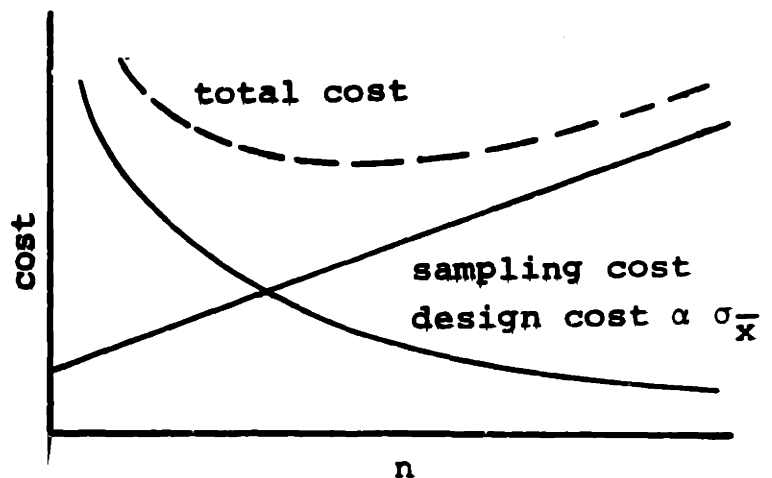


The probability that the true mean is further than  $k$  standard deviations away from the sample mean, from tables of the normal distribution, is

$k$	probability
1.0	1-0.68
1.28	1-0.80
1.64	1-0.90
1.96	1-0.95
2.58	1-0.99

Therefore, the true mean is further than 1.96 standard deviations from the sample mean only 5% of the time, and  $\bar{x} \pm 1.96\sigma_{\bar{x}}$  are said to be the 5% confidence limits of the true  $\bar{x}$  mean.

Figure A.2: Relationship of cost to sample size



$$1.96 V(\bar{x}) = 5 \text{ kg/cm}^2$$

$$n = 23$$

So, 23 specimens should be tested.

### Optimization

The only allocation to be made in simple random sampling is total sample size  $n$ . From equation A2 we know that the variance of the mean is related to sample size by

$$V(\bar{x}) = 1/n$$

Therefore, if a certain precision  $1/\sqrt{V(\bar{x})}$  is required, the appropriate sample size can be determined. Figure A.1b shows the relationships of sampling cost and variance to  $n$ . We could optimize the sampling plan for total cost by assuming some relationship between  $V(\bar{x})$  and cost in construction or design. A commonly made assumption which, for simplicity, has been made here is that this cost is proportional to  $\sqrt{V(\bar{x})}$ , the standard error of the mean. The optimal sample size is found by the minimum of the total cost curve.\*

---

\*To illustrate how a Bayesian might approach this same problem, consider the following. As discussed in Chapter 1, a Bayesian places probabilities on the truth of hypotheses themselves rather than determining confidence limits. From  $V(\bar{x})$  the pdf of an error of size  $e$  in the estimate of the true mean can be determined, as can the joint pdf of  $e$  and  $n$ ,  $f_{e,n}(e_0, n_0)$ , from sampling theory.

The "value" of an error can be expressed in terms of the loss,

Comments

1. Care must be taken to insure that the probability of each element being sampled is actually the same. If they are not, weighting factors must be applied.
2. The standard error of the mean,  $\sqrt{V(\bar{x})}$ , decreases as  $1/(n)^{1/2}$  with increasing sample size. Since standard error of the mean is inversely proportional to precision, quadrupling the sample size only doubles the precision of the estimate of the mean.
3. As long as the sampling fraction  $f=n/N$  is small, the fpc may be neglected. As  $n/N$  becomes larger,  $V(\bar{x}_i)$  decreases. Hence, neglecting the fpc is conservative since it overestimates the standard error of the mean. Equation A2 is interesting: it says that a sample of 100 from a population of 5000 gives essentially as precise an estimate of the mean as does the same sample from a population of 50,000 -- ten times the size of the first population -- a consequence which is not intuitive.

$l(e_o)$ , incurred by an error of magnitude  $e_o$ , and the expected loss from a sampling plan becomes

$$L(n_o) = E(l(e_o) | n_o) = \int l(e_o) f_{e,n}(e_o, n_o) de_o$$

if the cost of sampling  $n$  elements is  $C(n)$ , the expected "value" of a sample of size  $n_o$  is

$$\text{Expected value}(n_o) = C(n_o) + L(n_o)$$

Allocation is made on the basis of maximizing (algebraically) the expected value over  $n_o$ .

## Section 7A.2 Systematic Sampling

In systematic sampling the first sample element is chosen at random, and subsequent elements are chosen periodically throughout the population. Thus, grid sampling is a systematic plan.

### Estimators

An unbiased estimate of the mean from a systematic sample is,

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{A4}$$

The variance in this estimate is,

$$V(\bar{x}) = \sigma_{\bar{x}}^2 = \left(\frac{N-1}{N}\right) s^2 + \left(\frac{k(N-1)}{N}\right) \quad \text{A5}$$

where  $k$  is the interval between samples ( $k = N/n$ ), and  $s_w^2$  is the variance of elements within the same systematic sample

$$s_w^2 = \frac{1}{k} \sum_i^k \frac{\sum_j^n (x_{ij} - \bar{x})^2}{(n-1)} \quad \text{A6}$$

$x_{ij}$  is the  $j$  th member of the  $i$  th interval (Figure A.2).

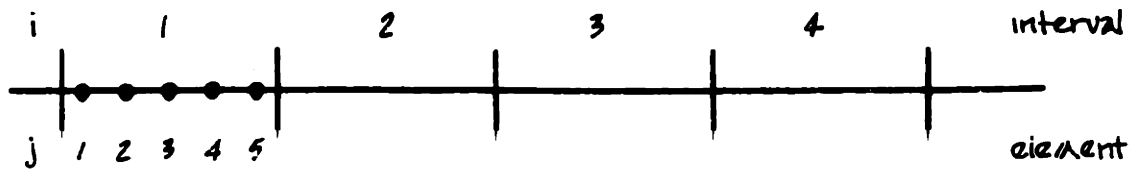


Fig.2

When only one systematic sample has been taken (i.e., one set of  $n$  observations at spacing  $k$ ) the variance of the mean cannot be evaluated unless assumptions are made about the nature of the sampled population. The conventional assumption is that the population can be modeled by a linear expression

$$x_i = \mu + e_i \quad A7$$

where each member of the population,  $x_i$ , can be thought of as the sum of the mean of the population and a zero-mean random component. This leads to the following results (Cochran, 1963):

Random Populations:  $\mu = \text{constant}$

$$V(\bar{x}) = \left(\frac{N-n}{N}\right) \frac{1}{n} \frac{\sum (x_i - \bar{x})^2}{(n-1)} \quad A8$$

Random within strata of size  $k$ :  $\mu = \text{constant for each stratum of size } k$

$$V(\bar{x}) = \left(\frac{N-n}{N}\right) \frac{\sum (x_i - \bar{x})^2}{2(n-1)} \quad A9$$

\*This estimate is not unbiased, but it overestimates the magnitude of  $V(\bar{x})$  and, hence, is conservative.

Linear Trending:  $\mu = \mu_0 + b_i$

$$V(\bar{x}) = \left(\frac{N-n}{N}\right) \frac{1}{n} \frac{\sum (x_i - 2x_{i+k} + x_{i+2k})}{6(n-2)} \quad \text{A10}$$

(This estimate is unbiased only if the variance of the population within each stratum is the same.)

Autocorrelated:

Autocorrelated populations are important in geology, and will be discussed in Section 7.A.3.

These estimates depend on the population being adequately described by the linear model, and therefore, contain inherent weakness. A general way of estimating  $V(\bar{x})$  from one systematic sample simply does not exist (Cochran, 1946; DeLury, 1950; Matern, 1947; Osborne, 1942; and Yates, 1949, have all dealt with this problem), since the use of other models produces somewhat different estimates. Standard convention in sampling practice is to adopt the linear model, although from professional considerations (i.e., experience and knowledge of geology) different choices might be made when the nature of the population is thought to deserve it.

### Optimization

The parameter over which allocation is optimized for systematic sampling is  $k$ , the spacing between observations.\* From equation A5,

$$V(\bar{x}) = \left(\frac{N-1}{N}\right) S^2 - \left(1 - \frac{k}{N}\right) S_w^2 \quad \text{A11}$$

---

\*In more than one dimension the pattern of systematic sampling may also be optimized (e.g. triangular vs, rectangular, etc.).



and if  $S^2$  and  $S_w^2$  are approximately the same,

$$V(\bar{x}) \cong \frac{k}{n} S_w^2$$

A12

so precision can be controlled only by changing the ratio of sample spacing to population size (e.g., bore hole spacing to formation or site dimension). The plot of  $V(\bar{x})$  vs.  $k$  is linear (Figure A.3).

Using the cost model,

$$C(n) = c_0 + c_1 n$$

where  $c_0$  is a set-up cost and  $c_1$  is the additional unit cost, a curve of sampling cost vs.  $k$  can be constructed (Figure A.3). The optimal  $k$  is determined, as before, by assuming a cost associated with  $\sqrt{V(\bar{x})}$ .

