R74-8 DSR 80628

RALPH M. PARSONS LABORATORY

for

WATER RESOURCES AND HYDRODYNAMICS

Department of Civil Engineering
Massachusetts Institute of Technology

THE METHODOLOGY OF BAYESIAN INFERENCE AND DECISION MAKING APPLIED TO EXTREME HYDROLOGIC EVENTS

by

Eric F. Wood

Ignacio Rodríguez-Iturbe

and

John C. Schaake, Jr.

January 1974

Report No.178

The work upon which this publication is based was supported in part by funds provided by the United States Department of the Interior as authorized under the Water Resources Research Act of 1964 as amended.

ABSTRACT

THE METHODOLOGY OF BAYESIAN INFERENCE AND DECISION MAKING APPLIED TO EXTREME HYDRO-LOGIC EVENTS

This study presents the methodology of Bayesian inference and decision making applied to extreme hydrologic events.

Inference procedures must consider both the natural or 'modelled' uncertainty of the hydrologic process and the statistical uncertainty due to a lack of information. Two types of statistical uncertainty were considered in this study. The first type is the uncertainty in modelling the hydrologic process, and the second type is the uncertainty in the values of the model parameters. The uncertainty is reduced by considering prior sources of information (regional regression, theoretical flood frequency analysis or subjective assessment) and historical flood data. A 'Bayesian distribution' of flood discharges is developed that fully accounts for parameter uncertainty. In an analogous manner, model uncertainty is analyzed, which leads to a 'composite Bayesian distribution'. The uncertainty in flood frequency curves from rainfall-runoff models is also analyzed, due to the uncertainty in the parameters of the models.

The Bayesian inference model is then applied to a Bayesian decision model, where the decision rule is the maximization of expected net monetary benefits. A case study of determining the optimal size of local flood protection for Woonsocket, Rhode Island, was considered, using realistic flood damage and cost functions.

The results indicate that Bayesian inference procedures can be used to fully account for statistical uncertainty and that Bayesian decision procedures provide a rational approach for making decisions under uncertainty.

ACKNOWLEDGEMENTS

This study is a part of a two-year project entitled "Estimation of Hydrologic and Engineering Parameters for Water Resource Management". The project was supported by the Office of Water Resources Research, Grant No. 14-31-0001-9021; this financial support is gratefully acknowledged. Administrative support was provided by M.I.T. through DSR 80628. Computer services were provided by the Information Processing Center and the Ralph M. Parsons Laboratory at M.I.T.

The work was performed by Eric F. Wood, Research Assistant in Civil Engineering, and, essentially as presented here, consists of his thesis presented in partial fulfillment of the requirements for the degree of Doctor of Science in Water Resources. The work was supervised by Prof. Ignacio Rodriguez-Iturbe.

The authors would like to express their gratitude to the many former and present faculty and students, whose discussions with the authors helped clarify many conceptual issues. In this regard, the discussions with Professors C. Allin Cornell and John C. Schaake; and Research Assistants Mr. Willie Vicens, Mr. Roberto Lenton, and Mr. Philip Liu have been most helpful.

A special note of thanks is due to Ms. Erika Babcock and Ms. Stephanie Demeris for their patient and skilful typing of the report.

TABLE OF CONTENTS

List of Figures

LIST OF TABLES

List of Symbols

$\underline{b}$	vector estimator for $\underline{\beta}$, in the regression equation, Equation 3.6
C_v	coefficient of variation
$E[\cdot]$	expectation operator
$f(x)$	pdf of the variable X
$f(\theta)$	pdf of parameter θ from sample observations
$f'(\theta)$	prior pdf for parameter θ
$f''(\theta)$	posterior pdf for parameter θ
$\tilde{f}(q)$	Bayesian pdf for q
$F_Q(q)$	cumulative density function for q
$F_Q(q)$	Bayesian CDF for Q
$G_Q(q)$	Exceedance probability j compliment of $F_Q(q)$, $= 1 - F_Q(q)$
h	precision (variance^{-1})
$\bar{i}$	average rainfall intensity in/hr
K_i	marginal density function of a set of observations (or marginal likelhihood) for model i
$L(\theta\|\underline{Q})$	likelihood function for parameter θ from the observation set $\underline{Q}$
m	sample mean (Chapter 4)
m'	prior mean
m"	posterior mean
n	number of sample observations (Chapter 4)
q or Q	discharge

q_m	peak discharge (Chapter 5)
Rq_m	region in $\bar{i}$, t_r space when q less than q_m
S^2	mean square error in regression (Eq. 3.7)
t_r	storm duration (hr)
T_e	recurrence interval in years
$u(d_i,q)$	utility for decision d_i and variable q
v	sample variance (Chapter 4)
v'	prior variance
v''	posterior variance
$V[\cdot]$	variance operator
Y_m	ratio of historical mean to Eagleson's mean (Eq. 3.26)
Y_v	ratio of historical variance to Eagleson's variance
$\underline{Y}$	vector of dependent variables for the linear regression (eq. 3.3)
$\underline{Z}$	vector of known physiographic characteristic for the linear regression (Eq. 3.3)
α	parameter of the exceedance model (Eq. 4.2)
$\underline{\beta}$	vector of regression parameters (Section 3.3)
β	parameter of rainfall parameter (Section 3.4 and Chapter 5)
σ^2	true variance of pdf
μ	true mean of pdf
$\underline{\varepsilon}$	vector of regression error residuals
ρ_{xy}	correlation between x and y

λ	parameter of rainfall pdf (Chapter 5) shadow premium (Chapter 6)
ν	degrees of freedom from a sample observation (Chapter 4) parameter of the exceedance model (Eq. 4.25)
ν'	prior number of degrees of freedom
ν''	posterior number of degrees of freedom
ϕ	water loss coefficient, in/nr
η	mean of the log-normal pdf for q.

CHAPTER 1

Introduction

Each year floods occur and result in large monetary damage, loss of life, and the disruption of community services. Since most historical records are short, the estimation of flood frequency curves is difficult. Such uncertainty can lead to poor flood design decisions.

While too often hydrologic analysis emphasizes hydrologic variables, the real decision variable is a general engineering variable, like the height of a dike or the size of a spillway. The hydrologic variables in flood estimation may be a T year flood or a flood frequency curve. These hydrologic variables should be used as an input into the decision process that leads to the 'best' engineering design. In the decision process, the hydrologic variables are related to design variables through a utility function which reflects the different economic and social implications of the project.

A decision process that is formulated to consider the occurrence of flood discharges as a stochastic process usually produces better designs. Due to limited data, the estimation of the extreme values for a stochastic process of flood discharges is extremely difficult, thus leading to significant statistical uncertainty. The statistical uncertainty is reflected in the uncertainty of the probability models as the correct ones to represent the process at hand; it is also reflected in the parameters of those models that are used to represent the true stochastic process. A decision process should be able to consider this statistical uncertainty, as well as the modelled uncertainty (the stochastic process), if rational designs are to be obtained.

Bayesian analysis, within the framework of statistical decision theory (Raiffa, 1968), prescribes a methodology for making decisions under the conditions of uncertainty and risk. Decision theory allows the decision maker to consider together the uncertainty of the modelled process, the quantifying of decision outcomes and the preferences towards these outcomes. Bayesian analysis is a probabilistic framework by which the uncertainty in any design variable and the knowledge about that variable can be considered jointly with the risk involved in the actual decision process.

Normally used procedures for flood design do not consider correctly the interaction among the hydrologic variables, the economic implications, and the preferences toward particular outcomes. The standard flood design procedure is to determine some project flood which leads to a design. No attempt to specify preferences towards particular outcomes is made, and the economic criterion for the project is that benefits exceed costs, at the specified design. The standard economic criterion for maximizing expected benefits is usually not considered.

The analysis of statistical uncertainty also is not analyzed correctly in current procedures of flood analysis. Usually, inferences on the occurrence of future floods are made by using a convenient probability model of flood discharges, where the unknown parameters are calculated from historical records. Such procedures for making inferences do not fully account for the uncertainty in the probability models of flood discharges or the uncertainty in the parameters of the models. Incorrect designs can result from an incomplete accounting of statistical uncertainty. Bayesian statistical procedures can account for all statistical uncertainty and moreover provide a procedure to pool together all the available information, be it historical, regional or subjective information.

1.1 Literature Review

Even though interest in Bayesian procedures is increasing, there are few reports of applying a full Bayesian analysis to hydrologic design.

A study by McGilchrist and Woodyer (1970) looks at the occurrence of floods greater than some discharge. A Bernoulli process is assumed, so that no distribution properties of the exceedance floods are specified. A beta pdf, for the parameter of the binomial pdf, is used. To estimate the parameters of the beta pdf, they use a combination of maximum likelihood point estimators and classical confidence interval procedures. Such procedures should not be passed off as Bayesian statistics. Since their model does not consider the distribution of flood discharges, the application to real decision problems would be limited.

The study by Shane and Gaver (1970) uses an exceedance model, similar to the exceedance model in this study. Their study finds the posterior pdf of the model parameters from the observed data and from a prior pdf. The prior is found from a regional regression. Using a linear cost function for flood protection, and using a constant damage, if a flood occurs, they find the optimal Bayes point estimator. This estimator is the one which minimizes expected damages plus the cost of protection. Point estimators will be shown to underestimate the uncertainty in the flood discharge pdf in Chapter 2 of this study. Shane and Gaver's procedure was not applied to a decision problem with real cost and damage functions, and thus the real issues in the decision process were not considered.

Andersen et al (1971) applied Bayesian statistical theory to an irrigation problem. The decision was to determine crop plantings, when future water supplies were unknown. The decision framework is extremely simplified. The future states of nature, water supplies, is represented by four discrete states; poor, fair, good and excellent water supplies. Information on the water supply is estimated by surveying snow pack levels. While the problem is interesting, the oversimplification of the future states of nature limits the procedures in more complex, real-world problems.

Davis et al (1972) considers the decision rule that minimizes risk. Conceptually, this decision rule can lead to difficulties, since risk implies that the 'correct' values for unknown parameters are known. In real-world decision problems, parameters are simply not known. Furthermore, the decision rule of maximizing expected utility seems to be a more appropriate decision rule. Davis et al do not use prior pdf for uncertain parameters, but only use observed data.

Tschannerl (1971) considers the optimal reservoir design using Bayesian procedures. His utility is a function of the design and uncertain parameters of the reservoir inflows. Such a utility function would be very difficult to find for real-world problems, where most utility functions depend upon the decisions and the uncertain discharge. To simplify the analysis, he assumes that the mean is the only unknown parameter. Finding an 'optimal' design on the basis of the uncertainty of the mean does not consider the complete pdf of the flows. Such a procedure can lead to incorrect decisions. His study is extremely interesting and also covers the effect of stream record length on opportunity losses inv ved in the design.

1.2 Summary of Proposed Methodology

This study advocates the use of Bayesian procedures for flood designs. The hydrologic variable, which is an input into the decision process, is the probability density function for q, the flood discharges. The true pdf for q is unknown, and is represented by some probability model. This model has parameters which are also unknown. The model uncertainty and the parameter uncertainty are statistical uncertainties, due to incomplete information.

The proposed methodology is to determine the probability density function for flood discharges, which completely accounts for all statistical uncertainty. To reduce the statistical uncertainty in the model parameters, historical data, as well as other sources of information, is analyzed. This analysis leads to a posterior pdf for the parameters. The Bayesian probability density function for flood discharges is obtained from the model, conditional upon the parameters, and the posterior pdf of the parameters. The Bayesian model for flood discharges fully accounts for parameter uncertainty. In an analogous manner to parameter uncertainty, model uncertainty is analyzed. The analysis leads to a 'composite Bayesian distribution' for flood discharges.

The resulting probability density function for the flood discharges is then used in the decision model to determine the optimal design, With the decision rule being the maximization of expected utility. The utility function is net monetary benefits, and it depends upon the design level and upon the occurrence of future flood discharges.

A case study for the Blackstone River, at Woonsocket, Rhode Island, demonstrates that the proposed methodology is practical in real-world problems. Both the procedures which account for statistical uncertainty and the decision process used to find optimal designs can be applied in a straightforward manner.

CHAPTER 2

Theoretical Considerations of Inference and Decision Procedures

2.1 Introduction.

Raiffa (1968) views the field of statistics as consisting of two main fields, that of decisions and that of inference.

The field of decisions is concerned with the solution of practical problems where a decision maker must decide upon some act among a set of alternatives. Inference is concerned with increasing the knowledge of how stochastic processes behave, separate to how they may affect decisions. This thesis is concerned with both inferences about extreme streamflow events (floods) and decisions concerning alternative designs to minimize the economic (or social) effects of such events.

The approach taken is that of Bayesian decision analysis which combines Bayesian subjective inference procedures with the decision framework of the Savage School (Savage, 1950)which employs subjective utilities for consequences and probabilities for states of nature.

This chapter discusses this framework and compares Bayesian inference procedures with non-Bayesian (classical, standard or orthodox) procedures and Bayesian decision making (Savage School) with the other principle schools of Fischer,Neyman-Pearson, and Wald (Raiffa, 1968). Only the Bayesian approach indicates to decision makers which act is the best because the methodology considers realistic loss functions for the decision variables and can consider, explicitly, all sources of uncertainty. Section 2.2.1 looks in detail at the Bayesian inference problem from two approaches. The first approach is finding 'optimal' Bayes' estimators for the uncertain parameters while the second approach finds the Bayesian distribution of the random variable of the decision process (in our case, flood events). The second approach considers and accounts for the whole probability distribution of the uncertain parameters. This is done in a manner analogous to compound distribution theory.

2.2 Inference Procedures.

This section discusses the main categories of inference procedures. Essentially the non-Bayesian or classical approach will be compared to Bayesian inference approach. The classical statistical school can further be broken down into those that support the Likelihood Principle - 'Likelihoodists', and those that do not, - 'non-likelihoodists'. There are several formulations that lead to the development of the likelihood principle of which the formulation by Birnbaum (1962), developed within the classical inference framework, is probably the most elegant. The principle states that observations which lead to the same likelihood function should lead to the same conclusions without specifying the relation between function and conclusions. Most classical procedures do not satisfy the likelihood principle. In fact, all procedures that necessitate reference to the random variable Q instead of the observation q violate this principle (Lindley, 1972). This includes such procedures as unbiased estimators and significance tests.

Bayesian methods satisfy the likelihood principle with the likelihood function representing the observations in updating probability density functions (pdf) of uncertain parameters, vis-a-vis Bayes' Rule.

Consider the problem of the hydrologist who must make a decision between a number of alternate designs that propose to prevent or decrease the occurrence of future floods. His first task is to make inferences about the underlying process that generates these events but, in addressing this problem, he is faced with a number of sources of uncertainty. These sources of uncertainty may be summarized into three categories (Benjamin and Cornell, 1970).

1. Natural uncertainty. This is the uncertainty in the stochastic process - the occurrence of streamflows (especially extreme streamflows for our problem)
2. Statistical Uncertainty. This is associated with the estimation of the parameters of the model of the stochastic process due to limited data.
3. Model Uncertainty. his is associated with the uncertainty that the probabilistic model of the stochastic process may not be the true model. Most hydrologic processes are so complex that no model yet devised may be the true model or that no hydrologic events follow one particular model.