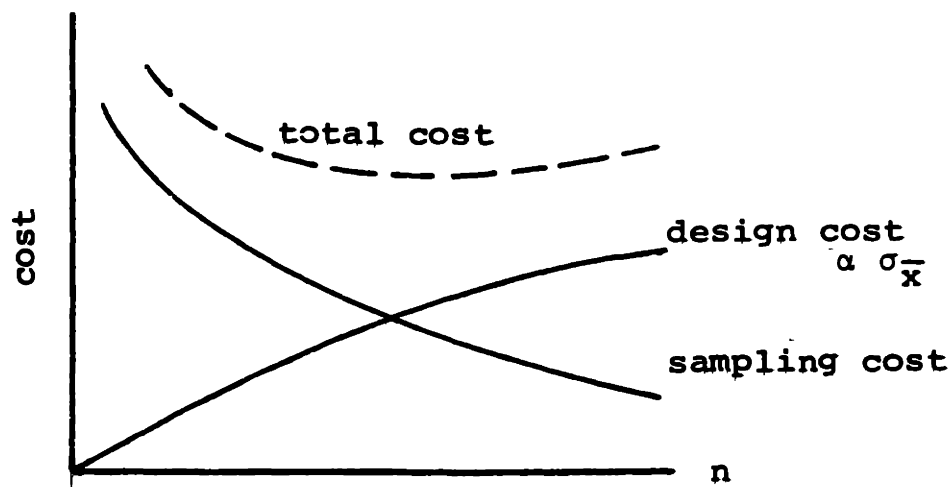


Figure A.3

Relationship of cost to sample size: systematic

Comments

1. Systematic sampling is the predominant sampling plan used in site investigation because it is easy to specify and administrate, coverage of the site is ensured, and trends are easily analyzed in the resulting data.
2. If the first sample element is randomly located, each element in the sampled population has an equal chance of being sampled, so systematic plans are self-weighting.
3. Remembering that the variance of the mean for random samples is,

$$V(\bar{x}) = \left(\frac{N-n}{N}\right) \frac{s^2}{n}$$

systematic sampling is more precise (for estimating the population mean) if and only if,

$$s_w^2 > s^2$$

In words, the elements within a systematic sample must be more heterogeneous than the population as a whole for systematic sampling to be more precise than simple random sampling. If,

$$s_w^2 = s^2$$

systematic and simple random sampling are equally precise.

Obviously, if positive correlation exists between elements within a systematic sample, as when periodicity exists in the population, the variance of the mean expands rapidly. Expressed in terms of this correlation (Cochran, 1963),

$$V(\bar{x}) = \frac{s^2}{n} \left(\frac{N-1}{N}\right) (1 + (n-1) \rho_w)$$

where  $\rho_w$  is the correlation between elements within the same sample

$$\rho_w = \frac{E[(x_{ij} - \bar{x})(x_{iu} - \bar{x})]}{E[(x_{ij} - \bar{x})^2]}$$

A14

(j and u denote elements).

### Section 7A.3 Stratified Random Sampling

A heterogeneous population can sometimes be naturally divided into subpopulations which are internally homogeneous. For each homogeneous subpopulation, called a stratum, precise estimates of the stratum characteristics can be obtained by random sampling. Estimates of the total population characteristics can then be made by combining the individual stratum estimates. For certain populations, stratifying before sampling is more efficient than taking samples directly from the total population. Sampling plans that specify a simple random sample in each stratum are called stratified random sampling plans.

#### Estimators

An unbiased estimator of the mean of the (total) sampled population is

$$\bar{\bar{x}} = \frac{\sum^m N_h \bar{x}_h}{N}$$

A15

where  $m$  is the number of strata, and the subscript  $h$  denotes the stratum (i.e.,  $N_h$  is the size of the  $h$ th stratum). The variance of this estimate is

$$V(\bar{x}) = \sum_h W_h^2 \frac{s_h^2}{n_h} (1 - f_h) \quad \text{A16}$$

where

$$W_h = N_h / N \quad \text{A17}$$

$$f_h = n_h / N_h \quad \text{A18}$$

Since the sample from each stratum is simple random, the estimate of the variance within each stratum is

$$s_h^2 = \frac{\sum (x_{hi} - \bar{x}_h)^2}{n_h - 1} \quad \text{A19}$$

where  $x_{hi}$  is the  $i$ th element of the sample from the  $h$ th stratum.

An estimate of the variance of the total population is

$$s^2 = \frac{1}{r} \sum_h s_h^2 + s_{\text{BETWEEN MEANS}}^2 \quad \text{A20}$$

### Example

Returning to the example of the shale-sandstone profile, each lithology can be considered as a subpopulation or stratum of the sampled population (i.e., the rock mass). Independent, simple random samples can be taken from both the shale and the sandstone and their sample statistics estimated separately. The estimate of the average strength of the total population is determined by combining the

average strata values and the size of the strata.

$$\left\{ \begin{array}{l} \text{average population} \\ \text{strength} \end{array} \right\} = (\% \text{ shale}) (\text{ave. shale str.}) \\ + (\% \text{ sandstone}) (\text{ave. ss. str.})$$

The advantage of stratified sampling over adjustment is that the allocation of effort to each stratum is controlled. This allocation may be optimized for precision or cost.

### Optimization

For the moment we will assume that the strata are obviously separable, and that the relative sizes of the sampled strata are known. How many elements do we take from each stratum? If the cost function is of the form

$$\text{cost} = C_0 + \sum_h C_h n_h \quad \text{A21}$$

the variance of  $\bar{x}$  is minimized when (Cochran, 1963)

$$n_h \propto N_h S_h / \sqrt{C_h} \quad \text{A22}$$

In words, the sample size within a stratum should be larger when:

- 1) the stratum is larger
- 2) the variance of the stratum is larger
- 3) the cost of sampling in the stratum is lower.

For the common special case when the cost per unit sampled is the same in each stratum,  $V(\bar{x})$  is minimized when (Neyman, 1934)

$$n_h = n \left( \frac{N_h S_h}{\sum N_h S_h} \right)$$

This is referred to as the Neyman allocation.\*

### Example

Consider the shale-sandstone profile once again, but assume that it also contains some limestone. The assumed proportions, variance, and sample allocations for a total of 100 samples are shown in Table A.2 (the cost of sampling each lithology is assumed to be the same).

Table A.2

	$W_h$	$S_h^2$	Neyman Allocation	Proportional Allocation
Sandstone	0.6	1.0	59	60
Shale	0.3	0.9	26	30
Limestone	0.1	1.5	<u>15</u>	<u>10</u>
			100	100

---

\*Another allocation frequently used is proportional allocation,

$$n_h = nN_h/N$$

which is sometimes more convenient than Neyman allocation, since prior estimates of the within-stratum variances are not required. However, Neyman allocation offers greater precision for the same sampling effort.

Table A.3

<u>Allocation Type</u>	<u><math>V(\bar{x})</math></u>
Simple Random	$0.01 + S_{bm}^2$
Neyman	0.0104
Proportional	0.0107

The variances in the estimates of the mean are shown in Table A.3. For very low variance between the means of the different lithologies simple random sampling is the most precise. Otherwise, Neyman-allocated stratified sampling is the most precise.

In real cases the values of  $N_h$  and  $S_h$  are seldom confidently known prior to sampling. Error, therefore, cannot be avoided in the evaluation of the  $n_h$ . Fortunately, the increase in  $V(\bar{x})$  caused by departures from the optimal  $n_h$  is small for substantial deviations (Cochran, 1963).

If  $g$  is defined as

$$g = \max_h (|n_h - n_h^*| / n_h)$$

A24

where  $n_h^*$  is the optimum value of  $n_h$ , then,

$$\frac{V(\bar{x}) - V(\bar{x})_{\min}}{V(\bar{x})_{\min}} \leq \frac{1}{n} \sum_h \frac{n_h^2 g^2}{n_h} = g^2$$

For example, if the proportional error in  $n_h$  is  $\leq 0.2$  then

$$\frac{V(\bar{x}) - V(\bar{x})_{\min}}{V(\bar{x})_{\min}} \leq 0.4$$

So we have seen that uncertainty in the estimations of the relative strata sizes  $N_h/N$  does not cause much decrease from the optimal precision. However, inaccuracy in the strata sizes does cause a bias error in the estimator of the mean. Therefore, the uncertainty in the estimate of the mean,  $\bar{x}$ , cannot be determined solely from the precision of the estimator (i.e., the spread of the  $\bar{x}$  from different samples about its average value).  $V(\bar{x})$  underestimates the real error in the estimate of  $\bar{x}$  since it does not include the effect of the bias. The mean square error (MSE) of the estimate  $\bar{x}$  about the true population mean is (Stephan, 1941)

$$MSE(\bar{x}) = \frac{\sum w_h s_h^2}{n_h} (1 - f_h) + \left[ \sum (w_h - W_h) \bar{x}_h \right]^2 \quad A25$$

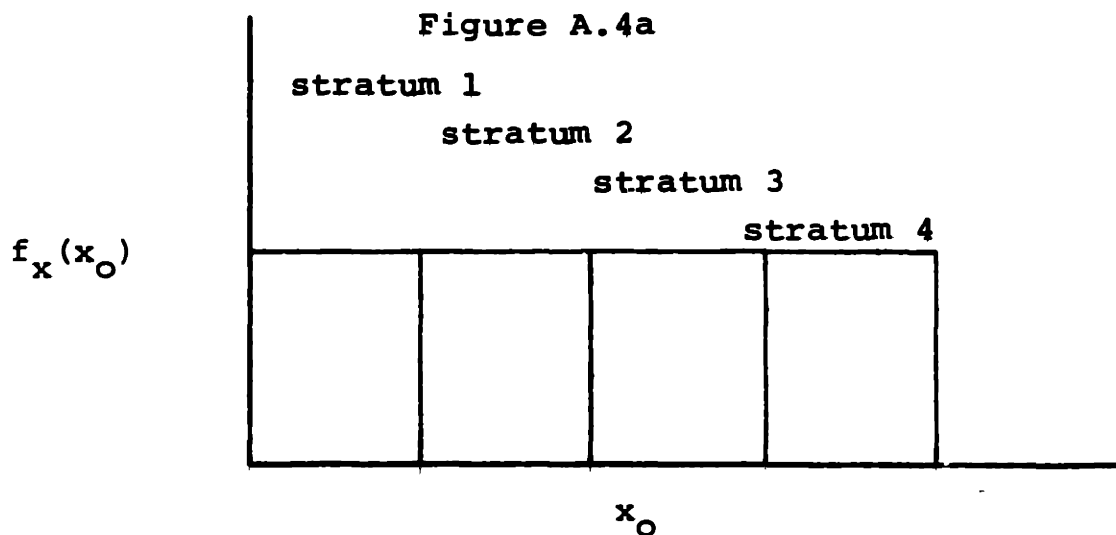
where  $w_h$  is the estimated  $N_h/N$ , and  $W_h$  is the true  $N_h/N$ . The first term is just  $V(\bar{x})$ , the second term accounts for the bias. As  $n$  increases the first term decreases, while the second term remains constant. For very large  $n$ , the precision of the estimate is entirely controlled by the bias error, and simple random sampling becomes more precise (equation A2).

When strata are not obviously separable, some criterion for stratification must be used. Dalenius (1957) and Dalenius and Hodges (1959) have derived an approximate rule which minimizes  $V(\bar{x})$  under Neyman allocation.

If the pdf of  $X$ ,  $f_x(x)$ , is known, the rule says to form the cumulative distribution (cdf) of  $\sqrt{f_x(x)}$  and stratify by



equal intervals of cum  $[\sqrt{f_x(x)}]$  . Naturally, the pdf of  $x$  is not known prior to sampling (otherwise why sample), and often it cannot even be guessed at. In this case the division can best be made, if at all, on the basis of some variable  $y$  which is correlated to  $x$  (assuming such a variable is known). Dalenius (1957) shows that dividing strata on the basis of equal divisions of cum  $[\sqrt{f_y(y)}]$  is an "efficient" stratification for  $x$ , if  $x$  and  $y$  are highly correlated (i.e., linearly related). An "efficient" stratification, in the statistical sense, means that given the existing knowledge of  $f(x)$  this stratification gives the minimum expected squared error of all possible stratifications.



Cochran (1963) presents some numerical data to support the rule in his discussion of the number of strata into which uniform distributions should be divided. The division of more complicated populations has not been extensively treated in the literature. The variance of the mean for populations as shown in Figure A.4a decreases

as the number of strata squared

$$V(\bar{x})_{k \text{ strata}} = \frac{V(\bar{x})_{\text{simple random}}}{k^2}$$

A24

According to Cochran, this relationship holds even for skewed populations as long as they have finite range. Figure A.4b is a plot of cost and variance vs  $k$  for the cost model

$$C = C_0 + C_1 k + C_2 k n$$

A25

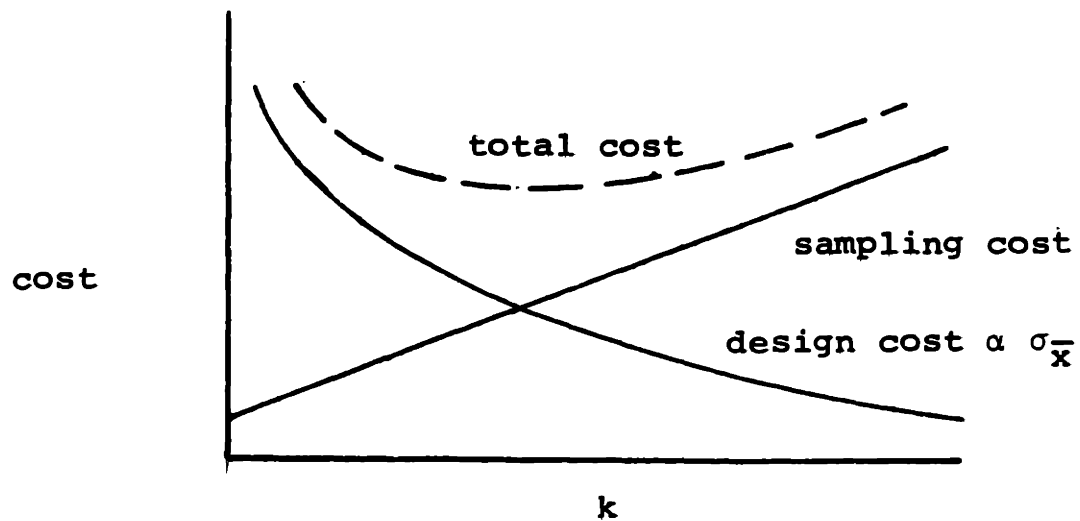


Figure A.4b

If the division is made on some variable whose regression on  $x$  is  $g(y) = E[x|y]$  (i.e.,

$$x = g(y) + \text{error})$$

A26

then

$$S_x^2 = S_g^2 + S_e^2$$

the  $S_g^2$  component of  $V(\bar{x})$  will decrease as

$$V(\bar{x}) = s^2/k^2$$

A27

if the regression of  $x$  on  $y$  is linear, less if non-linear (Cochran, 1963). The  $S_e^2$  component will remain unchanged as  $k$  increases, so, at some size  $k$ ,  $S_e^2$  will dominate and further increases in  $k$  will produce negligible changes in  $V(\bar{x})$ . Cochran presents a simple theoretical model which suggests that if the correlation coefficient between  $x$  and  $y$  is less than 0.95, no appreciable reduction in  $V(\bar{x})$  can be expected beyond  $K \cong 6$ .

#### Sampling one population for two independent variables

The sampling of joint orientations presents an example of the sampling of one population for two independent variables, in this case strike and dip. The difficulty with such plans is that the optimum allocation on one variable may not be the optimal allocation on the other.

For Neyman allocation the easiest approach is to determine the optimum allocation on each variable and average the results. Since the precision of a stratified sampling plan is insensitive to minor deviations from optimal allocation, the averaging procedure often offers almost as much precision as the optimum allocation for each variable. In the case of positive correlation between the two variables, averaging works even better.

For cases where the optimum allocations for each variable are widely different, two allocating techniques have been proposed

by Yates (1960). The first assumes that a loss function exists which depends on the precision of the estimates of the mean for each variable,

$$L(n_h) = a_{x_1} (V(\bar{x}_1)) + \dots + a_{x_k} (V(\bar{x}_k)) \quad A28$$

where  $a_{x_i}$  is the unit loss due to variance in the estimate of  $\bar{x}_i$ .

This equation is written to include the allocation of any number of variables  $x_i$ . The technique minimizes the total cost function

$$\text{total cost} = c(n_h) + L(n_h) \quad A29$$

where  $c(n_h)$  is the sampling cost. The second assumes a desired minimum precision for each variable and allots effort in the most economical way (which may produce greater than minimum precision for some  $x_i$ ).

If the cost of sampling is modeled by equation A25, the first procedure leads to the allocation

$$n_h = \frac{w_h}{\sqrt{c_h}} \sqrt{\sum a_{x_i} s_{x_i h}^2} \quad A30$$

where  $s_{x_i h}^2$  is the variance of  $x_i$  in the  $h$ th stratum.

The second method is approached by determining the cost of satisfying the precision requirements individually. If the allocation with the greatest cost also satisfies every other precision constraint, the problem is solved. If not, the optimum allocation may be determined graphically (Dalenius, 1957) for the condition of

two strata.\* The variance of  $\bar{x}_i$  can be expressed by the equation (Cochran, 1963).

$$v(\bar{x}_i) \geq \sum_{h=1}^k \frac{w_h^2 s_{xih}^2}{n_h} - \sum_{h=1}^k \frac{w_h^2 s_{xih}^2}{N}$$

A31

The locus of points  $(n_1, n_2)$  (i.e., the sizes of the 1<sup>st</sup> and 2<sup>nd</sup> strata) which satisfy equation A31 for each variable  $x_i$  is a hyperbola in the plane  $(n_1, n_2)$  (Figure A.5).

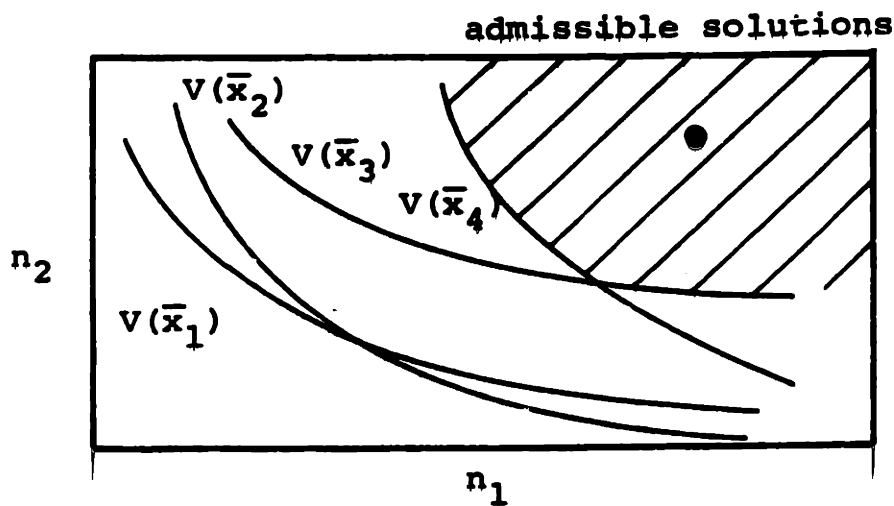


Figure A.5

The region above and to the right of all the curves is the locus of admissible solutions (i.e., strata sizes which satisfy all of the constraints). From this convex set of solutions the pair  $(n_1, n_2)$  with the minimum cost is selected.

\*Yates (1960) presents an analytical approach to this allocation which can handle any number of strata.

Comments

1. Stratified random sampling may be preferable to simple random sampling for several reasons:
  - a. Stratification may offer greater precision for the same sampling effort by allocating effort differently in different strata.
  - b. Field sampling may be easier since attention is concentrated on one stratum at a time.
  - c. Sampling problems and measurement errors may be different in different strata, and they are treated separately (e.g., measurement error in the strike direction of gently dipping joints is greater than in the strike direction of steeply dipping joints. By stratifying according to joint set the different magnitudes of errors are handled separately.).
2. Stratification is convenient in sampling certain populations, but the existence of homogeneous subpopulations is not a basic property of all populations, and stratification does not always lead to increased precision.

3. Since

$$s^2 = s_h^2 + s_{\text{between means}}^2 \quad \text{A32}$$

stratification only leads to greater precision than simple random sampling if the means of the strata are different (i.e.,  $s_{\text{between means}}^2 > 0$ ) -- assuming that the variances within strata are all the same.

4. Comparison of simple random, systematic, and stratified random sampling for some artificial populations:

a. Randomly ordered populations

For randomly ordered populations each of the sampling plans has similar precision. The average of  $V(\bar{x})$  over all possible systematic samples equals  $V(\bar{x})_{\text{simple random}}$ .

b. Linearly trending populations

$$V(\bar{x})_{\text{simple random}} > V(\bar{x})_{\text{systematic}} > V(\bar{x})_{\text{stratified}}$$

c. Periodic populations

When there is no prior information on the periodicity dimension of a periodic population,  $V(\bar{x})_{\text{sys.}}$  cannot be determined. The precision of a systematic sample from a periodic population is dependent on the relationship of  $k$  to  $\lambda$ . For  $k \propto \lambda$  the precision is the same as that obtained by sampling a single element (Figure A.6); for  $k \propto \lambda/2$  the estimate  $\bar{x}$  is exactly equal to the true value. Generally, systematic sampling plans should not be used when periodicity is suspected.