What information does the hydrologist have to help inferences in the light of the above difficulties? Often he will have information on the historical occurrence of flood events. The length of such records in the United States average 40-50 years and in many cases much less. There also exists in hydrology other techniques for estimating flood events. These techniques include regional flood-regression equations, analytical and empirical formulations.

How the information that the hydrologist has, is applied in making inferences, depends upon the school of statistics to which he precribes. Let us first discuss the Bayesian approach. These discussions will assume that there is no uncertainty in the probabilistic model of the stochastic process since a comparison between Bayesian and non-Bayesian statistical inference can be handled by looking at the uncertain parameters of the model. The uncertainties involved when choosing a model to represent the hydrologic process are dealt with in Chapter 7.

2.2.1 Bayesian Inference.

Bayesian inference lays its foundations upon the idea that states of nature can be and should be treated as random variables. Thus the mean annual flood, μ_q, is a random variable distributed with a mean m_{Q_A} and variance $S^2_{Q_A}$, estimated from the data. Considering streamflows as random variables, instead of deterministic variables, leads to better inferences and designs. The extension of the argument, by the Bayesian School, is that it is useful and professionally sound to treat all uncertain states of nature as random variables, whether these are the parameters of the models of streamflows, the capacity of a flood channel, or the area flooded by a particular sized flood.

Consider the problem set forth earlier about the hydrologist making inferences on flood events. Before making use of the data collected at the site, he will wish to set forth his 'information' concerning the uncertain parameter set of the model $\underline{\Theta}$. This information will be described by a prior pdf on $\underline{\Theta}$, $f'(\underline{\Theta}|I_o)$ - prior to applying the data. The basis upon which these priors are obtained will be discussed in detail in Chapter 3; for now it will be enough to say that they are based upon initial information I_o obtained from some source - maybe the regression relationship, theoretical studies, or eingineering judgement.

The hydrologist now has a set of observations $\underline{q}$ of annual floods which he assumes comes from the model pdf $f_Q(q|\underline{\Theta})$ which is conditional upon the parameter set $\underline{\Theta}$.

Bayes' Rule provides a procedure by which the prior distribution of the parameter set $\underline{\Theta}$ may be updated by the data to provide the posterior distribution of $\underline{\Theta}$. The proof is found in most probability tests (for example see Benjamin and Cornell, 1970, pages 64-65).

$$f''(\underline{\Theta}|\underline{q}, I_o) = f(\underline{q}|\underline{\Theta}) \cdot f'(\underline{\Theta}|I_o) / f(\underline{q}) \tag{2.1}$$

where

$f''(\underline{\Theta}|\underline{q},I_o)$ is the posterior pdf conditional upon a set of data $\underline{q}$ and initial information I_o.

$f(\underline{q}|\underline{\Theta}) \equiv L(\underline{q}|\underline{\Theta})$ is the sample likelihood function of the observations conditional upon the parameter set.

$f'(\underline{\Theta}|I_o)$ is the prior pdf conditional upon initial information.

$f(\underline{q})$ is a normalizing constant.

The posterior pdf of $\underline{\Theta}$ is therefore a function weighted by a prior pdf of $\underline{\Theta}$ and a data likelihood function in such a manner as to combine the information content of both. The effect of various priors and likelihoods upon the posterior is shown in Figure 2.1. If future observations $\underline{q}_F$ are available, Bayes' Rule can be used to update the probability density function on $\underline{\Theta}$. In this case, the former posterior pdf on $\underline{\Theta}$ now is the prior pdf since it is prior to the observation or utilization of the new data. Applying Bayes' Rule then yields the new posterior distribution $f''(\underline{\Theta}|\underline{q}_F, \underline{q}, I_o)$.

$$f''(\underline{\Theta}|\underline{q}_F,\underline{q},I_o) \propto L(\underline{q}_F|\Theta) \cdot f'(\Theta|\underline{q}, I_o) \qquad (2.2)$$

The new posterior pdf would have been obtained if the two samples $\underline{q}$ and $\underline{q}_F$ had been observed sequentially as one set of data. This is easily shown by looking at the prior pdf of $\underline{\Theta}$ conditional upon $\underline{q}$ and I_o.

$$f'(\underline{\Theta}|\underline{q}, I_o) \propto L(\underline{q}|\underline{\Theta})\ f'(\underline{\Theta}|I_o) \qquad (2.3)$$

Substituting into (2.2) yields

$$f''(\underline{\Theta}|\underline{q}_F,\underline{q}, I_o) \propto L(\underline{q}_F|\underline{\Theta}) \cdot L(\underline{q}|\underline{\Theta}) \cdot f'(\underline{\Theta}|I_o) \qquad (2.4)$$

The likelihood functions can be combined to give

$$f''(\underline{\Theta}|\underline{q}_F, \underline{q}, I_o) \propto L(\underline{q}_F,\underline{q}|\underline{\Theta}) \cdot f'(\underline{\Theta}|I_o) \qquad (2.5)$$

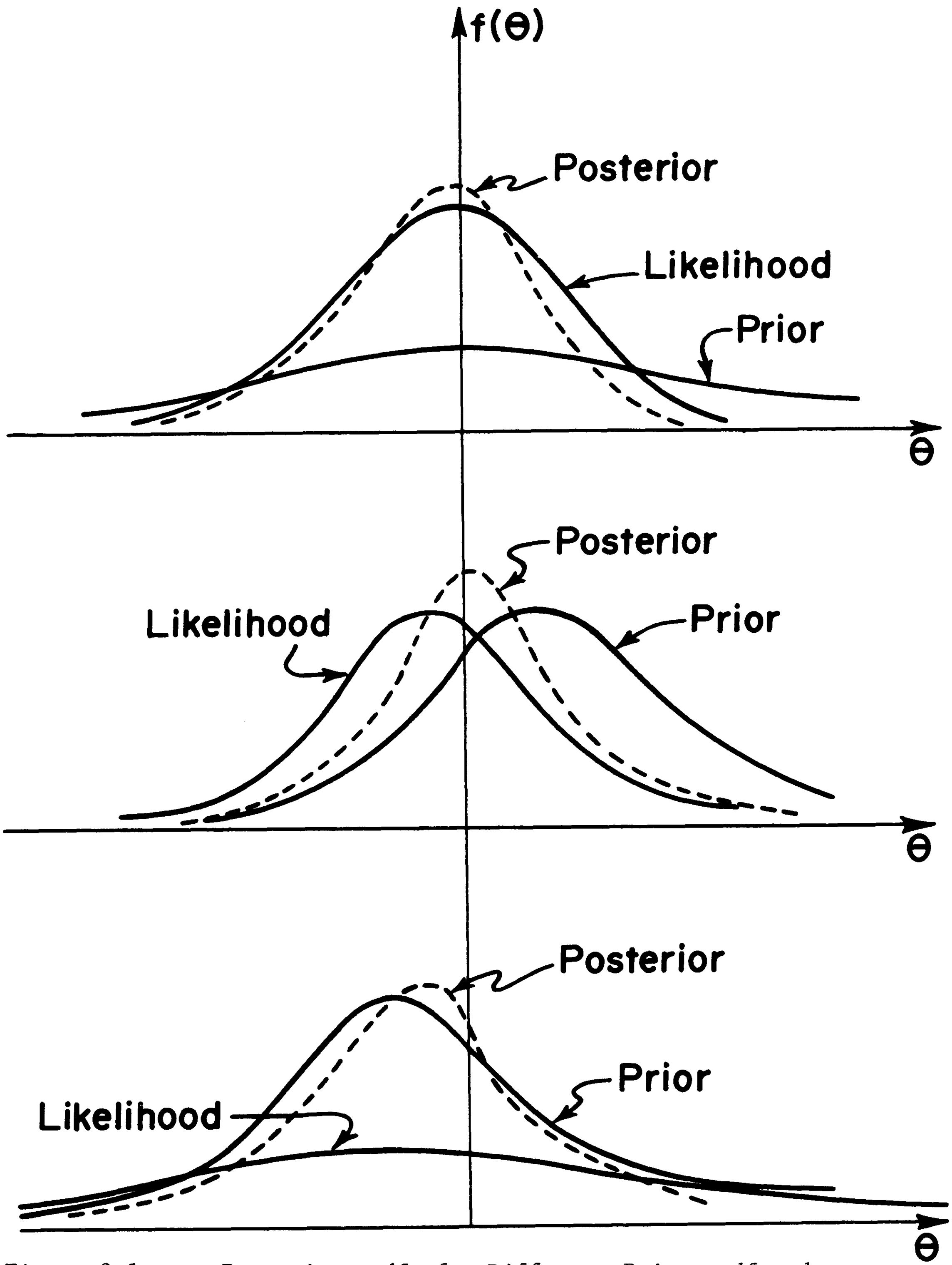


Figure 2.1: Posterior pdf for Different Priors pdf and Sample Likelihood Functions

which is the same result for the posterior of $\underline{\Theta}$ as if a sample $(\underline{q}_F, \underline{q})$ had been observed and initial information prior $f'(\underline{\Theta}|I_o)$ used.

The hydrologist wants to make inferences about the occurence of flood events assuming that such events Q can be modelled by a pdf $f_Q(q|\underline{\Theta})$ where the parameter set $\underline{\Theta}$ is unknown with certainty. His information has yielded a distribution for $\underline{\Theta}$ which will be represented by $f(\underline{\Theta})$.

The Bayesian analysis may proceed along one of two approaches. One approach is to find the 'best' point estimate of the parameter set denoted by $\underline{\Theta}^*$ and make inferences about flood events by using the model $f_Q(q|\underline{\Theta}^*)$. The other approach is to develop a model of flood events that is 'free' from any uncertain parameters. This model can be found by applying compound probability distribution theory. The unconditional distribution of Θ, $\tilde{f}_Q(\underline{q})$ will be found from

$$\tilde{f}_Q(q) = \int_{\underline{\Theta}} f_Q(q|\underline{\Theta})\, f_\Theta(\Theta)\, d\Theta \tag{2.6}$$

This distribution will be called the Bayesian pdf $\tilde{f}_Q(q)$ to distinguish it from the modeled pdf $f_Q(q|\underline{\Theta})$. Both approaches will be analyzed in detail.

2.2.1-a Bayes Point Estimation.

As stated earlier, Bayesian analysis looks upon the true parameter set $\underline{\Theta}$ as random variables with a distribution $f(\underline{\Theta})$ that is centered upon the expected value of the parameter distribution. If the estimator $\underline{\Theta}'$ is chosen when $\underline{\Theta}$ is the true value then a loss, $\ell(\underline{\Theta}'|\underline{\Theta})$ is encountered. Since the true estimator is a random variable the loss is also a random variable. The expected loss can be written as

$$E[\text{loss}|\underline{\Theta}'] = \int_{\underline{\Theta}} \ell(\underline{\Theta}'|\underline{\Theta})\, f_{\Theta}(\underline{\Theta})\, d\underline{\Theta} \qquad (2.7)$$

The optimal Bayes estimator $\underline{\Theta}^*$ is that estimator which minimizes the expected loss over all $\underline{\Theta}$. $f(\underline{\Theta})$ is the pdf of the parameter set $\underline{\Theta}$ and may be prior or posterior to data sampling. If it is prior, then inferences on q will be made using initial information; if $f(\underline{\Theta})$ is solely data based, then it will be the rescaled likelihood function.

Depending upon the form of the loss function $\ell(\underline{\Theta}'|\underline{\Theta})$, different Bayes estimators will minimize (2.7). Table 2.1 gives three commonly used loss functions and their corresponding Bayes estimator.

2.2.1-b Bayesian Distribution

The second approach in analyzing uncertain parameters is the application of compound distributions in a Bayesian framework. The definition of a compound distribution is: (see Benjamin and Cornell, 1970)

$$\tilde{f}_Q(q) = \int_{\underline{\Theta}} f_Q(q;\underline{\Theta})\, f_{\Theta}(\Theta)\, d\Theta \qquad (2.8)$$

where $f_Q(q;\Theta)$ is the pdf of Q which depends upon the random variables Θ.

Table 2.1

Optimal Bayes Estimator	Loss Function
$\Theta^* =$ mean of $f(\Theta)$	quadratic; $(\Theta^* - \Theta)^2$
$\Theta^* =$ median of $f(\Theta)$	linear; $\lvert\Theta^* - \Theta\rvert$
$\Theta^* =$ maximum likelihood of $f(\Theta)$	constant C if $\Theta \neq \Theta^*$ 0 if $\Theta = \Theta^*$

Table 2.1 Optimal Bayes Estimator for Various Loss Functions.

$f_\Theta(\Theta)$ is the pdf of the random variable Θ.

$\tilde{f}_Q(q)$ is the pdf of the random variable Q which does not depend upon any parameters.

The distribution $\tilde{f}_Q(q)$ can be interpreted as being the distribution of $(q;\Theta)$ weighted by the distribution of Θ. In Bayesian analysis, uncertain parameters are treated as random variables and the resulting distribution $\tilde{f}_Q(q)$ is the Bayesian distribution of Q. As the case with Bayes estimators, inference statements about Q can be made either prior to data observation $(f(\Theta) = f'(\Theta|I_o)$ or posterior to data observation $(f(\Theta) = f''(\Theta|\underline{q}, I_o))$. Inferences made by combining new information is achieved by updating the distributions of uncertain parametersthrough Bayes Rule, then by calculating the updated Bayesian distribution through the application of Equation (2.8). It is incorrect to try to update $\tilde{f}_Q(q)$ directly.

For the hydrologist it may not be clear which of the two procedures best represents the uncertainty of parameters. In fact, only the Bayesian pdf of q, $\tilde{f}_Q(q)$, and not the modeled pdf, $f_Q(q|\Theta^*)$ correctly accounts for the uncertainty in Θ. This is analyzed fully in Section 2.3.

2.2.2 Classical Inference.

If the hydrologist is a classical or non-Bayesian statistician, this section will discuss how he will make inferences for the problem discussed earlier from a Bayesian viewpoint.

First he will not use any other information except the information contained within the data sample to estimate uncertain parameters. The classical statistician does not view the uncertain parameter as a random variable but instead views the estimate of the parameter as a random variable. The classical procedure is to find a point estimate for the parameter and test as to whether the parameter falls within a confidence interval of acceptance at some level of significance.

Consider the case of a hydrologist with a data sample of length n from a shifted exponential of the form

$$f_{Q'}(q'|\lambda) = \lambda e^{\lambda q'} \qquad (2.9)$$

where

$$q' = \begin{cases} q - q_o & \text{for } q \geq q_o \\ 0 & q < q_o \end{cases}$$

An interval estimate for the unknown parameter λ may be established by noting that an estimate of $\hat{\lambda}$ can be obtained from $1/\bar{q}'$ where $\bar{q}'$ is the mean of $f_{Q'}(q'|\lambda)$. λ will be gamma distributed. To test if $\lambda \leq \lambda^*$ at a confidence level $1 - \alpha$, one may employ the distribution of $\bar{q}'$ in the following manner:

$$P[\hat{\lambda} \leq a] = P[\bar{q}' \geq \frac{1}{a}] = 1 - \alpha \qquad (2.10)$$

For a given α and n, tables of the gamma distribution will give $1/a$ as a function of λ or, for given $1/a$ and n , the confidence level α may be established. Rearrangement will isolate in the probability statement permitting the substitution of the observed value of $\hat{\lambda}$ and yielding the exact confidence interval on λ.