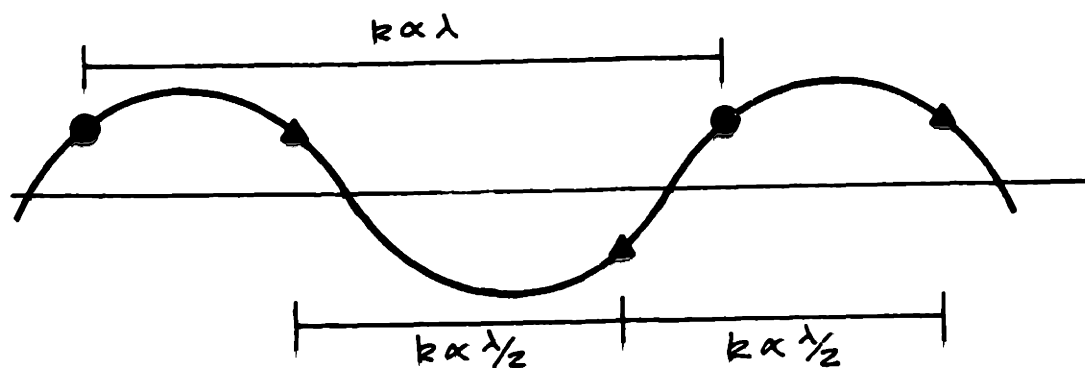


Fig. A.6

d. Autocorrelated Populations

For many geological populations there is reason to believe that elements will be more alike, the more proximal they are in space. Statistically this means the correlation  $\rho_{ij}$  between two elements  $x_i$  and  $x_j$  is a function of their separation distance. Such populations are said to be autocorrelated.

In autocorrelated populations one can show (Cochran, 1946) that stratified random samples have greater precision (i.e., lower  $V(\bar{x})$ ) than simple random samples. If the correlation function of distance is concave upward (Figure 7),

$$\rho_{i+1} + \rho_{i-1} - 2\rho_i > 0 \quad \text{A33}$$

Cochran (1946) has shown that systematic sampling has greater precision than either stratified or simple random sampling.

When

$$\rho_{i+1} + \rho_{i-1} - 2\rho_i = 0 \quad \text{A34}$$

the precisions of systematic and stratified random sampling are equal, and greater than the precision of simple random sampling.



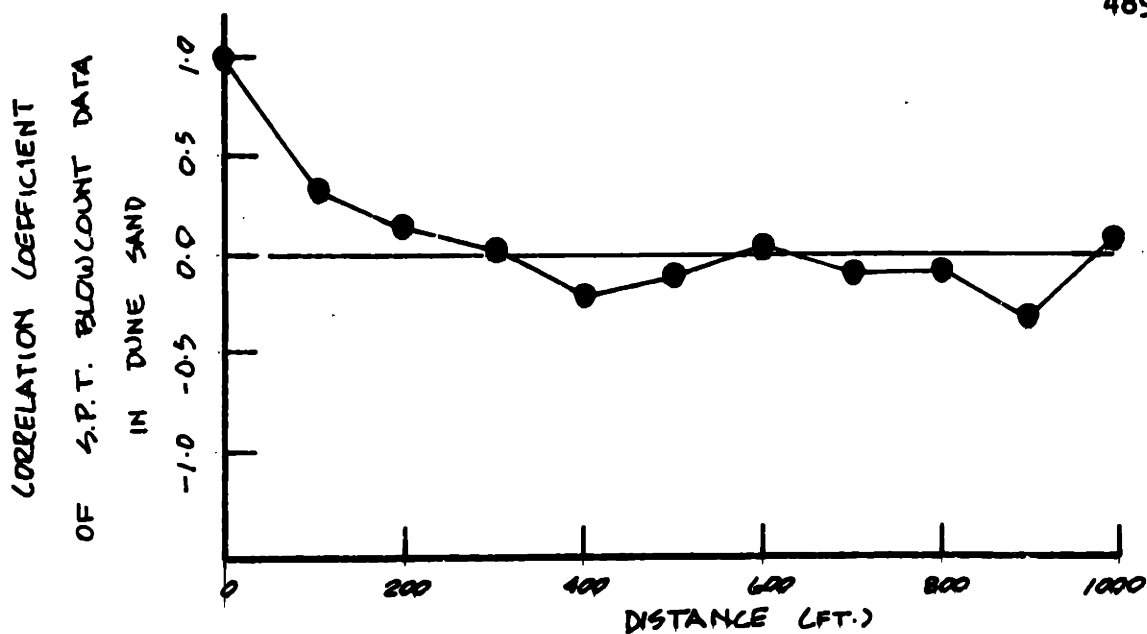


Fig. A.7

Hilldale (1971) presents evidence (Figure A.7) that at least some engineering properties of soils have concave autocorrelation functions. The conclusion one draws is that not only is systematic sampling cheaper, quicker, and easier to specify, but it may also be superior, in terms of greater precision, to stratified random or simple random sampling. As long as the correlogram Figure A.7 is of the form  $e^{-ad}$  ( $a$ , constant;  $d$ , distance -- i.e., concave) a square grid for sampling in two dimensions is similarly superior to two dimensional stratified random or simple random sampling. Matern (1960) discusses sampling with a triangular grid where samples are taken at the vertices.

### 7A.4 Cluster Sampling

In cluster sampling, aggregates, called "clusters," of elements are selected from the sampled population as units rather than as individual elements, and properties of the clusters are determined. From the properties of the clusters, inferences can be made on the total sampled population. Plans which specify to measure every element within clusters are called "single-stage cluster" plans, since they specify only one level of sampling; plans which specify that cluster properties be estimated by simple random sampling are called "two-stage cluster" plans, since they specify two levels of sampling. Higher order cluster plans are sometimes used.

#### Estimators

##### a. Equal-size clusters

We will consider the simplest case of cluster sampling first. Out of  $M$  possible clusters,  $m$  are selected; the ratio  $f_1 = m/M$  is called, as before, the "sampling fraction." Each cluster contains the same number of elements  $N$ ,  $n$  of which are selected for measurement ( $f_2 = n/N$ ).

An unbiased estimate of the true average of each cluster is

$$\bar{x}_i = \frac{\sum x_{ij}}{n_i}$$

A35

where  $x_{ij}$  is the  $j$  th element in the  $i$  th cluster. An unbiased estimate of the true average of the total population is

$$\bar{\bar{x}} = \frac{\sum \bar{x}_i}{m} = \frac{\sum \sum x_{ij}}{m n}$$

A36

and the variance of this estimate is

$$V(\bar{\bar{x}}) = \frac{(1 - f_1)}{m} s_1^2 + \frac{f_1 (1 - f_2)}{n m} s_2^2$$

A37

where  $s_1^2$  is the estimated variance between cluster means,

$$s_1^2 = \frac{\sum (\bar{x}_i - \bar{\bar{x}})^2}{n - 1}$$

A38

and  $s_2^2$  is the estimated variance within clusters

$$s_2^2 = \frac{\sum \sum (x_{ij} - \bar{x}_i)^2}{m(n-1)}$$

A39

An unbiased estimate of the total variance is obtained by summation of variances,

$$s^2 = s_1^2 + s_2^2$$

A40

#### b. Unequal-size clusters

In the more general case not all of the clusters are of equal size. For example, the numbers of joints displayed in different

outcrops are different. With unequal-sized clusters the selection plan for clusters is not as obvious as it was previously. The relative probability,  $z_i$ , of selecting different-sized clusters is now a parameter of the plan. Commonly the  $z_i$  are either taken all equal (simple random sampling of the clusters) or taken proportional to size. The precisions of these two plans are different.

For selection with equal probability an unbiased estimate of the true total population mean is

$$\bar{x} = \frac{\sum N_i \bar{x}_i}{\sum N_i} \quad \text{A41}$$

where  $N_i$  is the total number of elements in the  $i$ th cluster. The values of  $N_i$  need only to be of the correct magnitude relative to one another since they serve only as weighting factors. The estimated variance in  $\bar{x}$  is

$$V(\bar{x}) = \frac{1-f_1}{m-1} \frac{\sum (N_i \bar{x}_i - \bar{x}_m)^2}{(m-1)} + \frac{f_1}{m^2 \bar{N}^2} \sum \left[ \frac{N_i^2 (1-f_{2i}) s_{2i}^2}{n_i} \right] \quad \text{A42}$$

where  $\bar{N}$  is the average value of  $N_i$ , and  $\bar{x}_m = \sum N_i \bar{x}_i / m$ . If the assumption is made that  $n_i \propto N_i$  (i.e.,  $f_{2i} = n_i / N_i = \text{constant}$ ) the sampling plan is self-weighting and equation A42 simplifies to

$$V(\bar{x}) = \frac{1-f_1}{m \bar{N}^2} \frac{\sum (N_i \bar{x}_i - \bar{x}_m)^2}{(m-1)} + \frac{f_1(1-f_2)}{m \bar{n} \bar{N}} \sum N_i s_{zi}^2$$

A43

For our purposes this assumption (i.e.,  $n_i \propto N_i$ ) is frequently valid. Proportionally more joints are usually sampled from larger outcrops than smaller outcrops.

For selection with probability proportional to size an unbiased estimate of the true total population mean is

$$\bar{x} = \frac{1}{n} \sum \bar{x}_i$$

A44

and the estimated variance in  $\bar{x}$  is,

$$V(\bar{x}) = \frac{\sum (\bar{x}_i - \bar{x})^2}{m(n-1)}$$

A45

For selection with arbitrary probabilities or probabilities proportional to some measure of size an unbiased estimate of the true mean of the total population is

$$\bar{x} = \frac{1}{m N_0} \sum \frac{N_i \bar{x}_i}{z_i}$$

A46

where  $z_i$  is the probability of selecting the  $i$  th cluster, and

$N_0 = \sum N_i$  Again the magnitude of  $N_i$  need only be in the correct proportions relative to one another. The variance of  $\bar{x}$  is

$$V(\bar{x}) = \frac{\sum (N_i \bar{x}_i / z_i - \bar{x}')^2}{m(m-1)N_0}$$

A47

where  $\bar{x}'$  is the unweighted mean of the  $N_i \bar{x}_i / z_i$

$$\bar{x}' = \frac{1}{m} \sum N_i \bar{x}_i / z_i$$

A48

In all cases the variance of the total population can be estimated from the variances between elements within clusters and the variance between the means of the clusters.

$$\sigma^2_{(total)} = \sigma^2_{\left\{ \begin{array}{l} \text{between} \\ \text{cluster} \\ \text{means} \end{array} \right\}} + \sigma^2_{\left\{ \begin{array}{l} \text{between} \\ \text{elements} \\ \text{within} \\ \text{clusters} \end{array} \right\}}$$

A49

### Optimization

#### 1. Equal-sized clusters

There are two parameters to be allocated in cluster sampling: the number of clusters  $m$ , and the number of observations within each cluster,  $n$ .

Assuming a cost model of the form

$$C = C_1 m + C_2 m n$$

A50

( $C_1$  can be thought of as the set-up cost per cluster) the optimum  $m$  can be shown to be (Cameron, 1951)

$$m_{opt} = \frac{s_2}{(s_1^2 - s_2^2 / M)^{1/2}} (C_1 / C_2)^{1/2}$$

A51

Obviously,  $m$  must be an integer. Once  $m$  is determined,  $n$  is found by substituting into equation A50.

The evaluation of  $m_{opt}$  requires estimates of  $S_1^2$  and  $S_2^2$  which may not be accurately known prior to sampling. Fortunately, the precision vs.  $m$  function is relatively flat near  $m_{opt}$  so errors in the estimates of  $S_1^2$  and  $S_2^2$  cause only small losses of precision. Brooks (1955) has constructed tables of the limits of  $S_2^2/S_u^2$ , where

$$S_u^2 = S_1^2 - S_2^2/M$$

for 90% of optimum precision (Table A.4).

If an estimate of  $S_1^2 - S_2^2/M$  is known, Table A.4 can be used to select a value of  $m$ .

### Example

Samples are to be taken to determine accurately the average strike direction of a set of vertical joints in the foundation of a dam. The standard deviation within each cluster of sampled joints is expected to lie between  $15^\circ$  and  $30^\circ$  ( $S_2$ ) and the standard deviation between the means of the clusters is expected to lie between  $5^\circ$  and  $10^\circ$  ( $S_1$ ). Assuming that the number of outcrops which could be measured at each outcrop is very large, the range of  $S_2^2/S_u^2$  is expected to be

$$2.25 \leq S_2^2 / S_u^2 \leq 36$$

From Table A.4 with  $c_1/c_2 = 64$ :

$$m_o = 25$$

Therefore, use 25 clusters.

## 2. Unequal-sized clusters

For clusters selected with equal probability

$$m_{opt} \cong \frac{s_2 (C_1/C_2)^{1/2}}{(S_b^2 - s_2^2/\bar{M})^{1/2}}$$

which is identical to equation A51 with  $M$  replaced by  $\bar{M}$ , and  $S_1^2$  replaced by its weighted value,

$$S_b^2 = \frac{\sum M_i^2 (\bar{x}_i - \bar{x})^2}{\bar{M}^2 (N-1)}$$

Table A.4 can be used as before, once these substitutions are made.

The theory of optimal sample size is considerably more complicated for cluster sampling with non-equal selection probabilities than for equal probabilities. Hansen and Hurwitz (1949) developed a solution which simultaneously gives the optimum probabilities  $z_i$  and the optimum sampling and subsampling fractions  $f_1$  and  $f_2$ . The solution is complicated and will not be discussed here, but Kendall and Stuart (1968, v.3), Hansen, et al (1953), and Cochran (1963) all have discussions of the solution.

### Comments

1. Joint surveys are usually based on cluster plans because the cost (in time spent) of sampling many individual joints on one outcrop (which may be thought of as a cluster) is less than the cost of traveling between outcrops.



2. Cluster methods may require less sampling effort than other plans, but the sampling of smaller units (e.g., individual elements like joints) generally offers greater precision: a trade-off between cost and precision has to be considered. This section is only a review of the general application of cluster sampling. The theory has been extensively developed for use in the social sciences, and a reader wishing an extensive presentation should refer to Hansen, et al (1953), or Cochran (1963), both of whom devote several chapters to developing cluster sampling theory.
3. Nested Analysis of Variance

The variance of geological populations is usually a function of the spatial extent of the population. If we consider the strength of rock specimens taken from outcrops, the variance between specimens all taken from the same outcrop is usually smaller than the variance between specimens taken from many outcrops in one area of the site, which, in turn, is smaller than the variance between specimens taken from all across the site. These are nested levels of variance and cluster techniques allow us to evaluate variance as a function of the "extent" of the distribution. Excellent examples of the analysis of nested levels of variance in geology are to be found in Olson and Potter (1954) and in Potter and Siever (1955), in which the authors analyzed variance in cross-bedding direction as a function of spatial extent.

To analyze nested levels of variance we must first assume a model for the total population. Consider Figure 8 -- an abstract map of a population, which has been divided into many small elements, as illustrated by the cross-hatched one. The variance within each of these "lowest level" elements will be denoted  $s_4^2$ . This is the variance of the sampled population within the spatial extent of a lowest level element (we consider 4 levels of nested elements, but any number could have been considered). The means of lowest level elements within each 2<sup>nd</sup> lowest level element (identified by the subscript  $j$  in Figure A.8) are themselves distributed with some variance,  $s_3^2$ , which can be calculated from sample data; and so forth.

The value of a member of the population, therefore, equals some overall or "grand" mean plus random components accounting for the variance at all four levels. Mathematically this can be represented by

$$x_{ijkl} = \mu + a_i + b_{ij} + c_{ijk} + d_{ijkl}$$

A40

where  $\mu$  is the grand mean,  $a_i$  is the error associated with the highest level,  $b_{ij}$  the next highest level, ..., and  $d_{ijkl}$  the error with the lowest level. The subscripts  $i, \dots, l$  refer to the elements at the 1<sup>st</sup> through 4<sup>th</sup> level. The averages of the random components are all zero,

$$E[a_i] = E[b_{ij}] = E[c_{ijk}] = E[d_{ijkl}] = 0$$

A41

and the variances are,

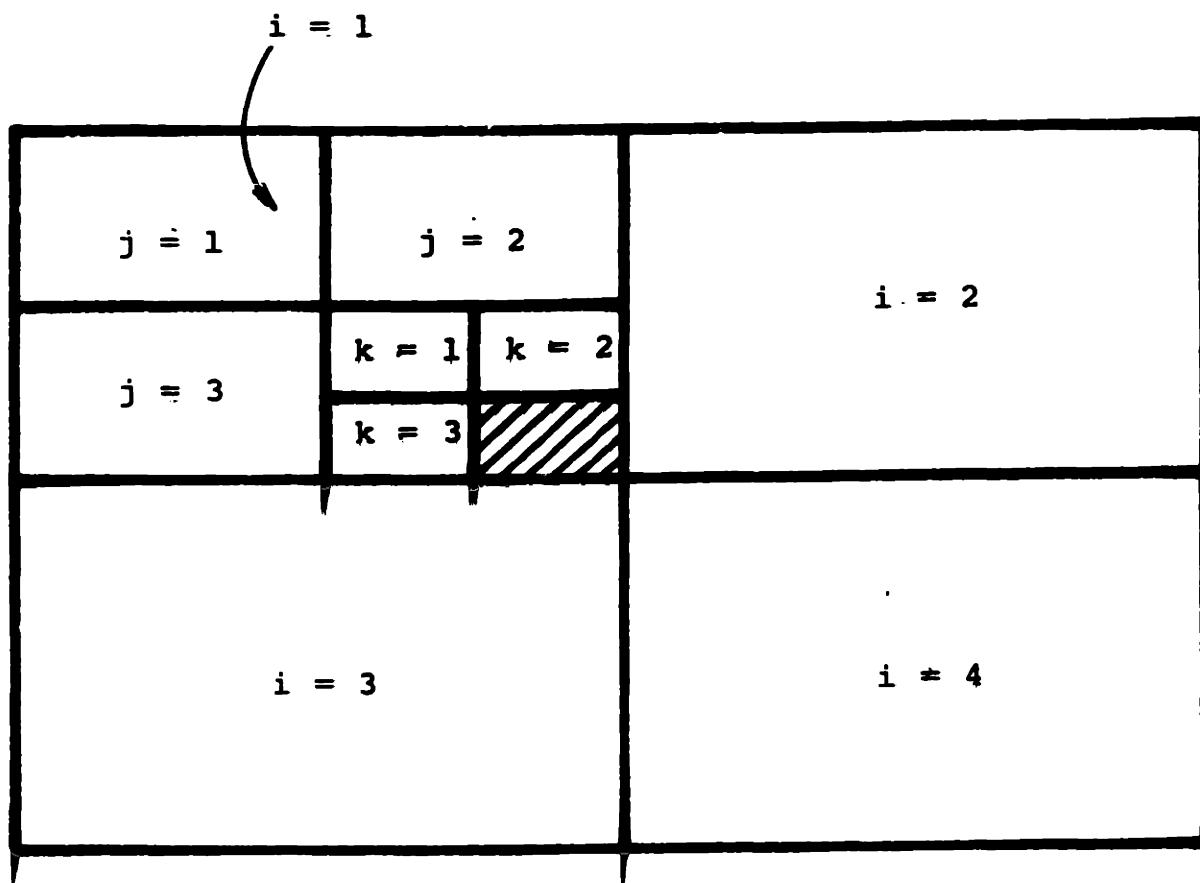


Figure A.8

Abstract map of a population divided into 64 "lowest level" elements (of which 4 are shown).

$$\begin{aligned}
 V(a_i) &= s_1^2 \\
 V(b_{ij}) &= s_2^2 \\
 V(c_{ijk}) &= s_3^2 \\
 V(d_{ijkl}) &= s_4^2
 \end{aligned}$$

A42

By summation of variances the total variance at any level equals the sum of the variances at lower levels. For example, the total variance between members of the population within one 3<sup>d</sup> level element is  $s_4^2 + s_3^2$ ; within one 2<sup>nd</sup> level element

$$s_4^2 + s_3^2 + s_2^2 .$$

The variance in the estimate of the grand mean is

$$V(\bar{\bar{X}}) = \frac{(1-f_1)}{n} s_1^2 + \frac{(1-f_2)f_1}{nm} s_2^2 + \frac{(1-f_3)f_2f_1}{nm} s_3^2 + \dots$$

A43

where  $f_i$  is the sampling fraction at level  $i$ , and  $n$ ,  $m$ ,  $k$  are the sizes of the samples at the 1<sup>st</sup>, 2<sup>nd</sup>, and 3<sup>d</sup> levels, respectively. Formulas for estimating variances and means at each level are given in Table A.5.