The confidence interval will therefore cover the true value $(1 - \alpha)$ 100 percent of the time.

Fundamental problems with classical confidence intervals have been reported which can not be resolved except by applying a Bayesian interpretation to confidence intervals. These problems may be summarized as following:

1) A test for the region $\lambda \leq k$ $(P[\lambda \leq k] = 1 - \alpha)$ could, as discussed above, provide a bound for λ which is dependent upont the data, i.e. $k - k(q')$, or, fixing the bound on λ, develop a level of confidence that depends upon the data, i.e. $1 - \alpha = 1 - \alpha(q')$. For the classical statistician this should be disturbing since the confidence levels are not measures of any fundamental kind of confidence but are data dependent. (Pratt, 1965). This can be resolved in a Bayesian framework since the parameter (and not the estimate) is the random variable and such confidence interval probabilities are interpreted as being estimates of the posterior distribution that is conditional upon the estimate.

$$P[\lambda \leq k \mid \hat{\lambda}] = 1 - \alpha$$

2. Another disturbing result arising in the classical procedures is when a most powerful test on a parameter may be significant at one level but not at a less extreme level (Stein, 1951). Examples have been presented in the literature which exhibit this property, but, when the test is interpreted within a Bayesian framework, this undesirable property will not occur. (Chambers, 1970).

Discussion so far has centered upon confidence intervals. The classical statistician is also interested in the best point estimate, and procedures used to obtain this estimate include the method of moments and the maximum likelihood criterion. Both methods are widely discussed in statistics textbooks. The later procedure chooses that value of the parameter, as a best estimator, by the simple rule that states: the best estimate of the true parameter is the one with the highest relative likelihood of being observed, given the sample. The likelihood function is defined as

$$L\,(\underline{\Theta}\,|q_1 \ldots q_n) = \prod^{n} f_Q\,(q_i|\,\underline{\Theta}) \tag{2.11}$$

and the estimator $\underline{\Theta}^*$ is when

$$\frac{\partial}{\partial \underline{\Theta}} L(\underline{\Theta}^*\,|q_1 \ldots q_n) = 0 \tag{2.12}$$

The likelihood function gives an approximate distribution of the parameter, given the sample, from which confidence intervals can be established. Such confidence interval testing suffers from the problems discussed earlier and,as for point estimation, the next section will show that such procedures significantly underestimate the uncertainty in the flood events when uncertainty in parameters is present.

2.2.3 Comparison of the Bayesian Distribution with Bayesian Parameter Estimation.

This section compares the two procedures discussed under Section

2.2.1 to determine if optimal Bayes point estimators or the Bayesian pdf better represent the uncertainty towards unknown parameters and therefore will yield better inferences on the decision variable of interest - namely flood events.

Consider the process of annual flood events, Q, which has a distribution $f_Q(q|\mu_Q,\sigma^2{}_Q)$ which is conditional upon the mean μ_Q and variance σ_Q^2. It will be assumed for simplicity that the distribution of flood events has only one parameter, Θ , which is unknown with certainty and that the hydrologists knowledge about Θ is represented by a pdf $f_\Theta(\Theta)$ with mean μ_Θ and variance σ_Θ^2 .

The decision role for the Bayesian parameter estimation is from Equation (2.7) where the optimal Bayes' estimate Θ^* will be

$$\Theta^* = \min_{*\Theta} E[\text{loss}|\Theta'] = \int_\Theta \ell(\Theta'|\Theta')\, f_\Theta(\Theta)\, d\Theta \tag{2.13}$$

and the distribution of Q in the analysis will be $f_Q(q|\ \Theta^*)$. The mean and variance of $f_Q(q|\Theta^*)$ will be written as $\mu_{Q|\Theta^*}$ and $\sigma_Q{}^2{}_{|\Theta^*}$ respectively. If it is assumed that a quadratic loss function on the parameters is used, the choice of Θ^* will be the mean of $f_\Theta(\Theta)$ and will be independent of $\sigma_\Theta{}^2$. This procedure does not employ the full knowledge of Θ represented in $f_\Theta(\Theta)$, but only employs the first moment. Intuitively it would seem the point estimation procedure will underestimate the variance in Q, and inferences about Q will not reflect the true uncertainty that exists. This is in fact what happens,as will be shown below.

The Bayesian distribution of Q, $\tilde{f}_Q(q)$ follows from Equation

(2.6) or (2.8) and is

$$\tilde{f}_Q(q) = \int_{\Theta} f_Q(q|\Theta)\ f(\Theta)\ d\Theta \qquad (2.14)$$

The moments of the Bayesian pdf are easily found to be :

$$\tilde{\mu}_Q = \int_{\Theta}\int_{q} q\ f_Q(q|\Theta)\ f\ (\Theta)\ d\Theta\ dq = \int_{\Theta} \mu_Q\ f(\Theta)\ d\Theta \qquad (2.15)$$

Thus the mean of $\tilde{f}_Q(q)$ is the mean of the modelled distribution of Q weighted by the distribution of Θ . This leads to $\tilde{\mu}_Q = \mu_{Q|\mu\Theta}$. If a quadratic loss function is used, as in the Bayesian parameter estimation which leads to $\Theta^* = \mu_\Theta$, then $\tilde{\mu}_Q = \mu_{Q|\Theta^*}$. The mean of the Bayesian distribution is equal to the mean of the model distribution using the optimal Bayes' estimator.

The variance of the Bayesian distribution is by definition

$$\tilde{\sigma}_Q^{\ 2} = E[Q^2] - \tilde{\mu}_Q^{\ 2} \qquad (2.16)$$

Using Equation (2.15) this leads to

$$\tilde{\sigma}_Q^{\ 2} = \int_{\Theta} \sigma_Q^{\ 2} f(\Theta)\ d\Theta + \int_{\Theta} (\mu_Q - \tilde{\mu}_Q)^2\ f\ (\Theta)\ d\Theta \qquad (2.17)$$

or

$$\tilde{V}[Q] = E_\Theta[V[Q|\Theta]\ ] + V_\Theta[E[Q|\Theta]\ ] \qquad (2.18)$$

where $V[\cdot]$ is the variance operator and $E[\cdot]$ is expectation. Thus the variance of the Bayesian distribution of Q is the sum of two parts. The first is the expected variance of Q given Θ and the second is the variance (over Θ) of the expectation of Q given Θ.

The first term of the variance is the variance of the model distribution using the mean of $f_\Theta(\Theta)$ as in the Bayes estimator. It is this part of the total variance for which a point estimator procedure accounts. The second term of (2.18) is not accounted for in the Bayes estimator procedure, and it is by this quantity that the preceding procedure underestimates the variance of Q (given the uncertainty in Θ) and will lead to misleading inference statements.

The effect of this can be shown in a simple example displayed in Figure 2.2. In this example, the unknown parameter Θ is the mean, the variance is known. The unknown parameter Θ may take on a value of either Θ_1 or Θ_2 with equal probability, and the optimal Bayes estimator is assumed to be the mean of $P_\Theta(\Theta)$, i.e. $\Theta^* = (\Theta_1 + \Theta_2)/2$.

The error introduced by not adequately taking the uncertainty of Θ into account is significant, especially for extreme events.

In most flood design, extreme events play an important role. Inferences drawn from the Bayes estimator procedures could lead to serious design mistakes.

This analysis can be extended to all classical procedures that employ point estimators. There is no classical procedure to alleviate the full variance accounting; the only valid approach seems to be the Bayesian pdf.

2.2.4 Conclusions to Inference Procedures.

The past few sections have looked at inference procedures from

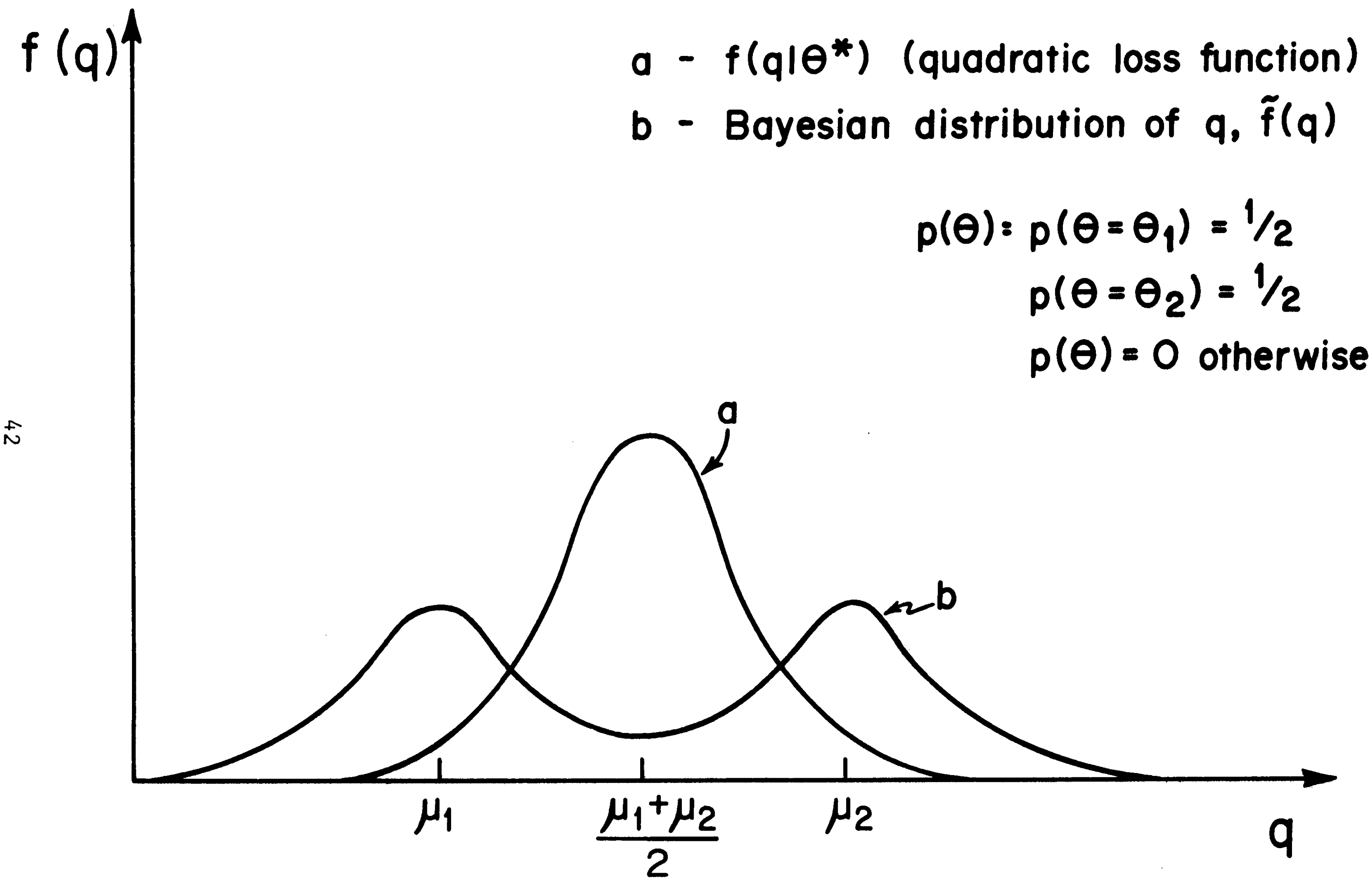


Figure 2.2: A Comparison of a Bayesian pdf with a pdf Employing an Optimal Bayes Estimator

both the classical and Bayesian viewpoints. The following things were found:

1. Classical procedures usually violated the likelihood principle which stated, in essence, that sets of observations leading to the same likelihood function should lead to the same inference, decisions, etc.

2. Classical confidence intervals lead to acceptance regions or levels of significance that are data dependent; also such confidence tests often given acceptance regions that are rejected at higher significance ($1-\alpha$) levels.

3. Under Bayesian procedures, significance levels are interpreted as estimates of the posterior distribution and such an approach clears up the difficulties of conclusion 2.

4. Of the Bayesian procedures, only the Bayesian pdf fully and completely estimates the variance of the process due to modelled uncertainty and parameter uncertainty.

5. Conclusion 4 can and should be extended to any procedure that relies upon point estimation,whether Bayesian or non-Bayesian.

2.3 Decision Analysis

The second field of statistics defined by Raiffa (1968) is that of decisions. In many statistical problems, the two fields overlap the two categories. This is especially true in the analysis and design of flood events. This Section will discuss the major schools of decision methodology identified by Raiffa (1968). These major schools are Fisher, Neyman-Pearson, Wald and Savage.

Before each approach is discussed, the decision problem, as perceived by the hydrologist, is as follows: a hydrologist must decide whether a flood protection design d_1 should be implemented or d_2. Suppose the design is a channel improvement to protect against the 100 year event. Design d_1 proposes to use a design discharge Q_C even though the true value of the 100 year event for design d_1 is unknown with certainty. The information the hydrologist has includes costs and damage curves as well as the information available to him under the inference procedures presented in Section 2.2 - namely historical data and analytical/empirical flood estimating techniques.

What information the hydrologist uses depends upon what decision school he follows. The discussion will divide the schools into non-Bayesian and Bayesian (Savage) schools.

2.3.1 Non-Bayesian Decision Analysis

This Section will look at the simple decision problem from the Fisherian, Neyman-Pearson, and Wald decision perspective.

The Fisherian school perceives the role of choosing between d_1 or d_2 as a hypothesis test, not as a decision act. The analysis does not employ losses associated with various errors and does not assign probabilities to possible states of nature. It is only concerned with rejecting a null hypothesis if the null hypothesis is true and feels additional sampling should be done until the null hypothesis can be accepted or rejected beyond a reasonable level of doubt. In flood design this additional sampling is not feasible since the events are rare and other sources of information are not accepted by the Fisherian School.

The Neyman-Pearson School frames the decision problem by a hypothesis test where acceptance or rejection of the hypothesis is an action problem. The evaluation between actions is determined by setting the level of type I error, α, and choosing that strategy which has the lowest type II errors, β. This comparison is made by plotting β against the true state of nature. This is the operating characteristic of the hypothesis rule (Benjamin and Cornell, 1970). The power function of the strategies is a plot of $(1-\beta)$ versus the state of nature. An operating characteristic for our problems may be represented by Figure 2.3. The problems with this approach are

1) Better error control only comes through more experimentation which usually is not feasible in flood design.
2) The best strategy (experimentation and action) depends upon the whole operating characteristic which depends upon relative seriousness of each type of error and upon "guesses" as to where the true value of Q_C lies, thus reducing the 'decision' to one of pure judgement.