Table A.5  
Nested analysis of variance

---


$$v(\bar{x}) = \frac{(1-f_1) s_1^2}{n} + \frac{(1-f_2) f_1 s_2^2}{nm} + \frac{(1-f_3) f_2 f_1 s_3^2}{nmk} + \dots$$


---

Variance among the  
 $n$  primary means

$$s_1^2 = \frac{\sum (\bar{x}_i - \bar{\bar{x}})^2}{n-1}$$


---

Variance among the  
 $nm$  2<sup>d</sup> level means

$$s_2^2 = \frac{\sum \sum (\bar{x}_{ij} - \bar{\bar{x}})^2}{n(m-1)}$$


---

Variance among the  
 $nmk$  3<sup>d</sup> level means

$$s_3^2 = \frac{\sum \sum \sum (x_{ijk} - \bar{\bar{x}})^2}{nm(k-1)}$$


---

Variance among measur-  
ments at  $h^{\text{th}}$  level element

$$s_h^2 = \frac{\sum \dots \sum (x_{ijk\dots} - \bar{\bar{x}})^2}{nmk \dots (q-1)}$$


---

## REFERENCES

- Agterberg, E.P. (1966) "Trend surfaces with autocorrelated residuals", Pa. St. Univ., Min. Ind., 1: H-1 - H-19.
- Agterberg, F.P. (1970) "Autocorrelation functions in geology", in D.F. Merriam, ed., Geostatistics, Plenum Press, New York, pp. 113-142.
- Alpert, M. and H. Raiffa (1969) A Progress Rept. on Training Probability Assessors, (first draft).
- Anderson, T.W. (1960) "A modification of the sequential probability ratio test to reduce sample size", Ann. Math'l Stat., 31.
- Arnold, K.J. (1941) On Spherical Probability Distributions, Ph.D. dissertation, M.I.T.
- Attneave, F. (1953) "Psychological probability as a function of experienced frequency", Jour. Exp. Psych., 46: 81-86.
- Baecher, G.B. (1971) Oil Exploration in Shoestring Sands: A Bayesian Approach, Geology 147 Term Paper, Harvard University.
- Barnard, G.A. (1958) "Thomas Bayes - A biographical note", in E.S. Pearson and M.G. Kendall (eds.), Studies in the History of Statistics and Probability, Griffin, London, 481 p.
- Bartlett, M.S. (1966) Introduction to Stochastic Processes, Cambridge Univ. Press, Cambridge, England, 362 pp, 2 ed.
- Bayes, T. (1763) "An essay towards solving a problem in the doctrine of chances", 46, 1958.
- Bellman, R. (1957) Dynamic Programming, Princeton Univ. Press, N.J., 112 p.
- Benjamin, J.R., and C.A. Cornell (1970) Probability, Statistics, and Decision for Civil Engineers, McGraw-Hill, N.Y., 684 pp.
- Bernoulli, Jacob (1713) Ars Confectandi, Basel.
- Black, W.L. (1962) Sequential Search, Masters thesis, Dept. of Elec. Eng'g., M.I.T.
- Blackwell, D. (1951) "Comparison of experiments", Proc. 2nd Berkeley Symp. Math. Statist. Probability, Univ. of Calif. Press, Berkeley, Calif., pp. 93-102.

- Blackwell, D. (1953) "Equivalent comparison of experiments", Ann. Math'l. Statist., 24: 265-272.
- Bradt, R.N., and S. Karlin (1956) "On the design and comparison of certain dichotomous experiments", Ann. Math. Statist., 27: 390-409.
- Brillouin, L. (1964) Scientific Uncertainty and Information, Academic Press, N.Y. 351 p.
- Brooks, S. (1955) "The estimation of an optimum subsampling number", Jour. Amer. Stat. Assn., 50: 398-415.
- Brown, A.A. (1960) "Search theory and problems of exploration boring", Penn. State Univ., Ind. Exp. Sta. Bull., 72: 33-37.
- Busk, H.G. (1929) Earth Flexures, Cambridge Univ. Press, 106 p.
- Cardillo, G.P., and K.S. Fu (1967) "A dynamic programming procedure for sequential pattern classification and feature selection", Math'l Biosciences, 1: 463-491.
- Charnes, A. and W.W. Cooper (1958) "The theory of search: optimum distribution of search effort", Management Sci., 5: 44-50.
- Chayes, F. (1949) "Statistical analysis of fabric diagrams", in H.W. Fairbairn, Structural Petrology of Deformed Rocks, Addison-Wesley Press, Cambridge, Mass. pp.297-326.
- Chew, M.C. Jr. (1967) "A sequential search procedure", Ann. Math'l Stat., 38: 494-502.
- Chien, Y.T., and K.S. Fu (1967) "An optimal pattern classification system using dynamic programming", Math. Biosciences, 1:439-461.
- Churchman, C.W., R.L. Ackoff, E.L. Arnoff (1957) Introduction to Operations Research, Wiley, N.Y., 645 p.
- Clapham, A.R. (1936) "Over-dispersion in grassland communities and the use of statistical methods in plant ecology", Jour. Ecology, 24: 232-251.
- Clark, P.J. (1956) "Grouping in spatial distributions", Sci., 123: 373-374.
- Cochran, W.G. (1936) "The  $\chi^2$  distribution for the binomial and Poisson series, with small expectations", Annals of Eugenics, 7: 207-217.

- Cochran, W.G. (1946) "Relative accuracy of systematic and stratified samples for a certain class of population", Ann. Math'l. Statist., 17: 164-177.
- \_\_\_\_\_ (1963) Sampling Techniques, John Wiley & Sons, N.Y., 413 p.
- \_\_\_\_\_, F. Mosteller, and J.W. Tukey (1954) Statistical Problems of the Kinsey Rept., Am. Stat. Assn., Washington, D.C., 280 p.
- Cohen, J. and Hansel, G.E.M. (1955) "Gambling and subjective probability", Br. Jour. Psych., 46: 178.
- Coomb, C.H., and D.C. Beardsley (1954) "Decision making under uncertainty", in, R.M. Thrall, C.H. Coombs, and R.L. Davis (eds.), Decision Processes, Wiley, N.Y.
- Cornell, C.A. (1971) "First order uncertainty analysis of soils, deformation and stability", Conference on Applications of Statistics and Probability to Soil and Structural Engineering, University of Hong Kong (in press).
- Cox, D.R., and H.D. Miller (1965) The Theory of Stochastic Processes, Methuen, London, 398 p.
- Cramer, H. (1955) The Elements of Probability Theory, Wiley, N.Y., 281 p.
- Dale, H.C.A. (1958) "Subjective probability, gambling, and intelligence", Nature, 181: 363-364.
- \_\_\_\_\_ (1959) "A priori probabilities in gambling", Nature, 183: 842-843.
- Dalenius, T. (1957) Sampling in Sweden, Almqvist and Wicksell, Stockholm.
- \_\_\_\_\_ and J.L. Hodges, Jr. (1959) "Minimum variance stratification", Jour. Amer. Stat. Assn., 54: 88-101.
- Danskin, J.M. (1962) "A theory of reconnaissance I", Operations Research, 10: 285-299.
- \_\_\_\_\_ (1964) "On Koopman's addition theorem in search theory", Institute of Naval Studies, Cambridge, Mass.



- Davidson, D., P. Suppes, and S. Siegel (1957) "Decision-making: an experimental approach", Stanford University Press, 121 p.
- Deere, D.U., and R.P. Miller (1966) "Engineering classification and index properties for intact rock", Tech. Rept. No. AFWL-TR-65-116, Air Force Weapons Lab, Kirkland Air Force Base, New Mexico.
- de Finette, B. (1937) "Foresight: its logical laws, its subjective sources", Translation in H.E. Kyburg and H.E. Smokler, Studies in Subjective Probability, John Wiley and Sons, N.Y. pp. 93-158, 1964.
- De Groot, M.H. (1970) Optimal Statistical Decisions, McGraw-Hill, N.Y. 489 p.
- de Guenin, J. (1961) "Optimum distribution of effort: an extension of the Koopman basic theory", Operations Research, 9: 1-7.
- De Lury, D.B. (1950) Values and Integrals of the Orthogonal Polynomials Up to  $n = 26$ , Univ. of Toronto Press, 33 p.
- Dobbie, J.M. (1963) "Search theory: a sequential approach", Naval Res. Logist. Qt., 10:323-334.
- Dowds, J.P. (1961) "Mathematical probability is an oil-search tool", World Oil, 153(4): 99-106.
- Drew, J.L. (1966) Grid Drilling Exploration and its Application to the Search for Petroleum, Ph.D. thesis, Penn. State Univ.
- Duda, R.O., and H. Fossum (1966) "Pattern classification by iteratively determined linear and piece wire linear discriminant functions", I.E.E.E. Trans. Elec. Comp., EC-15: 221-232.
- Edwards, W. (1953) "Probability-preferences in gambling", Amer. Jour. Psych., 66: 349-364.
- \_\_\_\_\_ (1955) "The prediction of decisions among bets", Jour. Exp. Psych., 51: 201-214.
- \_\_\_\_\_ (1956) "Reward probability, amount, and information as determiners of sequential two-alternative decisions", Jour. Exp. Psych., 52: 177-188.
- Ellsberg, D. (1961) "Risk, ambiguity, and the Savage axioms", Qt. Jour. Economics, 75: 643-669.