While hypothesis testing on strategies is often used by hydrologists, the cogent information for decision making is the magnitude of the

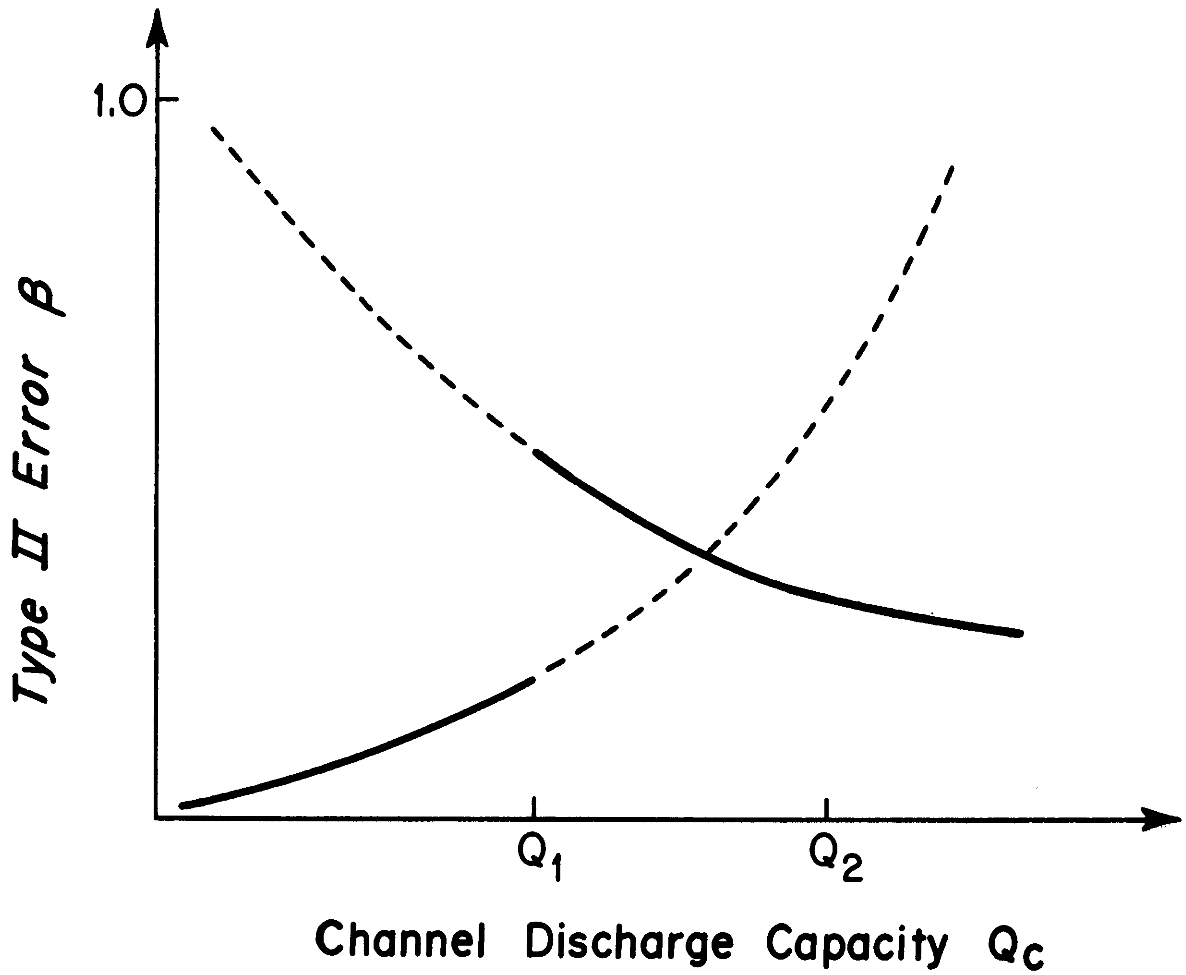


Figure 2.3: Neyman-Pearson Operating Characteristic for Decision Problems

state of nature (or the probability of its magnitude), not its level of significance. This can be shown by looking at Figure 2.4 which shows the expected damages and costs versus the true 100 year discharge for each of two decision strategies.

Classical procedures would test the null hypothesis $P[Q_C \geq Q_1] = 1 - \alpha$. If α is set quite low to reflect the high cost of decision d_2, then the error of rejecting the null hypothesis, when $Q \geq Q_1$, is the more serious of the two types of error.

A decision procedure which would use an expected damage and costs criterion would test the mean of $f_Q(q)$ to see if it is larger than Q_b, the break even point between the economies of decisions d_1 and d_2. This is the Bayesian approach with linear utility functions. The equivalent classical procedure is to test if $Q \geq Q_b$ at $\alpha = .50$ and use strategy d_1 if rejected. The former classical procedure puts too much weight in the ad hoc procedure of setting α very small.

Neither classical procedure (if the test $P[Q>Q_b] = .50$ can even be considered from the classical viewpoint) will lead to the best design since either d_1 or d_2 will be chosen. Many different designs will fall into the 'accept' category and hypothesis tests will not indicate the best design.

Wald (1950) extended the concepts of Neyman and Pearson by considering concepts as cost, loss, value, and worth of consequences in his formulation of the statistical decision problem. Wald did not assign probabilities to states of nature but found the value of a

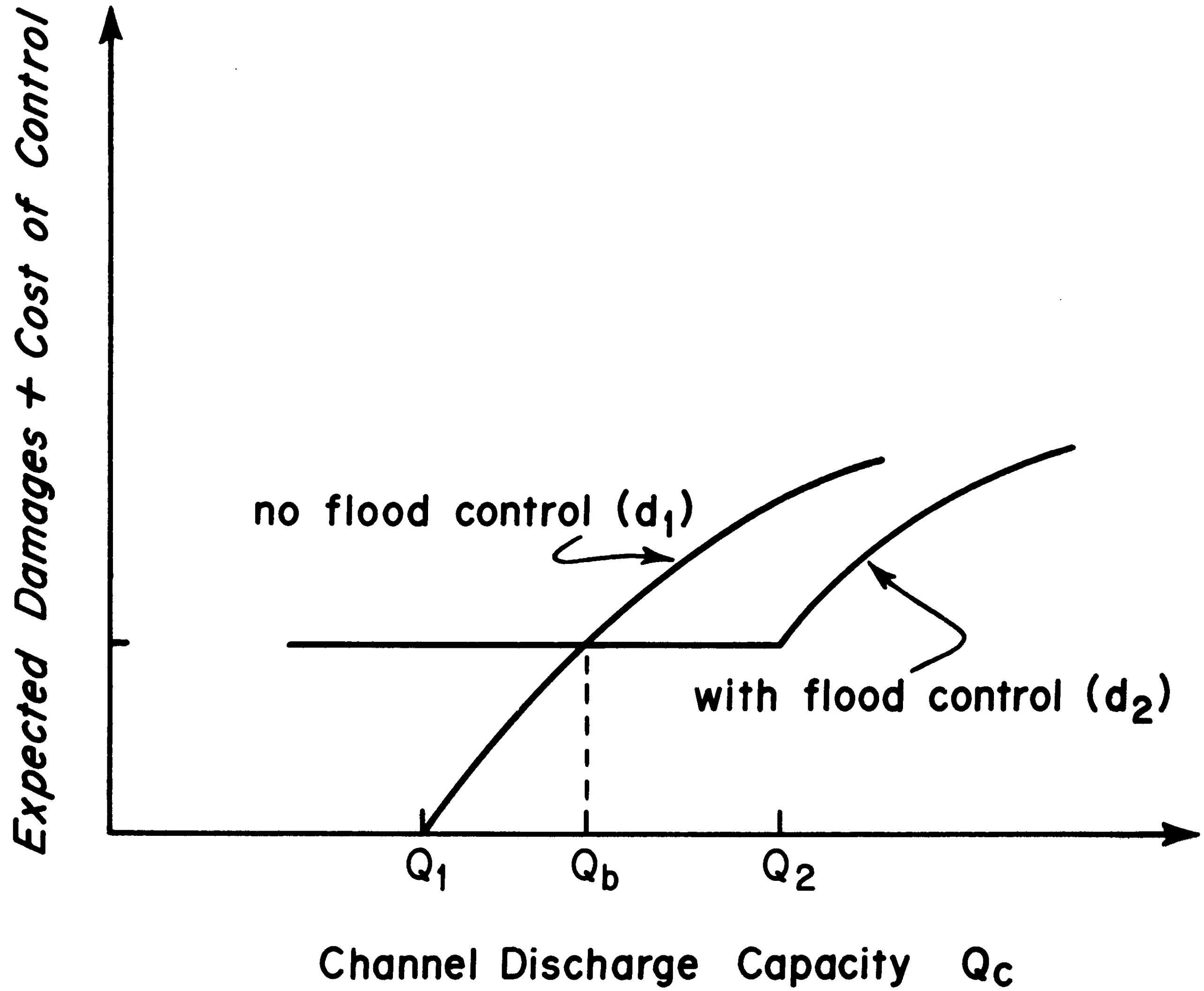


Figure 2.4: Expected Damages plus Costs for the 100 Year Flood Versus Flood Density Capacity for a Hypothetical Example

strategy conditional upon a state of nature being true, $v(d_i|Q)$. Wald would find those strategies that were not dominated by other strategies (a strategy at least as good or better for all conditional states of nature). From the set of efficient strategies, the criterion on choosing the best strategy was not addressed. Many criteria have been proposed and a popular approach is that of minimax. The minimax rule states that the best strategy is the one that minimizes the maximum loss. This rule has been shown to have the following deficiencies:

1) Minimax is equivalent to $\alpha = \beta$ which has been shown to give an incoherent procedure on the basis of the substitutability principle between simple lotteries (Lindley, 1971).
2) The violation of the principle of irrelevant alternative [if $\ell_1 > \ell_2$ of (ℓ_1, ℓ_2), then ℓ_2 can not be the best of (ℓ_1, ℓ_2, ℓ_3) when ℓ_3 is irrelevant to the choice between ℓ_1, ℓ_2], has been demonstrated by Savage (1954).
3) The optimal minimax strategy may be based upon a magnitude of the state of nature which has almost zero probability of occurrence.

Despite these demonstrations as to the unsoundness of minimax criterion, it is still used. Other procedures similar to the minimax criterion suffer from similar arguments. The problems stated above would not occur under an expected utility criterion.

2.3.2 Bayesian Decision Analysis.

The Bayesian decision analysis extends Wald's results by introducing both subjective utilities and probabilities for states of nature. Savage (1954) developed the theory based upon subjective probabilities and the theory of utility of von Neumann and Morgenstern (1947), even though the concept of subjective probabilities and utility in decision making goes back to Ramsey (1926).

The decision problem will be considered in detail since from our previous discussion it has been shown to provide a framework that does not lead to incoherent results and does not rely upon arbitrary procedures.

The consequence of any decision depends upon the decision taken and upon the subsequent outcome of the future state of nature, which is unknown. Thus, for each decision strategy a_i, from a set $\underline{A}$, and for each possible outcome event λ, from a set Λ, the decision maker can define a utility or value function, $u(a_i, \lambda_j)$. When the decision action a_i is taken, the outcome event or state of nature λ_j is unknown; therefore, the utility function of the action-event set $u(a_i, \lambda_j)$ is unknown and, thus, a random variable. The utility expected of any action a_i is

$$E[u|a_i] = \int_{\Lambda} u(a_i, \lambda_i) \; f_{\Lambda}(\lambda) \; d\lambda \qquad (2.19)$$

The decision rule is to choose that act a^* that maximizes the expected utility, that is

$$E[u|a^*] = \max_{\text{all } a \varepsilon \underline{A}} E[u|a] \qquad (2.20)$$

The decision strategy a_i may include (for the flood control problem) the building of dikes to some elevation, building a reservoir of some particular volume, a flood channel of a particular size, the restriction of land use in the flood plain, or any combination of these actions; or to do nothing. The outcome event or state of nature λ_j can be a single or multidimensional array depending upon the complexity of the analysis. It may include the discharge in the river, the height that the dike fails, the area flooded, the damage caused or any outcome which is uncertain and can be considered a random variable.

The field of utility theory has received a considerable amount of attention but much more research needs to be done-especially in the area of the assessment of decision maker's utility preferences. Recent work in the area of utility theory include Fishburn (1964, 1969), Pratt, Raiffa and Schlaifer (1965), Keeney (1969), Kirkwood (1972), Keeney and Raiffa (1971), and others.

This thesis will not investigate in detail the utility aspect of the decision problem but instead will employ an expected cost criterion (linear utility function in costs) with the decision rule being: choose that strategy a^* which minimizes the expected damages plus the cost of protection. Thus the main emphasis will be upon the inferences of the states of nature $f_\Lambda(\underline{\lambda})$.

For most of this work, the state of nature λ will be the peak flood discharge, Q. If one is concerned with the problem of flooding due to levee failure from seepage as well as overtopping them $\underline{\lambda} = (Q,R)$ where R is a measure of levee reliability and $f_\Lambda(\lambda)$

becomes a joint PDF $f_{Q,R}(q,r)$. The utility function $u(a_i;q_j,r_k)$ is assigned for all strategies and states of nature. It is by this procedure that the simple decision problem is expanded.

2.3.3 Conclusions to Decision Procedures

In this section, the field of statistics dealing with decisions was discussed. The non-Bayesian approaches were concerned with either hypotheses tests or decision rules that did not address the cogent issue of what values of the decision variables will lead to the 'best' decisions given the information at hand, where 'best' is expressed as the most preferred (which may or may not be the most economical). The non-Bayesian procedures often required dubious ad hoc approaches and produced strategies that could not be evaluated with the above criterion. The Bayesian decision approach of the Savage school appears to be a more consistent procedure for making decisions under uncertainty and is the approach advocated by this thesis. Problems of application exist but there are no problems in the concepts as with non-Bayesian procedures.

Chapter 3

Assessment of Prior Information

3.1 Introduction

In Chapter 2, the discussion found upon many of the theoretical issues surrounding inference and decision making. It was shown that the Bayesian probability density function of the random variable, upon which decisions are based, provides a procedure that completely accounts for parameter uncertainty. On this basis, the Bayesian probability density function seems to be a rational approach for making inferences. Furthermore, applying the Bayesian probability density function and an appropriate utility function indicates to the decision maker a rational decision strategy. This procedure is recognized as Bayesian decision theory. An important part of the Bayesian approach is the assessment of prior information about uncertain variables. Such information is reflected in a prior probability density function. Within the Bayesian school, there exist two groups. One group, the 'objectivists' (Keynes, 1921 ; Jefferies, 1961) feels that probability statements should reflect a logical or necessary state of knowledge. Prior density functions should be based, therefore, on hard facts, and people, given the same information, should arrive at the same priors.

The other group, the 'subjectivists' (probably led by de Finetti and Savage), feel that probability statements reflect the beliefs and the willingness to act of the person making the inferences and dedisions. These beliefs may come from an empirical study, theoretical analysis or intuition. There is no requirement that, given the same information, two people should reach the same prior probability density distributions. One may have certain feelings, which he can not formally express, but which may have bearing upon the problem at hand.

In water resource planning, there exists a large body of theoretical and empirical procedures for the estimation of uncertain quantities in the absence of observed data. Also, in addition to such procedures, experienced engineers have often applied their judgement in making evaluations. This chapter will look at some methods for evaluating prior probability density distributions in absence of observed data.

3.2 Subjective Prior Assessment.

Engineers have often applied 'engineering judgement' to obtain insights and assessments to parameters which could not be readily obtained by other means. For example, a hydrologist may try to estimate the percentage of rainfall that is infiltrated (or the parameters of infiltration models) by applying his judgement, based on his knowledge of the physical characteristics of the basin. In essence, he is looking for some point estimates such as the mean on the probability density function (pdf) for that variable. The last chapter showed that one could better account for the effects of uncertainty by considering the complete pdf rather than its central moment.

Subjective evaluation of the pdf is difficult because a decision maker often can not explicitly state a particular fractile value. Instead, the value may be implicitly evaluated by his actions. This leads back to the notion that subjective probabilities, which are based upon the decision maker's beliefs, are inseparable from his actions.

Probability assessments require experimentation that lets the decision maker choose between actions. This is often in the form of one lottery, L_1, which is compared to another lottery, L_2, with known probabilities of winning or losing. If both lotteries have the same payoffs and the decision maker is indifferent to whether he 'plays' one lottery or the other, the subjective probability of lottery L_1 is equal to the known probability of L_2.

Such experimentation, though, is influenced by the procedural setup and, as such, may affect the results. Other problems may arise due to the following difficulties:

1) The questioning of the experiment or lottery

2) The order of the questions

3) The order in which random variables are estimated.

If more than one parameter of a distribution is desired (mean and variance), the order of their evaluation may affect the results since previous answers influence current assessments.

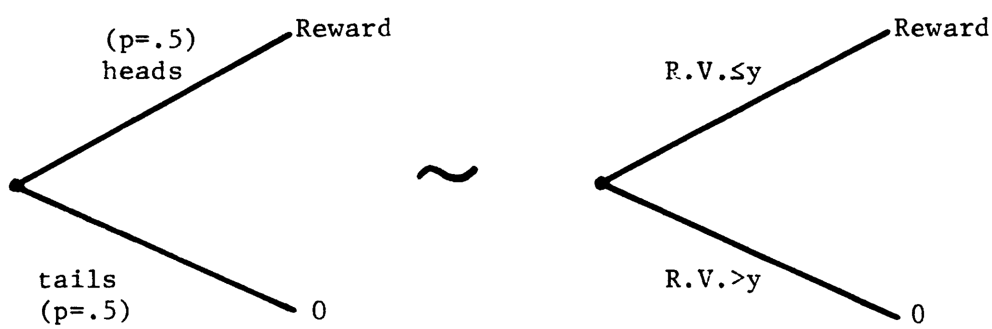
The assessments of prior distributions may be done on a single future state of nature which is the random variable - for example, the maximum peak discharge from spring runoff after observing the winter snow fall, but before melting takes place. Assessment may also be used for the distribution of the state of nature where the random variables are the parameters of the distribution. The distribution of annual peak flows may be known, but the mean and variance will be unknown parameters, and therefore treated as random variables.

There are two effective methods for directly assessing subjective distribution on random variables. These are:

a) The Fractile Method.

The fractiles of the cumulative distribution may be assessed directly or indirectly by breaking the random variables into intervals that are equally likely. It is most convenient for the decision maker to assess fractiles that do not require a high degree of refinement. Thus, the lower quartile (.25) is 'easier' to assess than, let's say, the .31 fractile. Usually about 5 fractiles are estimated, and a pdf is fitted through the points. A common procedure (Raiffa, 1968) is to approach the problem by having the decision maker consider the following set of questions:

1) "At what value of the random variable (R.V.) do you feel that there is an equally likely chance that the true value of the R.V. will be above or below?" This implies indifference between the following two lotteries



Let's assume he chooses $y = \bar{y}$. This will establish the .5 fractile.

2) "Suppose that the true value of the R.V. is greater than $\bar{y}$; at what value of y are you indifferent to the above set of lotteries?" This value will establish the .75 fractile. The same procedure can be used to estimate the .25 fractile.

3) "At what value of y do you feel the true value will exceed it only very rarely?" Such a procedure will try to establish the .99 fractile. A similar procedure is used for the .01 fractile. Extreme fractiles are very difficult to establish.

Further questions should be asked to assure consistency. If the values are inconsistent, this variance should be pointed out to the decision maker and be resolved by further thought. Other procedures, besides the above method, may be used to assess fractiles (Schlaifer, 1961) , but they are all based upon lottery actions.

The pdf can be also constructed from the cumulative fractiles. A direct way to obtain the pdf is first to use the lottery actions to determine the most probable value (mode) and values half as high as the mode, and then to find points which divide the area of the pdf into relative sizes.

In either the cumulative fractile method or probability density method, curves are fitted through the points.

b) Restrictive pdf Method.

Another approach to the estimation of the prior pdf is the restriction of the density function to a particular family and the estimation of the parameters of the function. The choice of the density function is not restricted by theory, but often a natural conjugate family is chosen. A natural conjugate prior density function has the property in which the posterior distribution is of the same form as the prior. This property simplifies the inference analysis and often permits analytical derivations of the Bayesian distribution. Since most of the weight of the uncertain parameter is near the mean of its pdf and not in the tails, the resulting inferences will be relatively insensitive to the functional form of the prior as long as the location and scale parameters are closely estimated.

3.2.1 Assessment of Joint Distributions

The methods of subjective assessment for joint multivariate priors are more complex than for the density function of a single random variable. Consider the probability function $f(Q|\theta_1,\theta_2)$, where θ_1 and θ_2 are unknown. The subjective joint probability $f(\theta_1 . \theta_2)$ must be assessed. But if independence can be assumed, then

$$f(\theta_1 \theta_2) = f(\theta_1) f(\theta_2) \tag{3.1}$$

and each marginal density function can be assessed by applying either of the methods above.

If θ_1, θ_2 are not independent, then $f(\theta_1 \theta_2)$ may be evaluated by one of the following procedures:

1. Evaluate $f(\theta_1 \theta_2)$ by assessing the relative probabilities of combinations of variables by the fractile method. This evaluation requires a large number of separate assessments.

If the form of $f(\theta_1\theta_2)$ is restricted to a particular family of density functions, then the assessment of its parameters may be easier.

2) Evaluate $f(\theta_1\theta_2)$ by

$$f(\theta_1\theta_2) = f(\theta_1|\theta_2)\ f\ (\theta_2) \qquad (3.2)$$

If the fractile method is used, successive assessments would be required for $f(\theta_1|\theta_2)$. Again, an easier approach may be the restriction of the prior pdf to a particular family and the assessment of the parameters of the pdf.

Making assessments on joint distributions of many variables is extremely difficult unless independence among some of the variables can be assumed.

3.3 Empirically Based Priors.

3.3.1 Introduction.

Empirical formulas for estimating peak flood discharges have been used for many years. A good summary is given in the Handbook on the Principals of Hydrology (Gray, 1970). All flood formulae take the general form of a regression involving physiographic characteristics, such as basin area and rainfall-runoff characteristics, such as excess precipitation or time of concentration. The coefficients must be evaluated for a given region where the formula is to be applied.

These formulae supposedly provide a best estimate as expected value of the flood peak. The variance of the estimator can not be obtained directly. One procedure for obtaining the variance of a formula is to determine the estimated flood peak, $\hat{q}$, for a number of sites where historical values, q, are known. A histogram of the ratio $y = q/\hat{q}$ would provide a basis for the estimation of $f(y)$. For ungauged sites, $f(y)$ and $\hat{q}$ (estimated for the ungauged sites) would be combined to yield $f(q)$.

In applying a flood formula and estimating its coefficients, it should be realized that there exists a set of coefficients that will minimize the dispersion of the predicted flood peak about the observed peaks. Choosing the minimum variance estimators is, in essence, the procedure of regression. Linear regression predicts values of the dependent variable (flood peaks) from a linear equation of independent variables (physiographic and meteorological factors.) The set of coefficients of the linear equation is chosen so as to minimize the squared difference between predicted and observed values.

Linear regression can indicate how much of the variability in the observed data is 'explained' by each independent variable and how much is left 'unexplained'. It provides an estimate of the variance of a predicted value of the dependent variables, both for sites included in the regression and for new or ungauged sites.

Classical multivariate regression has been used to estimate streamflow characteristics by Matalas and Benson (1961), Matalas and Gilroy (1968), and Thomas and Benson (1970). Benson (1962) used linear regression to estimate peak discharges for the T year flood.

Regression is a minimum variance estimator that provides both the expected value and the variance of the dependent variable at an ungauged site. Its application may provide useful prior information for estimating flood discharges.

The classical multivariate regression model is of the form:

$$\underline{Y} = \beta_o + \underline{\underline{Z}}\,\underline{\beta}_Z + \underline{\varepsilon} \qquad (3.3)$$

where $\underline{Y}$ is an nx1 vector of observations

$\underline{\underline{Z}}$ is an nxk matrix of known basin physiographic characteristics

$\underline{\beta} = \beta_o + \underline{\beta}_Z$ is a (k+1) x 1 vector of parameters

$\underline{\varepsilon}$ is an nx1 vector of error residuals such that $E[\underline{\varepsilon}] = \underline{0}$ and $V[\underline{\varepsilon}] = \underline{\underline{I}}\,\sigma^2$

and n is the number of basins included in the regression

k is the number of physiographic and meteorologic variables in the regression

$\underline{\underline{I}}$ is the identity matrix which implies that the error terms are uncorrelated.

For ease in notation, define $\underline{\underline{X}}$ as

$$\underline{\underline{X}} = (\underline{1}^T + \underline{\underline{Z}}) \qquad (3.4)$$

where $\underline{1}^T$ is a column vector of 1's and the regression model represented by (3.3) may be rewritten as

$$\underline{Y} = \underline{\beta} \cdot \underline{\underline{X}} + \underline{\varepsilon} \qquad (3.5)$$

The use of the classical regression model implies that the following assumptions hold that:

1) the independent variables have fixed known values.

2) The values of the dependent variables $\underline{Y}$ are mutually independent.

3) The variance of the error terms is homoscedastic - that is, constant for all y_i .

A further assumption that the error terms distributed $N(0,\sigma^2)$ is often made to facilitate the application of confidence tests. Generally the errors will be normally distributed due to the Central Limit Theorem, but the assumption is not required to perform the regression.

The assumption that the dependent variables are independent will, in general, hold for mean annual floods between basins. It will also be true for the variance of mean annual floods. The assumption that the error terms are homoscedastic and uncorrelated is usually checked after the regression is performed, with the assumption included.

Equation (3.5) can be solved for an estimate of $\underline{\beta}$, which will minimize the error sum of squares, $\underline{\varepsilon}^T\underline{\varepsilon}$. It is well known that this estimate, $\underline{b}$ is

$$\underline{b} = (\underline{\underline{X}}^T\underline{\underline{X}})^{-1}\underline{\underline{X}}^T\underline{Y} \qquad (3.6)$$

and has the following properties (Draper and Smith, 1966):

1. $\underline{b}$ is the estimate of $\underline{\beta}$ which minimizes the error sum

of squares $\varepsilon^T\varepsilon$ irrespective of any distributional properties of $\underline{\varepsilon}$.

2. $\underline{b}$ provides a minimum variance unbiased estimator for $\underline{\beta}$ irrespective of distribution properties of $\underline{\varepsilon}$.
3. If the errors are independent and distributed $N(0,\sigma^2)$, then $\underline{b}$ is the maximum likelihood estimate of $\underline{\beta}$.

An estimate of σ^2, s^2, can be found from the mean square error, thus

$$s^2 = \underline{Y}^T \underline{Y} - \underline{b}^T \underline{\underline{X}}^T \underline{Y} / (n - k) \tag{3.7}$$

Without assuming any distributional properties of s^2 (or σ^2) and observing $\underline{X}_{n+1}$, the vector of known physiographic factors for the 'next' or ungaged basin, the following hold true:

1. $$E[\tilde{Y}_{n+1}] = \underline{X}_{n+1} \underline{b} \tag{3.8}$$

2. $$V[\tilde{Y}_{n+1}] = [1 + \underline{X}_{n+1}^T (\underline{\underline{X}}^T \underline{\underline{X}})^{-1} \underline{X}_{n+1}] s^2 \tag{3.9}$$

where $\tilde{y}_{n+1}$ is an estimate of the dependent variable for the $(n+1)^{th}$ basin

$E[\cdot]$ is the expectation operator

$V[\cdot]$ is the variance operator

These moments can therefore be used to determine the parameters of the prior distributions by applying the restrictive pdf method of Section 3.2.

3.3.2 Flood Regression for Southern New England

The regression formulae of the preceding section is applied to

36 basins in southern New England. The resulting regression equation is then applied to an ungaged basin to obtain prior information on the parameters of the pdf for flood events.

Assume that the moments (mean and variance) of the pdf of annual floods are affected by the following physiographic and meteorological characteristics

1. Basin Area, A; in square miles
2. Mean main channel slope, $s\ell$; in feet per mile
3. Surface storage area + 0.5, St, in percent of drainage area.
4. Orographic factor, 0; dimensionless. This factor takes into account, empirically, the effect of mountain ranges and prevailing storm patterns upon the meteorological processes which effect flood peaks.
5. Main channel length, L_s; in miles
6. Rates of the 100 year 24-hour rainfall intensity to the mean annual 24-hour rainfall intensity, $R_{h/m}$
7. Mean January temperature below freezing, T_J in degrees Fahrenheit. This is used as a measure of the potential effects of large snowmelts to flood peaks.

These seven characteristics were adapted from Benson (1962).

The effect of these seven characteristics upon the mean or variance of the series of annual flood events may be in an additive manner or may be in a multiplicative manner. The assumption that the effect is additive leads to a linear regression while the multiplicative assumption leads to a log-linear regression. Each will be discussed.

3.3.2(a) Linear Regression.

Let Y, which is the dependent variable, mean or variance, be related in an additive manner to the given basin characteristics. Thus

$$Y = b_1 + b_2 \cdot A + b_3 \cdot s\ell + b_4 \cdot St + b_5 \cdot 0 + b_6 \cdot L_s + b_7 \cdot R_{h/m} + b_8 \cdot T_J + \varepsilon \tag{3.9}$$

where $\varepsilon \sim N(0,\sigma^2)$

Equation (3.9) is in the form of the classical regression Equation of (3.5).

For the 36 basins, the dependent and independent variables are given in Table 3.1.

Table 3.2 gives the vector of regression coefficients $\underline{b}$. The regression for the mean of the annual flood had an $R^2 = .791$, and the regression for the variance $R^2 = .615$. This implies that 79.1% of the variance of the observed data of the mean was explained by the regression and 61.5% for the variance.