- Engle, J.H. (1957) "Use of clustering in mineralogical and other surveys", Proc. 1st Intl. Conf. on Operations Res., Oxford, The English Univ. Press, pp. 176-192.
- Enslow, P.H., Jr. (1966) "A bibliography of search theory and reconnaissance theory literature", Nav. Res. Logist. Qt., 13: 177-202.
- Feller, W. (1957) An Introduction to Probability Theory and Its Application, 2 ed, John Wiley & Sons, N.Y., 2 vols.
- Fellmer, W. (1961) "Distortion of subjective probabilities as a reaction to uncertainty", Qt. Jour. Econ., 75: 670-689.
- Fishburn, P.C. (1964) Decision and Value Theory, John Wiley and Sons, N.Y., 459 p.
- Fisher, R.A. (1922) "On the interpretation of  $\chi^2$  from contingency tables, and on the calculation of p", Jour. Roy. Stat. Soc., 85: 87-89.
- \_\_\_\_\_ (1932) Statistical Methods for Research Workers, Oliver and Boyd, London, 354 p.
- \_\_\_\_\_ (1953) "Dispersion on a sphere", Proc. Roy. Soc. of London, Series A, 217: 295-305.
- \_\_\_\_\_ (1959) Statistical Methods and Scientific Inference, Hafner, N.Y., 178 p.
- \_\_\_\_\_ (1970) Statistical Methods for Research Workers, Hafner, Darien, Conn., 362 p.
- Fisher, R.A., H.G. Thornton, & W.A. Mackenzie (1922) "The accuracy of the plating method of estimating bacterial populations", Ann. Appl. Biol., 9: 325.
- Folayan, J.I., K. Hoeg, and J.R. Benjamin (1970) "Decision theory applied to settlement predictions", Jour. Soil Mech. and Found Div., A.S.C.E., 96: 1127-1141.
- Fookes, P.G. (1969) "Geotechnical mapping of soils and sedimentary rock for engineering purposes with examples of practice from the Mangla Dam project", Geotechnique, 19: 52-74.
- Friedman, M., and L.J. Savage (1947) Selected Techniques of Statistical Anal., McGraw-Hill, N.Y.

Fu, K.S. (1968a) Sequential Methods in Pattern Recognition and Machine Learning, Academic Press, N.Y., 227 p.

\_\_\_\_\_ (1968b) "Relationship among various learning techniques in pattern recognition systems", in, Pattern Recognition, L.N. Kanal, ed., Thompson Book Co., Washington, D.C., 399-408.

\_\_\_\_\_ (1970) "Statistical pattern recognition", in, J.M. Mendal and K.S. Fu (eds.), Adaptive, Learning, and Pattern Recognition Systems, Academic Press, N.Y.

\_\_\_\_\_, Y.T. Chien, and G.P. Cardillo (1967) "A dynamic programming approach to sequential pattern recognition", I.E.E.E. Trans. on Elec. Comp., EC-16: 790-803.

Galanter, E.H. & W.A.S. Smith (1958) "Some experiments on a simple thought problem", Amer. Jour. Psych., 71: 359-366.

Gluss, B. (1959) "An optimum policy for detecting a fault in a complex system", Operations Research, 7: 468-477.

Good, I.J. (1950) Probability and the Weighting of Evidence, Griffin, London.

Goodnow, J.J. (1955) "Determinants of choice-distribution in two-choice situations", Amer. Jour. Psych., 68: 106-116.

Grayson, C.J., Jr. (1960) Decisions Under Uncertainty: Drilling Decisions by Oil and Gas Operators, Harvard Univ. Graduate School of Business Admin., Boston, Mass., 402 p.

Gumbel, E.J., J.A. Greenwood, and D. Durand (1953) "The circular normal distribution", Jour. Amer. Statist. Assn., 48: 131-152.

Hansen, M.H., and W.N. Hurwitz (1949) "On the determination of the optimum probabilities in sampling", Ann. Math'l. Statist., 20: 426-432.

\_\_\_\_\_, W.N. Hurwitz, and W.G. Madow (1966) Sample Survey Methods and Theory, 2ed., John Wiley and Sons, N.Y., 2 v.

Haralick, R.M., and I. Dinstein (1971) "An iterative clustering procedure", I.E.E.E. Trans., SMC-1: 275-289.

Harbaugh, J.W., and D.F. Merriam (1968) Computer Applications in Stratiagraphic Analysis, John Wiley and Sons, N.Y., 282 p.

- Hazen, S.W., Jr. (1967) "Some statistical techniques for analyzing mine and mineral deposit sample and assay data", U.S. Bureau of Mines Bull., #621.
- Hilldale, C. (1971) A Probabilistic Approach to Estimating Differential Settlement of Footings on Sand, Masters Thesis, M.I.T., Dept. of Civil Eng'g.
- Huijbregts, C., and G. Matheron (1970) "University kriging", 9th Intl. Sym. on Tech. for Decision Making in the Min. Industry (in press).
- Hume, D. (1739) A Treatise on Human Nature, London.
- Irwin, F.W. (1953) "Stated expectations as functions of probability and desirability of outcomes", Jour. Personality, 21: 329-335.
- Jeffreys, H. (1961) Theory of Probability, Oxford Univ. Press, 459 p.
- Keeney, R.L. (1969) "Multidimensional utility functions", Operations Research Center, M.I.T., Tech. Rept. #43, 178 p.
- Kendall, M.G., and A. Stuart (1969) The Advanced Theory of Statistics, Hafner, N.Y., VI-3.
- Keynes, J.M. (1921) A Treatise on Probability, Macmillan, London, 466 p.
- Kiraly, L. (1969) Statistical Analysis of Fractures, Sonderdruck aus der Geol. Rundschau, 59: 124-151.
- Knill, J.L. & K.S. Jones (1965) "The recording and interpretation of geological conditions in the foundations of the Roseires, Kariba, and Latiyan dams", Geotechnique, V15: 94-124.
- Koch, G.W. and R.F. Link (1971) Statistical Analysis of Geological Data, Wiley, N.Y., V.2.
- Kolmogoroff, A.N. (1933) The Foundations of the Theory of Probability (trans., 1950) Chelsea Publ. Co., N.Y.
- \_\_\_\_\_ (1941) "Interpolation and extrapolation", Bulletin de l'academie des sciences de U.S.S.R., Ser Math 5:3-14.
- Koopman, B.O. (1940) "The bases of probability", Reproduced in Kyburg, H.E., and H.E. Smokler, Studies in Subjective Probability, John Wiley and Sons, N.Y., 1963.

Koopman, B.O. (1946) "Search and screening", Operations Evaluation Group, Office of the Chief of Naval Operations, Washington, D.C., OEG Rept. #56 (ATI 64 627), 172 pp.

\_\_\_\_\_ (1956a) "The theory of search: I. Kinematic bases", Operations Research, 4:324-346.

\_\_\_\_\_ (1956b) "The theory of search: II. Target detection", Operations Research, 4: 503-531.

\_\_\_\_\_ (1957) "The theory of search: III. The optimum distribution of searching effort", Operations Research, 5: 613-626.

Krumbein, W.C., and L.L. Sloss (1963) Stratigraphy and Sedimentation, Freeman and Co., San Francisco, 660 p.

\_\_\_\_\_ and F.A. Graybill (1965) An Intro. to Statistical Models in Geology, McGraw-Hill, N.Y., 475 p.

Lahee, F.H. (1961) Field Geology, 6th ed., McGraw-Hill, N.Y., 926 p.

Laplace (1825) A Philosophical Essay on Probabilities, 5th ed., Bachelier, Paris.

Larsson, I. (1952) "A graphic testing procedure for point diagrams", Am. Jour. Sci., 250: 586-593.

Le Roy, L.W. (ed.) (1951) Subsurface Exploration, Colo. Sch. of Mines, Golden, Colo., 1156 p.

Levorsen, A.I. (1967) Geology of Petroleum, 2ed., Freeman and Co., San Francisco, 695 p.

Lewis, D.M. (1962) "The characteristic selection problem in recognition systems", I.R.E. Trans on Information Theory, 8: 171.

Lovering, T.S., and E.N. Goddard (1950) "Geology and ore deposits of the Front Range, Colorado", U.S.G.S. Prof. Paper 223.

Luce, R.D. & P. Suppes (1965) "Preference, utility, and subjective probability", in, R.D. Luce, et al (eds.), Handbook of Math'l. Psychology, John Wiley and Sons, N.Y., V3: 249-410.

McMahon, B.K. (1968) A statistical study of rock slopes in jointed gneiss with reference to highway rock slope design, Colo. School of Mines, Basic Engineering Dept.

- MaMahon, B. (1971) "A statistical method for the design of rock slopes", Dames and Moore Bulletin, No. 39: 1-9.
- Matern, B. (1947) "Methods of estimating the accuracy of line and sample plot surveys", Medd. fr. Statens Skogsforsknings Institut, 36: 1-138.
- \_\_\_\_\_ (1960) "Spatial Variation", Medd. fr. Statens Skogsforsknings Institute, 49(5): 1-144.
- Mattson, R.L., and J.E. Dammann (1965) "A technique for determining and coding subclasses in pattern recognition problems", I.B.M. Jour. of R and D, 9:294-302.
- Matula, D. (1964) "A periodic optimal search", Amer. Math. Monthly, 71: 15-21.
- Medawar, P.B. (1968) Induction and Intuition in Scientific Thought, Methuen, London, 62 p.
- Mela, D.F. (1961) "Information theory and search theory as special cases of decision theory", Operations Research, 9: 907-909.
- Miller, V.C. (1962) "Statistical prediction by discriminant analysis", Meteorological Monographs, 4 #25.
- Mitchel, J.K. (1963) "Engineering properties of San Francisco Bay Mud", Calif. Div. Mines, Special Rept. #82: 25-32.
- Morse, P.M. (1970) "On browsing: the use of search theory in the search for information", Operations Research Center, M.I.T., Tech. Rept. #50, 35 p.
- Mosteller, F. & P. Noguee (1951) "An experimental measurement of utility", Jour. Polit. Econ., 59: 371-404.
- Müller, L. (1933) "untersuchungen über statistische Kluftmessung", Geol. u. Bauw., 5: 185.
- Myers, J.L., J.G. Fort, L. Katz, and M.M. Suydam (1963) "Differential monetary gains and losses and event probability in a two-choice situation", Jour. Exp. Psych., 66: 521-522.
- Nagel, E. (1938) "Principles of the theory of probability", Intl. Encyclopedia of Unified Science, V.1, #6.
- Nagy, G. (1968) "State of the art in pattern recognition", I.E.E.E. Proc. 56: 836-862.

Neyman, J. (1934) "On the different aspects of the representative method", Jour. Roy. Stat. Soc., 97: 558-606.