3.3.2(b) Log-Linear Regressions

Assume that Y is the dependent variable and is affected jointly by the seven basin characteristics. Thus

Table 3.1 Flood Regression Variables for Southern New England

MEAN OF ANNUAL FLOOD (CFS)	STANDARD DEVIATION OF ANNUAL FLOOD (CFS)	AREA (SQ MI)	CHANNEL SLOPE (FT/MI)	STORAGE % AREA	TEMPERATURE BELOW 32F (JAN)	OROGRAPHIC FACTOR	RATIO - 100 YR/2.33 YR 24 HR RAINFALL INTENSITY	CHANNEL LENGTH (MI)
312.	152.	12.	21.50	1.82	9.	1.00	2.03	7.8
2262.	1113.	183.	12.10	2.16	9.	1.00	2.11	38.6
1253.	566.	68.	33.90	3.30	11.	1.00	2.12	14.0
633.	224.	47.	37.30	6.26	11.	1.00	2.12	16.6
1361.	1537.	55.	24.40	4.08	13.	1.00	2.03	23.3
2217.	1026.	146.	31.80	2.51	13.	1.00	2.14	23.6
3611.	1673.	202.	26.60	1.50	11.	1.00	2.07	24.8
3384.	3352.	107.	40.70	3.80	9.	1.00	2.16	22.6
224.	95.	22.	6.23	3.19	7.	1.00	2.12	10.8
1112.	500.	124.	2.50	1.86	6.	1.00	2.14	27.1
356.	243.	35.	23.80	3.88	5.	1.00	2.27	11.0
481.	157.	42.	10.70	1.11	5.	1.00	2.31	16.3
835.	266.	72.	16.20	2.61	4.	1.05	2.06	14.0
3179.	4485.	121.	16.00	2.08	6.	1.85	2.25	25.6
1129.	1074.	29.	72.70	1.74	6.	2.00	2.29	10.3
2161.	2901.	157.	14.20	2.61	7.	1.70	2.44	26.5
1536.	834.	84.	15.10	1.49	5.	1.15	1.94	18.6
5436.	2644.	269.	27.40	4.17	15.	1.00	1.92	33.8
4029.	1889.	103.	56.50	0.50	14.	1.10	1.85	21.6
2101.	1436.	83.	49.00	1.36	14.	1.00	1.96	20.2
1813.	580.	71.	38.50	3.06	13.	1.00	1.93	22.6
1257.	1126.	36.	100.00	1.97	11.	1.00	2.04	9.6
322.	253.	12.	48.10	1.56	10.	1.00	2.03	7.1
4271.	2503.	375.	17.70	2.83	10.	1.00	2.04	42.8
2254.	1265.	53.	94.80	1.31	9.	0.95	2.22	16.3
941.	1185.	44.	38.20	2.83	9.	1.00	2.28	14.3
7233.	9001.	688.	14.79	1.94	8.	1.00	2.34	56.1
16890.	13173.	497.	28.80	1.14	7.	1.65	2.35	53.6
8742.	7448.	216.	38.90	2.76	7.	1.70	2.46	26.5
2003.	2182.	45.	61.20	1.74	5.	2.00	2.22	13.3
2970.	2985.	74.	27.00	1.69	5.	2.00	2.31	14.5
3097.	2236.	105.	29.70	1.51	4.	1.65	1.94	10.5
495.	295.	12.	47.00	2.48	2.	2.00	1.88	11.3
2024.	1483.	57.	47.70	1.71	9.	1.10	2.13	14.6
4709.	8074.	133.	34.00	2.29	6.	2.00	2.19	27.6
13478.	17961.	246.	20.70	2.23	7.	2.55	2.28	43.6

Table 3.2

Vector Element	Value	Variable
b_o	-9759.46	constant
b_1	7.40876	area
b_2	14.9054	slope
b_3	- 211.932	storage
b_4	167.318	temperature below 32°F
b_5	3314.25	orographic factor
b_6	1591.66	rainfall intensity ratio
b_7	125.944	channel length

Table 3.2: Value of Vector $\underline{b}$ for Linear Regression

$$Y = b_1 \cdot A^{b_2} \cdot s\ell^{b_3} \cdot St^{b_4} \cdot 0^{b_5} \cdot L_s^{b_6} \cdot R_{h/m}^{b_7} \cdot T_J^{b_8} \cdot u \qquad (3.10)$$

Taking natural logarithms of (3.10) transforms the regression equation into the linear form of (3.5). The same 36 basins were used for the estimate of the parameters of the log-linear regression. The independent and dependent variables are the log transforms of the values given in Table 3.1, and the regression coefficients are given in Table 3.3. The log-linear regression for the mean of the annual flood explained 95.7% (R = .957) of the observed variance of the logs of the data. The log-linear regression for the variance of annual floods had an $R^2 = .902$.

3.3.3 Checking the Regression Assumptions.

It is important to check that the assumptions implied in the regression formulation hold. These assumptions are given in Section 3.3.1.

The assumption of error terms being homoscedastic and uncorrelated is of particular importance and can be checked by studying the residuals.

A scattergram of the residuals can usually indicate whether the standard deviation of the residuals is constant over the range of the dependent variables $\underline{y}$ - indicating homoscedasticity. A scattergram of the residuals from the regression of the log of the mean of the annual flood against the log of the basin and meteorological characteristics is presented in Figure 3.1. The scattergram

Table 3.3

Vector Element	Value	Variable
b_o	-2.069×10^8	constant
b_1	-8.530×10^4	ℓn area
b_2	7.727×10^4	ℓn slope
b_3	-1.041×10^6	ℓn storage
b_4	1.375×10^6	ℓn (temperature below 32°F)
b_5	6.774×10^7	ℓn orographic factor
b_6	2.992×10^7	ℓn rainfall intensity ratio
b_7	3.485×10^6	ℓn channel length

Table 3.3: Values of Vector $\underline{b}$ for Log-Linear Regression

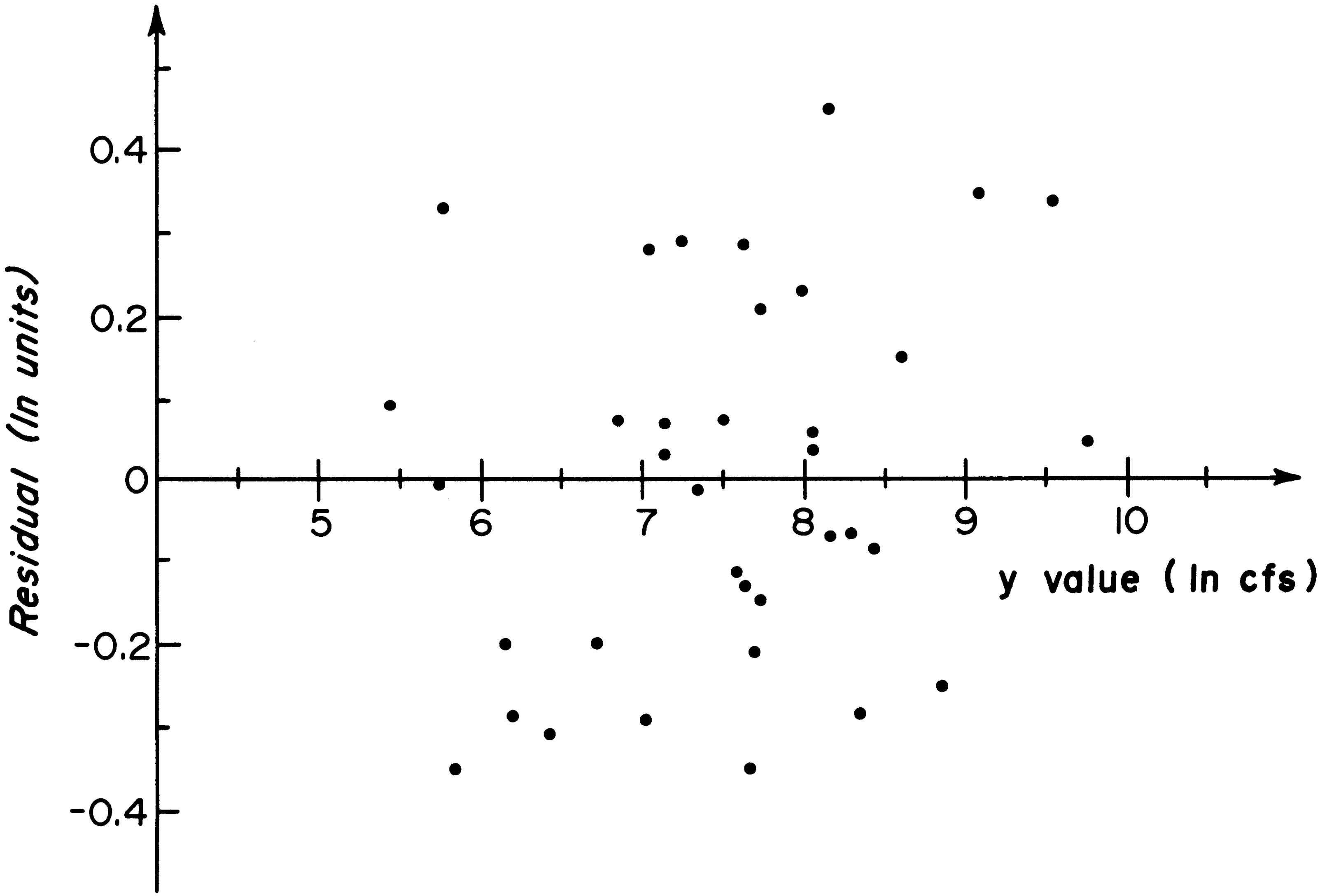


Figure 3.1: Scattergram of the Residuals from the Log-Linear Regression for Southern New England.

shows that the range of the residuals is constant over the range of $\ln \mu_Q$. Therefore, there are no indications that the assumption of homoscedasticity is invalid. Certain other statistics may be computed, but it is usually felt that a scattergram is more informative (Draper and Smith, 1966).

Correlation among the residuals is investigated by studying the covariance matrix of the residuals, which can be shown to be

$$\underline{\underline{V}}\ (\underline{\underline{e}}) = (\underline{\underline{I}} - R)\sigma^2 \tag{3.11}$$

where $\underline{\underline{V}}\ (\underline{\underline{e}})$ is defined as the covariance matrix of the residual matrix $\underline{\underline{e}}$

$\underline{\underline{I}}$ is the identity matrix

$\underline{\underline{R}} = \underline{\underline{X}}\ (\underline{\underline{X}}^T\ \underline{\underline{X}})^{-1}\ \underline{\underline{X}}^T$

σ^2 is the scalar variance of the error term $(v[\underline{\varepsilon}] = \underline{\underline{I}}\ \sigma^2)$

The variance of any residual e_i is given by the i^{th} diagonal and the covariance between $(e_i e_j)$ is given by the (i,j) the element of $(\underline{\underline{I}} - \underline{\underline{R}})\sigma^2$. The values of the correlation of the residuals depend entirely upon the independent variables $\underline{\underline{X}}$. Table 3.4 gives the matrix of the residual correlations. It is seen that the residuals are independent, as the regression assumption requires.

It is felt that there is no reason to think the regression on the mean annual flood violates the assumptions upon which it is based. A similar study for the variance of annual floods shows that no basis exists for concluding that it violates the assumptions implied in classical regression.

1.000	-.007	-.057	-.042	-.099	-.017	-.003	0.054	-.255	-.149	-.022	-.134	0.030	-.034	-.088	-.027	-.025	0.045	-.131	-.077	-.053	-.065	-.232	0.086	0.001	-.059	0.162	0.125	0.088	-.022	-.047	0.036	0.004	-.052	0.036	0.018
-.007	1.000	0.026	-.025	-.105	-.030	-.038	-.043	-.053	-.174	0.006	-.116	-.020	-.037	0.090	-.020	-.072	-.066	-.074	-.050	-.079	0.070	0.030	-.112	-.035	-.011	-.150	-.102	0.025	0.052	0.060	0.151	-.098	0.001	-.026	-.063
-.057	0.026	1.000	-.118	-.042	-.087	-.057	-.068	-.080	0.013	-.089	0.068	-.053	0.003	-.039	-.047	0.020	-.119	0.053	-.026	-.042	-.117	-.065	-.037	0.008	-.070	0.011	0.076	-.070	-.007	-.042	-.137	0.234	-.047	0.013	0.039
-.042	-.025	-.118	1.000	-.194	-.073	0.033	-.147	-.090	0.072	-.129	0.104	-.016	0.001	-.023	-.035	0.067	-.163	0.212	-.001	-.139	-.070	-.037	-.060	-.010	-.111	0.022	0.100	-.061	-.004	0.039	0.136	-.103	-.020	-.033	-.020
-.099	-.105	-.042	-.194	1.000	-.054	0.027	-.075	-.122	-.044	0.010	0.041	0.080	-.037	0.016	-.021	0.028	-.155	0.005	-.073	-.190	0.007	-.062	-.067	0.016	-.053	0.040	0.048	0.038	0.026	0.077	0.277	-.166	-.002	-.051	-.116
-.017	-.030	-.087	-.073	-.054	1.000	-.076	-.059	-.016	-.003	-.019	0.034	0.001	-.007	-.006	-.042	0.015	-.118	-.052	-.063	-.057	-.075	-.030	-.070	-.034	-.050	-.064	-.023	-.061	0.011	-.010	-.027	0.253	-.043	-.005	-.013
-.003	-.038	-.057	0.033	0.027	-.076	1.000	-.018	0.016	-.062	0.017	-.013	-.052	0.000	0.026	-.007	-.066	-.095	-.187	-.086	-.021	-.073	-.020	-.087	-.040	-.001	-.101	-.066	-.025	0.018	-.021	-.180	0.248	-.049	0.012	0.025
0.054	-.043	-.068	-.147	-.075	-.059	-.018	1.000	0.039	0.077	-.128	0.022	-.082	0.029	0.014	0.003	0.005	-.102	0.116	-.006	-.072	-.067	0.024	-.092	-.110	-.089	-.117	-.026	-.080	-.005	0.042	0.042	-.111	-.036	-.020	0.042
-.255	-.053	-.080	-.090	-.122	-.016	0.016	0.039	1.000	-.345	-.099	-.149	-.057	-.085	-.010	-.116	-.048	-.005	0.113	0.026	-.024	0.042	-.139	0.030	0.162	-.051	0.103	0.167	0.060	0.037	-.064	0.002	0.072	0.007	0.035	-.020
-.149	-.174	0.013	0.072	-.044	-.003	-.062	0.077	-.345	1.000	-.024	-.274	-.117	-.116	0.130	-.121	-.170	-.002	-.027	0.018	0.024	0.169	-.004	-.082	0.175	0.034	-.114	-.016	0.083	0.105	-.039	-.051	0.123	0.039	0.030	-.059
-.022	0.006	-.089	-.129	0.010	-.019	0.017	-.128	-.099	-.024	1.000	-.129	-.203	0.029	-.030	-.037	-.021	0.044	0.292	0.092	0.041	-.070	-.026	-.007	-.110	-.142	-.096	0.050	-.096	-.029	-.032	-.060	-.140	-.045	0.034	0.158
-.134	-.116	0.068	0.104	0.041	0.034	-.013	0.022	-.149	-.274	-.129	1.000	-.078	-.028	-.015	-.052	-.089	0.264	-.073	0.001	0.096	0.054	-.101	0.055	-.213	-.104	-.177	-.139	0.019	-.007	-.031	0.196	-.083	-.058	0.056	0.090
0.030	-.020	-.053	-.016	0.080	0.001	-.052	-.082	-.057	-.117	-.203	-.078	1.000	0.027	0.063	0.035	-.172	-.036	0.126	0.061	0.035	-.043	0.038	-.092	-.013	-.019	-.113	0.032	-.017	-.002	-.023	-.365	-.290	-.020	0.024	0.161
-.034	-.037	0.003	0.001	-.037	-.007	0.000	0.029	-.085	-.116	0.029	-.028	0.027	1.000	-.084	-.134	-.015	-.013	-.006	0.006	-.003	0.059	0.006	-.009	0.082	0.011	-.005	-.080	-.071	-.071	-.113	-.037	-.059	0.009	-.105	-.219
-.088	0.090	-.039	-.023	0.016	-.006	0.026	0.014	-.010	0.130	-.030	-.015	0.063	-.084	1.000	-.106	0.066	0.095	-.029	-.016	0.028	-.098	-.124	0.120	-.076	-.070	0.108	-.040	-.125	-.183	-.172	-.063	-.091	-.065	-.094	-.136
-.027	-.020	-.047	-.035	-.021	-.042	-.007	0.003	-.116	-.121	-.037	-.052	0.035	-.134	-.106	1.000	0.055	0.005	0.118	0.044	0.044	0.045	0.012	0.004	0.066	-.050	-.060	-.086	-.155	-.061	-.146	0.014	0.246	-.003	-.080	-.199
-.025	-.072	0.020	0.067	0.028	0.015	-.066	0.005	-.048	-.170	-.021	-.089	-.172	-.015	0.066	0.055	1.000	-.040	-.144	-.036	-.023	0.006	-.001	-.085	0.012	0.053	-.056	-.025	0.075	0.001	-.003	-.264	-.334	-.011	-.011	0.037
0.045	-.066	-.119	-.163	-.155	-.118	-.095	-.102	-.005	-.002	0.044	0.264	-.036	-.013	0.095	0.005	-.040	1.000	-.042	-.087	-.187	-.057	0.054	-.203	0.130	0.033	-.023	0.049	-.014	0.059	0.056	-.164	0.085	0.005	-.054	-.086
-.131	-.074	0.053	0.212	0.005	-.052	-.187	0.116	0.113	-.027	0.292	-.073	0.126	-.006	-.029	0.118	-.144	-.042	1.000	-.313	-.113	-.120	-.185	-.035	-.155	0.091	0.023	-.162	0.150	-.032	0.007	-.130	0.053	-.102	-.016	-.053
-.077	-.050	-.026	-.001	-.073	-.063	-.086	-.006	0.026	0.018	0.092	0.001	0.061	0.006	-.016	0.044	-.036	-.087	-.313	1.000	-.109	-.092	-.105	-.045	-.093	-.006	0.004	-.045	0.046	-.007	0.027	0.008	0.046	-.062	-.011	-.021
-.053	-.079	-.042	-.139	-.190	-.057	-.021	-.072	-.024	0.024	0.041	0.096	0.035	-.003	0.028	0.044	-.023	-.187	-.113	-.109	1.000	-.046	-.050	-.094	-.009	-.013	0.034	0.031	0.052	0.014	0.080	0.093	-.205	-.018	-.047	-.066
-.065	0.070	-.117	-.070	0.007	-.075	-.073	-.067	0.042	0.169	-.070	0.054	-.043	0.059	-.098	0.045	0.006	-.057	-.120	-.092	-.046	1.000	-.138	0.009	-.143	-.081	0.045	0.049	-.044	-.063	-.032	-.176	0.083	-.100	0.023	0.119
-.232	0.030	-.065	-.037	-.062	-.030	-.020	0.024	-.139	-.004	-.026	-.101	0.038	0.006	-.124	0.012	-.001	0.054	-.185	-.105	-.050	-.138	1.000	0.095	-.100	-.081	0.146	0.091	0.055	-.056	-.043	0.013	0.008	-.086	0.036	0.063
0.086	-.112	-.037	-.060	-.067	-.070	-.087	-.092	0.030	-.082	-.007	0.055	-.092	-.009	0.120	0.004	-.085	-.203	-.035	-.045	-.094	0.009	0.095	1.000	0.016	0.016	-.166	-.073	-.072	0.058	0.057	-.085	-.017	0.002	-.035	-.034
0.001	-.035	0.008	-.010	0.016	-.034	-.040	-.110	0.162	0.175	-.110	-.213	-.013	0.082	-.076	0.066	0.012	0.130	-.155	-.093	-.009	-.143	-.100	0.016	1.000	-.152	-.172	-.167	-.062	-.067	0.042	0.205	-.129	-.124	0.023	0.147
-.059	-.011	-.070	-.111	-.053	-.050	-.001	-.089	-.051	0.034	-.142	-.104	-.019	0.011	-.070	-.050	0.053	0.033	0.091	-.006	-.013	-.081	-.081	0.016	-.152	1.000	-.063	-.012	-.083	-.032	-.016	0.136	0.065	-.063	0.014	0.056
0.162	-.150	0.011	0.022	0.040	-.064	-.101	-.117	0.103	-.114	-.096	-.177	-.113	-.005	0.108	-.060	-.056	-.023	0.023	0.004	0.034	0.045	0.146	-.166	-.172	-.063	1.000	-.261	-.130	0.049	0.031	0.075	0.166	-.029	-.008	0.017
0.125	-.102	0.076	0.100	0.048	-.023	-.066	-.026	0.167	-.016	0.050	-.139	0.032	-.080	-.040	-.086	-.025	0.049	-.162	-.045	0.031	0.049	0.091	-.073	-.167	-.012	-.261	1.000	-.132	-.068	-.054	0.086	-.006	-.034	-.104	-.170
0.088	0.025	-.070	-.061	0.038	-.061	-.025	-.080	0.060	0.083	-.096	0.019	-.017	-.071	-.125	-.155	0.075	-.014	0.150	0.046	0.052	-.044	0.055	-.022	-.062	-.083	-.130	-.132	1.000	-.098	-.129	-.062	0.185	-.037	-.086	-.120
-.022	0.052	-.007	-.004	0.026	0.011	0.018	-.005	0.037	0.105	-.029	-.007	-.002	-.071	-.183	-.061	0.001	0.059	-.032	-.007	0.014	-.063	-.056	0.058	-.067	-.032	0.049	-.068	-.098	1.000	-.130	-.115	-.268	-.045	-.103	-.121
-.047	0.060	-.042	0.039	0.077	-.010	-.021	0.042	-.064	-.039	-.032	-.031	-.023	-.113	-.172	-.146	-.003	0.056	0.007	0.027	0.080	-.032	-.043	0.057	0.042	-.016	0.031	-.054	-.129	-.130	1.000	-.235	0.068	-.030	-.080	-.141
0.036	0.151	-.137	0.136	0.277	-.027	-.180	0.042	0.002	-.051	-.060	0.196	-.365	-.037	-.063	0.014	-.264	-.164	-.130	0.008	0.093	-.176	0.013	-.085	0.205	0.136	0.075	0.086	-.062	-.115	-.235	1.000	-.077	-.039	-.039	0.067
0.004	-.098	0.234	-.103	-.166	0.253	0.248	-.111	0.072	0.123	-.140	-.083	-.290	-.059	-.091	0.246	-.334	0.085	0.053	0.046	-.205	0.083	0.008	-.017	-.129	0.065	0.166	-.007	0.185	-.268	0.068	-.077	1.000	0.031	-.271	-.136
-.052	0.001	-.047	-.020	-.002	-.043	-.049	-.036	0.007	0.039	-.045	-.058	-.020	0.009	-.065	-.003	-.011	0.005	-.102	-.062	-.018	-.100	-.086	0.002	-.124	-.063	-.029	-.034	-.037	-.045	-.030	-.039	0.031	1.000	0.001	0.038
0.036	-.026	0.013	-.033	-.051	-.005	0.012	-.020	0.035	0.030	0.034	0.056	0.024	-.105	-.094	-.080	-.011	-.054	-.016	-.011	-.047	0.023	0.036	-.035	0.023	0.014	-.008	-.104	-.086	-.103	-.080	-.039	-.271	0.001	1.000	-.224
0.018	-.063	0.039	-.020	-.116	-.013	0.025	0.042	-.020	-.059	0.158	0.090	0.161	-.219	-.136	-.199	0.037	-.086	-.053	-.021	-.066	0.119	0.063	-.034	0.147	0.056	0.017	-.170	-.120	-.121	-.141	0.067	-.136	0.038	-.274	1.000

Table 3.4 Table of Residual Correlations from the Log-Linear Regression for Southern New England

3.3.4 Applying the Regression Formulae to 'Ungauged' Basins.

The log-linear and the linear regression equations developed in Section 3.3.2 can be applied to other basins in New England to estimate the first two moments of the mean and variance of the annual flood series. The regression formulae were applied to the Blackstone River at Woonsocket, Rhode Island. The values of the independent variables for the two basins are given in Table 3.5.

For the Blackstone River, the linear regression for the mean, m_y, of the annual flood provided the following moments.

$$\begin{aligned} E[m_y] &= 5188 \text{ cfs} \\ V[m_y] &= (2246)^2 \text{ cfs}^2 \end{aligned} \tag{3.12}$$

and for the regression on the variance, σ^2_y,

$$\begin{aligned} E[\sigma^2_y] &= 45 \times 10^6 \text{ cfs}^2 \\ V[\sigma^2 \ ] &= 48.5 \times 10^6 \text{ cfs}^4 \end{aligned} \tag{3.13}$$

The log-linear regression applied to the Blackstone River predicted the moments of the log-mean of the annual flood, $\ell n\ m_y$, and the log-variance of the annual flood, $\ell n\, \sigma^2_y$ which are:

$$\begin{aligned} E[\ell n\ m_y] &= 8.27 \\ V[\ell n\ m_y] &= .073 \end{aligned} \tag{3.14}$$

$$\begin{aligned} E[\ell n \sigma^2_y] &= 15.52 \\ V[\ell n\, \sigma^2_y] &= 1.031 \end{aligned} \tag{3.15}$$

The regression assumes that these moments are for the $\ell n\ m_y$ and $\ell n\sigma^2_y$ normally distributed and thus m_y, σ^2_y are log-normally dis-

Table 3.5

Area = 416 sq. miles

Channel slope = 11.50 ft/mi.

Storage area = 3.51 % of area

Temperature below 32°F (January) = 6°F

Orographic factor 1.0

Rates of 100 year to 2.33 year, 24. hr rainfall intensity = 2.22

Length of main channel = 42.6 miles

Table 3.5: Independent Variables for Blackstone River at Woonsocket, R.I.

tributed, To obtain the moments of m_y and σ^2_y, the following transform is used (Benjamin and Cornell, 1970):

$$\left.\begin{aligned} E[\omega] &= \exp\left(E[\ell n\ \omega] + .5\ V[\ell n\ \omega]\right) \\ V[\omega] &= E^2[\omega] \cdot \{\exp\left(V[\ell n \omega]\right) - 1.0\} \end{aligned}\right. \qquad (3.16)$$

where $E[\cdot]$ = expectation operator

$V[\cdot]$ = variance operator.

Letting $\omega = m_y$ and substituting (3.14) into (3.16) gives

$$\left.\begin{aligned} E[m_y] &= 4042 \text{ cfs} \\ V[m_y] &= 1.24 \times 10^6 \text{ cfs}^2 \end{aligned}\right. \qquad (3.17)$$

and for $\omega = \sigma^2_y$, substituting (3.15) into (3.16) gives

$$\left.\begin{aligned} E[\sigma^2_y] &= 9.22 \times 10^6 \text{ cfs}^2 \\ V[\sigma^2_y] &= 1.537 \times 10^{14} \text{ cfs}^2 \end{aligned}\right\} \qquad (3.18)$$

The information from the regression, represented by Equations (3.12), (3.13), (3.17) and (3.18), can be utilized in determining the parameters of the prior probability density function. If the form of the prior is restricted to a particular family of density functions, then the procedures set up in Section 3.2 can be used. Essentially, the procedure is fitting the information on the numerical values of the moments to the algebraic expressions for the moments of the density function and solving for the parameters of the distribution.

3.3.5 Summary.

This section studied the use of empirical procedures for obtaining information about the mean and variance of the annual series of flood events. Empirical procedures are widely used in flood estimation and rely on the assumption that the basin under study will 'act' in a similar manner to those near by.

Empirical procedures lack causal validation. Two empirical formulations may fit the data quite well, but they may predict different values for the "ungaged" basin. If both fit the data equally well, there is no method to distinguish between them.

It will be shown in Chapter 4, when the regression model is used to provide prior information that the regression provided information equivalent to 4 to 7 years of data. Chapter 4 applies both sample and prior information to a number of probability distribution models of the occurrence of floods.

3.4 Evaluation of Theoretically Based Prior Information

3.4.1 Introduction

The evaluation of prior information from empirical procedures, such as regression equations or flood formulae, suffers from a deficiency due to the procedures being based upon data from similar or near-by areas. Therefore, empirical procedures may not capture all the congent components of the rainfall-runoff process for the basin being studied.

The analysis of the runoff characterictics, using rainfall and basin characteristics, can be performed in the following manner (Eagleson, 1972; Leclerc and Schaake, 1973).

Let Q be the peak discharge and $\underline{\Theta}$ a vector of variables representing rainfall and antecedent moisture conditions. It is well known that Q is a function of $\underline{\Theta}$ even though this functional relationship is a complex process. This process depends upon the basin's characteristics and its response to the elements of $\underline{\Theta}$. From the response and from the probability density function for $\underline{\Theta}$, $f(\underline{\Theta})$ the cumulative density function (CDF) for the peak discharge can be calculated by:

$$F_Q(q) = \int_{R_q} f(\underline{\Theta})\ d\underline{\Theta} \qquad (3.16)$$

where

$f(\underline{\Theta})$ is the probability density function of the matrix of rainfall and antecedent moisture condition variables

R_q is the region in the $\underline{\Theta}$ space where $Q(\underline{\Theta})$ is less than or equal to q. This region is defined by the basin response to $\underline{\Theta}$.

Eagleson (1972) simplifies the relationship between runoff and rainfall by applying kinematic wave theory for hydrograph forecasting. Leclerc and Schaake (1973) use deterministic simulation of the catchment to estimate this relationship. By using a storm selection procedure, which only considers the worst storms of a family of synthetically generated storms, Leclerc and Schaake determined the flood frequency curve.

Both procedures may be useful in providing prior information on uncertain parameters. This information can then be combined with historical information. Eagleson's procedure will be considered in detail here.

3.4.2 Eagleson's Analytical Derivation of the Flood Frequency Curve

Eagleson (1972) analytically derives the flood frequency curve by taking a stochastic model of the rainfall process and applying the rainfall to an idealized overland flow plane. The mechanics of that flow are analyzed by kinematic wave theory.

The stochastic model of individual rainfall events is formulated in terms of two random variables; the rainfall intensity, i, and the storm duration, t_r. Each storm is assumed to have a rectangular storm interior, and the excess rainfall intensity is assumed to be constant throughout the duration of the storm.