\_\_\_\_\_ and E.S. Pearson (1931) "Further notes on the  $\chi^2$  distribution", Biometrika, 22: 298-305.

\_\_\_\_\_ (1933) "On the Problem of the Most Efficient Tests of Statistical Hypotheses", Phil. Trans. Roy. Soc., A, 231: 289-337.

Nicks, D.C. (1959) "Prediction of sequential two-choice decisions from event runs", Jour. Exp. Psych. 57: 105-114.

Noren, N. (1969) Long-Range Decisions in Mining, The Economic Research Institute, Stockholm School of Economics, 351 p.

Olson, J.S., and P.E. Potter (1954) "Variance components of cross-bedding direction in some basal Pennsylvanian sandstones of the Eastern Interior Basin", J. Geol., 62: 26-49.

Osborne, J.G. (1942) "Sampling errors of systematic and random surveys of cover-type areas", Jour. Amer. Stat. Assn., 37: 256-264.

Phillips, F.C. (1956) The Use of Stereographic Projection in Structural Geology, Edward Arnold, London, 86 p.

Phillips, L.D. & W. Edwards (1966) "Conservatism in a simple probability inference task", Jour. Exp. Psych., 72: 346-354.

Pierce, C. (1960) Collected Papers, V4, Harvard Univ. Press, Cambridge, Mass.

Pincus, H.J. (1961) "Statistical Methods applied to rock fractures", Bull. Geol. Soc. Amer., 62: 81-130.

Poincare, H. (1914) Science and Method, Nelson and Sons, London.

Posner, E.C. (1963) "Optimal search procedures", I.E.E.E. Trans. on Info. Theory, IT-9: 157-160.

Potter, P.E. and R. Siever (1955) "A comparative study of Upper Chester and Lower Pennsylvanian stratigraphic variability", J. Geol., 63: 429-451.

Pratt, J.W., H. Raiffa, and R. Schlaifer (1965) Introduction to Statistical Decision Theory, McGraw-Hill, N.Y., (preliminary ed.).

- Preston, M.G. & P. Baratta (1948) "An experimental study of the auction-value of an uncertain outcome", Amer. Jour. Psych., 61: 183-193.
- Price, N.Y. (1966) Fault and Joint Development in Brittle and Semi-Brittle Rock, Pergamon, N.Y., 176 p.
- Raiffa, H. (1968) Decision Analysis, Addison-Wesley, Reading, Mass., 309 p.
- \_\_\_\_\_ and R. Schlaifer (1961) Applied Statistical Decision Theory (student edition), M.I.T. Press, Cambridge, Mass., 356 p.
- Ramsay, J.G. (1967) Folding and Fracturing of Rocks, McGraw-Hill, N.Y., 568 p.
- Ramsey, F.P. (1926) "Truth and Probability", Reproduced in Kyberg, H.E. and H.E. Smokler, Studies in Subjective Probability, John Wiley and Sons, N.Y., 1963.
- Rao, C.R. (1965) Linear Statistical Inference and its Application, Wiley, N.Y., 522 p.
- Robertson, A.M., and D.R. Piteau (1971) "The determination of joint populations and their significance for tunnel stability", in The Technology and Potential of Tunnelling, N. Look, ed. (in press).
- Robinson, C.S., and F.T. Lee (1965) "The validity of geologic projection; a successful example: the Straight Creek Tunnel pilot bore, Colo.", U.S.G.S. Repts., Open file ser. #803, 65 p.
- Robinson, W.O. (1971) "Multivariate probability assessment", Operations Research Center, M.I.T., Tech. Rept. #60, 121.
- Romanova, M.A. (1970) "Trend analysis of geological data--basic literature", in Topics in Math'l Geol. (Romanova and Sarfmnov, eds.), Consultants Bureau, p. 273-278.
- Salmon, W.C. (1966) The Foundations of Scientific Inference, Univ. of Pitts. Press, 157 p.
- Sander, B. (1948) Einführung in die Gefuegekunde, part I, Springer-Verlag, Vienna.
- Savage, L.J. (1954) The Foundations of Statistics, John Wiley and Sons, N.Y., 294 p.



- Savage, L.J. (1961) "The foundations of probability reconsidered", Proc. of the Fourth Berkeley Symp. on Math. and Probability, 1: 575-586.
- Savinskii, I.D. (1965) Probability Tables for Locating Elliptical Underground Masses with a Rectangular Grid, Consultants Bureau, N.Y., 110 p.
- Scheidegger, A.E. (1960) The Physics of Flow Through Porous Media, Univ. of Toronto Press, 2nd ed., 313 p.
- Schlaifer, R. (1959) Probability and Statistics for Business Decisions, Wiley, N.Y., 732 p.
- Shanawany, M.R. (1936) "An illustration of the accuracy of the  $\chi^2$  approximation", Biometrika, 28: 179-187.
- Shannon, C. (1948) "The Mathematical theory of communication", Bell System Tech Jour., July and Oct.
- Siegel, S. (1961) "Decision making and learning under varying conditions of reinforcement", Ann. N.Y. Acad. Sci., 89: 766-783.
- \_\_\_\_\_ and D.A. Goldstein (1959) "Decision-making behavior in a two-choice uncertain outcome situation", Jour. Exp. Psych. 57:37-42.
- Skellum, J.G. (1952) "Studies in plant ecology. I. Spatial Pattern", Biometrika, 39: 346-362.
- Slichter, L.B. (1955) "Geophysics applied to prospecting for ores", Econ. Geol. 50th Anniv. Vol: 885-915.
- Snow, D. (1966) Discussion (theme 3) Intl. Cong. on Rock Mechanics, Lisbon, 3: 243-244.
- Sowers, G.B., and G.F. Sowers (1970) Introductory Soil Mechanics and Foundations, 3d ed. Macmillan, N.Y., 556 p.
- Sprowls, R.C. (1953) "Psychology-mathematical probability in relationships of lottery gambles", Amer. Jour. Psych., 66: 126-130.
- "Student" (1906) "On the error of counting with a haemocytometer", Biometrika, 5: 351-360.

- Suppes, P., & R.C. Atkinson (1960) Markov Learning Models for Multiperson Interactions, Stanford Univ. Press, 296 p.
- Switzer, P. (1967) "Reconstructing patterns from sample data", Ann. Math. Stat., 38: 138-154.
- Teller, P. (1968) "Goodman theory of projection", Br. Jour. Phil. of Sci., 20.
- Terzaghi, K. (1929) "Effect of minor geological details on the safety of dams", A.I.M.M.E. tech. publ. #215: 31-44.
- \_\_\_\_\_ & R.B. Peck (1947) Soil Mechanics in Engineering Practice, Wiley, N.Y., 727 p.
- Terzaghi, R. (1965) "Sources of error in joint surveys", Geotechnique, 15: 287-303.
- Tribus, M. (1962) "The maximum entropy estimate in reliability", in Recent Developments in Information and Decision Processes, R.E. Machol and P. Gray, eds., Macmillan, N.Y., 102-140.
- \_\_\_\_\_ (1969) Rational Descriptions, Decisions, and Design, Pergamon Press, 487 p.
- Turkstra, C.J. (1970) "Applications of Bayesian decision theory", in, N.C. Lind (ed.) Lectures on Structural Reliability and Codified Design, Univ. of Waterloo, Canada, pp. 49-71.
- Tversky, A. (1967) "Additivity, utility and subjective probability", Jour. Math. Psych., 4: 175-202.
- Vanmarke, E. (1970) Some basic results for 1-D random processes, Course notes for course 1.965, M.I.T.
- Vistelius, A.B. (1966) Structural Diagrams, Pergamon Press, 176 p.
- Van Neumann, J., and O. Morgenstern (1947) Theory of Games and Economic Behavior, 2nd ed., Princeton Univ. Press, Princeton, N.Y., 641 p.
- Wahlstrom, E.E. (1964) "The validity of geological projection: a case history", Econ. Geol., 59: 465-474.
- Wald, A. (1947) Sequential Analysis, Wiley, N.Y., 212.p.
- \_\_\_\_\_ (1950) Statistical Decision Functions, Wiley, N.Y., 179 p.

- Wallis, W.A., & H.V. Roberts (1956) Statistics: A New Approach, Free Press, N.Y., 418 p.
- Warner, J. (1969) "Fortran IV program for constructing pi diagrams with the Univac 1108 computer", Computer contribution #33, Kansas State Geol. Survey, Univ. of Kansas, Lawrence.
- Watanabe, S. (1969) Knowing and Guessing, Wiley, N.Y., 592 p.
- Watson, G.S. (1956) "Analysis of dispersion on sphere", Monthly Notices Roy. Astron. Soc., Geophys. Suppl. 7(4): 153-159.
- \_\_\_\_\_ (1966) "The statistics of orientation data", J. Geol., 74: 786-794.
- \_\_\_\_\_ (1971) "Trend-surface analysis", Math. Geol., 3: 215-226.
- West, Q.M. (1951) The results of applying a simple random sampling process to farm management data, Agricultural Experimentation Station, Cornell University.
- Whittle, P. (1963) "Stochastic processes in several dimensions", Bull. 34th Sess. of I.S.I.: 974-985.
- Wiener, N. (1949) Extrapolation, Interpolation and Smoothing of Stationary Time Series, M.I.T. Press, Cambridge, Mass., 163 p.
- Wilde, D.J. (1964) Optimum Seeking Methods, Prentice-Hall, Englewood Cliffs, N.J., 202 p.
- Wilson, E.B. (1952) An Introduction to Scientific Research, McGraw-Hill, N.Y., 373 p.
- Winkler, R.L. (1967a) "The assessment of prior distributions in Bayesian analysis", Jour. Amer. Statist. Assn., 62: 776-800.
- \_\_\_\_\_ (1967b) "The quantification of judgement: some methodological suggestions", Jour. Amer. Statist. Assn., 62: 1105-1120.
- Yates, F. (1948) "Systematic sampling", Phil. Trans. Roy. Soc. London, Ser. A, 241: 345-377.
- \_\_\_\_\_ (1960) Sampling Methods for Censuses and Surveys, 3d ed., Chas. Griffin and Co., London, 440 p.
- Zellner, A. (1971) An Introduction to Bayesian Inference in Econometrics, Wiley, N.Y., 431 p.



1