Eagleson, under the assumption that the time of concentration for the stream is greater than the time of concentration for the overland flow catchment, derives the exceedance probability of a flood peak discharge, q_p, as

$$G_Q(q_p) = e^{-2\sigma}\sigma^{-\sigma+1}\Gamma(\sigma)\exp[-\beta\frac{(q_p-\bar{q}_b)}{645KA_r}] \quad (3.17)$$

where $G_Q(\cdot)$ is the compliment of the cumulative density function, $F_Q(\cdot)$ and where σ, β, $\bar{q}_b$, K and A_r are parameters obtained from either the stochastic rainfall model or the deterministic runoff model. Appendix E gives a summary of Eagleson's derived flood formula, and all parameters are defined therein.

3.4.3 Application of $G(q_p)$ for the Assessment of Priors

The flood frequency formula, represented by Equation (3.19), is for the complete series of flood peaks. These peaks are represented by the n·N rainfall events that produce direct runoff, where n is the average annual number of such events and N is the number of years of record.

Consider a series of the largest N events from the series of n·N events. This series is the annual exceedance series. It can be shown that the exceedance probability of the annual exceedance series is related to the complete series by

$$G_e(Q \geq q_p) = n \cdot G_Q(q_p) \quad (3.18)$$

The annual series of flood peaks consist of the largest event in each of the N years of record. The exceedance probability of the annual series, $G_a(q \geq q_p)$, is related to the exceedance probability of the annual exceedance series by, (Chow, 1964),

$$G_q(q \geq q_p) = 1 - \exp[-G_e(q \geq q_p)] \tag{3.19}$$

for events of the same rank. Eagleson's flood frequency formula can be transformed to give the cumulative probability of the annual series, $F_a(q) = 1 - G_a(q)$. This transformation is performed by the substitution of $G_e(q_p)$ from (3.18) into (3.19) and gives

$$F_a(q_p) = \exp\{-nI_o \cdot \exp[-b(q_p - q_b)]\} \tag{3.20}$$

where

$$I_o = e^{-2\sigma}\sigma^{-\sigma+1}\Gamma(\sigma)$$

$$b = \beta/(645\ K\ A_r)$$

Manipulation of (3.20) yields

$$F_a(q_p) = \exp\{-\ \exp[-b(q_p - u]\} \tag{3.21}$$

where

$$u = q_b + \frac{1}{b}\,\ell n(nI_o)$$

Equation (3.21) is of the form of a Gumbel Type I Extreme Value distribution with mean and standard deviation as

$$\hat{m} = u + \frac{\gamma\sqrt{6}}{\pi} \sigma_q \qquad (3.22)$$

$$\sigma_q = \frac{\pi}{\sqrt{6}\, b} \qquad (3.23)$$

where $\gamma = .5772157$. (Euler's constant)

Thus, Eagleson's flood frequency formula can be considered as a Gumbel extreme value distribution, where the parameters are estimated from physiographic characteristics rather than from observed data.

To obtain prior information on the mean and variance of the distribution of annual floods for an 'ungaged' basin within the region, the following procedure is applied:

1. From historical records and from Eagleson's analysis (Equations 3.22 and 3.23), obtain estimates of the mean and variance of the annual series of flood events for a number of basins within the region.

2. Consider the random variable Y_m, whose definition is:

$$Y_m = \frac{\text{historical mean}}{\text{Eagleson's predicted mean}} \qquad (3.24)$$

which should have $E[Y_m] = 1$. Similarly, obtain a ratio of the variances, Y_v. From the observed basin ratios, the sample moments are calculated and a density function fitted for the random varialbes Y_m and Y_v.

3. For the 'ungaged' basin, calculate Eagleson's estimate of the mean of the annual series of flood events. If it is assumed that the historical estimate of the mean is a true estimate (no variance), then the distribution of the mean μ for the ungaged

basin can be found from the pdf for Y_m and Eagleson's predicted mean $\hat{m}$. This pdf for μ is

$$f(\mu) = \frac{1}{\hat{m}} f(Y_m) \tag{3.25}$$

If $f(Y_m) \sim N(1,\sigma^2)$ then $f(\mu) \sim N(\hat{m},(\hat{m}\sigma)^2)$

4. Using a procedure similar to 3, calculations for the distribution on the variance are performed.

The procedures 1 through 4 were followed for the same 36 New England basins that were used in the regression analysis. Their location is shown in Figure 3.2. An index to the plate numbers is given in Table 3.6. Using the values for the rainfall parameters given by Eagleson (1972), Table 3.7 presents the ratio of historical to predicted means, Y_m, for the 36 basins. Similarly, the ratios of the variances, Y_v, are given in Table 3.8. The moments for the ratio of the means are

$$\begin{aligned} E[Y_m] &= .995 \\ V[Y_m] &= .235 \end{aligned} \tag{3.26}$$

The moments for the ratio of the variances, Y_v, are

$$\begin{aligned} E[Y_v] &= 5.86 \\ E[Y_v] &= 85.3 \end{aligned} \tag{3.27}$$

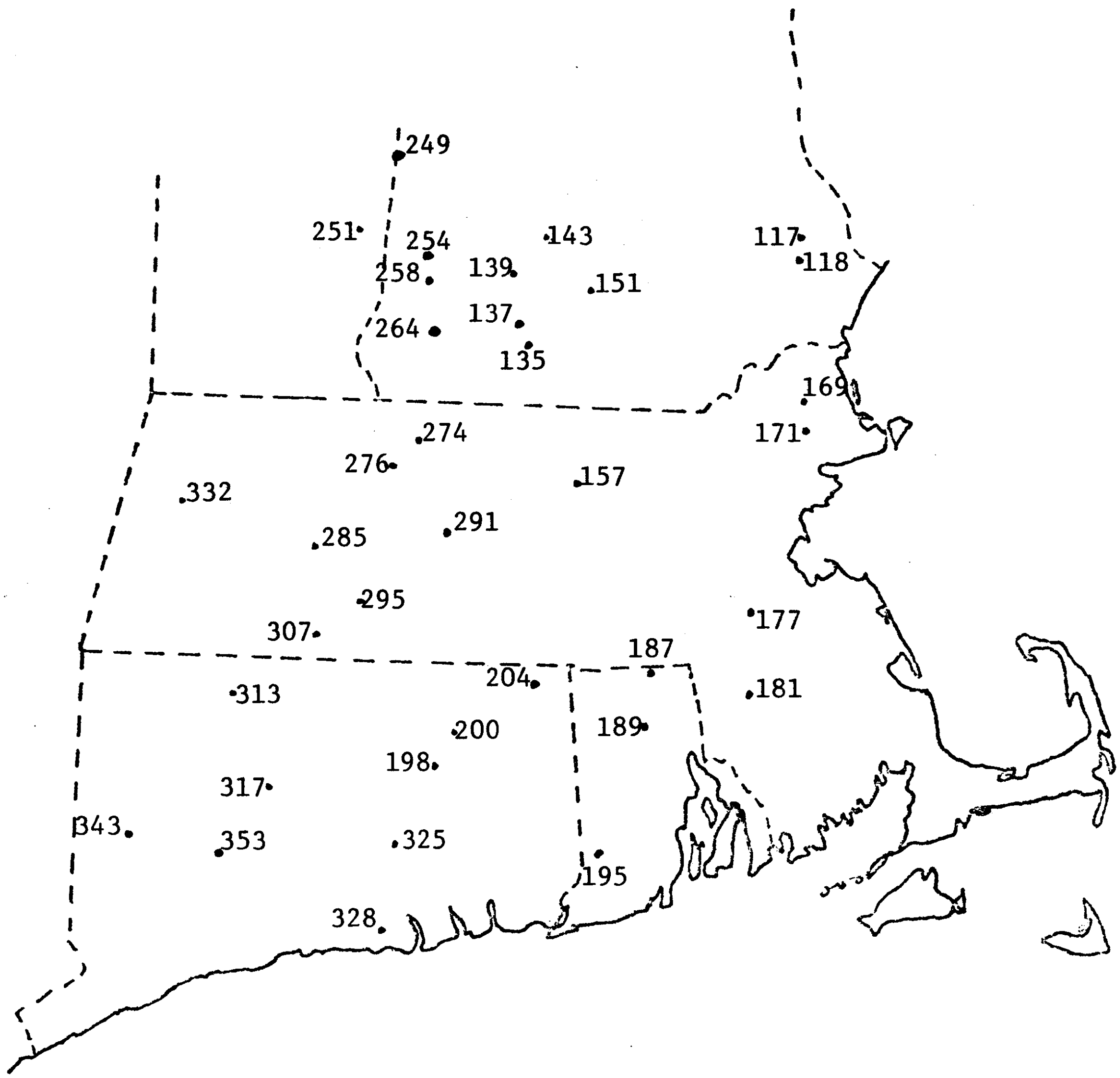


Figure 3.2: Location of Basins in Southern New England for the Assessment of Prior Information

Station No.	Location
117	Oyster River near Durham, N.H.
118	Lamprey River near Newmarket, N.H.
135	Contoocock River at Peterboro, N.H.
137	Naranusit Brook near Peterboro, N.H.
139	North Branch Contoocock River near Antrim, N.H.
143	Warner River at Davisville, N.H.
151	Piscataquog River near Gofftown, N.H.
157	North Nashua River near Loominster, Mass.
169	Parker River Basin at Ryfield, Mass.
171	Ipswich River near Ipswich, Mass.
177	Neponset River at Norwood, Mass.
181	Wading River Near Norton, Mass.
195	Wood River at Hope Valley, R.I.
198	Willimantic River near South Coventry, Conn.
200	Mount Hipe River near Warrenville, Conn.
204	Quinebaug River at Quinebaug, Conn.
209	Moosup River at Moosup, Conn.
249	Sugar River at West Claremont, N.H.
251	Williams River at Brockway Mills, Vt.
254	Cold River at Drewville, N.H.
258	Ashuelot River near Gilsum, Vt.
264	South Branch Ashuelot River at Webr. Near Marlboro, N.H.
274	Moss Brook at Wendell Depot, Mass.
276	Millers River at Frving, Mass.
285	Mill River at Northampton, Mass.
291	East Branch Swift River near Hardwick, Mass.
295	Chicopee River at Indian Orchard, Mass.
307	Westfield River near Westfield, Mass.
313	West Branch Farmington River at Riverton, Conn.
317	Pequabuck River at Forestville, Conn.
322	Park River at Hartford, Conn.
325	Salmon River near East Hampton, Conn
328	Manunketesuck River near Clinton, Conn.
332	East Branch Housatonic River at Coltsville, Mass.
343	Shepaug River near Roxbury, Conn.
353	Naugatuck River near Beacon Falls, Conn.

Table 3.6 Location of Stream Gauges for Figure 3.2

Station No.	Predicted Mean	Historical Mean	Ratio Y_M
117	375	312	0.8322
118	4137	2262	0.5468
135	1855	1253	0.6755
137	1327	633	0.4769
139	1598	1361	0.8518
143	3666	2217	0.6047
151	4548	3611	0.7939
157	2683	3384	1.2612
169	610	224	0.3675
171	2602	1112	0.4274
177	1107	356	0.3215
181	1070	481	0.4495
195	1910	835	0.4373
198	2888	3179	1.1007
200	907	1129	1.2454
204	3567	2161	0.6058
209	1997	1538	0.7700
249	5748	5436	0.9457
251	2763	4029	1.4581
254	2304	2101	0.9120
258	1974	1813	0.9184
264	992	1257	1.2673
274	389	322	0.8288
276	6930	4271	0.6163
285	1529	2254	1.4744
291	1263	941	0.7452
295	11082	7233	0.6527
307	9487	16890	1.7803
313	4719	8742	1.8525
317	1272	2003	1.5750
322	1872	2970	1.5869
325	3317	3097	0.9337
328	379	495	1.3057
332	1586	2024	1.2762
343	3237	4709	1.4549
353	5437	13478	2.4790

Table 3.7 Ratio of Means, Boston Parameters

Station No.	Predicted Variance	Historical Variance	Ratio Y_V
117	19921	23000	1.1546
118	3922824	1238000	0.3156
135	598393	343000	0.5732
137	292637	50000	0.1709
139	397093	2362000	5.9482
143	2566149	1053000	0.4103
151	4719990	2798000	0.5928
157	1425173	11238000	7.8854
169	66094	9000	0.1362
171	1885135	250000	0.1326
177	164606	59000	0.3584
181	235057	39000	0.1659
195	668014	71000	0.1063
198	1799714	20118992	11.1790
200	113841	1153000	10.1282
204	2942131	8418000	2.8612
209	898056	695000	0.7739
249	8062896	6990000	0.8669
251	1325536	3568000	2.6917
254	877686	2062000	2.3494
258	650275	961000	1.4778
264	173936	1267000	7.2843
274	19921	64000	3.2128
276	15027104	6264000	0.4168
285	369572	1599000	4.3266
291	257370	1405000	5.4591
295	47878288	81024992	1.6923
307	25612016	173538000	6.7756
313	5350316	55468992	10.3674
317	268887	4759000	17.6989
322	704160	8911000	12.6548
325	1374941	5000000	3.6365
328	19921	87000	4.3673
332	425543	2199000	5.1675
343	2152184	65184992	30.2878
353	6822592	322579968	47.2811

Table 3.8 Ratio of Variances, Boston Parameters

To obtain unbiased estimates of the expected value of the ratio of the means, Y_m, the fraction of area that contributed to direct runoff, FA_r, had to be adjusted. For the predictions using the Boston rainfall parameters, as being typical for all New England, FA_r was adjusted to .43.

The results for the estimation of the variance shows that Eagleson's procedure, on the average, significantly underestimates the variance. The estimate for the ratio of the historical to predicted variances is 5.86, which is highly biased. The variance in the estimate of Y_v is 85.3. The large variance, taken together with the biased $E[Y_v]$, signifies that a large amount of risk is associated with using the variance ratio.

The reason for the poor estimates in the variance of the annual flood series may be found by rewriting the equations for the mean and standard deviation of the cumulative annual flood discharge as

$$\hat{m} = \frac{645\ KA_r}{\beta}\,[\ell n(nI_o) + \gamma] \tag{3.28}$$

$$\sigma_q = \frac{645KA_r}{\beta} \cdot \frac{\pi}{\sqrt{6}} \quad \text{(3.29}$$

All terms have been defined previously and the reasonable assumption of no base flow has been made. Both β and I_o depend upon the statistics of rainfall pdf. $1/\beta$ is the mean rainfall intensity, $1/\lambda$ is the mean storm duration and I_o is a function of both β and λ, as well as the area contributing to direct runoff and other basin parameters. Better estimates of the rainfall parameters could lead to better prediction of the mean and variance, if the underlying structure of the model is at least reasonably correct.

Rainfall data for twelve gages around New England were analysed to find the best values for the rainfall parameters. Figure 3.3 shows the location of these gages. Tables 3.9 through 3.20 give the monthly statistics for the time between storms, storm duration, storm depth, and average intensity. In general, significant variations occur between stations, and significant seasonal variations occur within the year at each station. Average annual statistics were calculated and a surface fitted, using multiquadric cones, (Shaw and Lynn, 1972). The surface for average intensity is shown in Figure 3.4, for average storm duration in Figure 3.5 and for the average number of storms per year in Figure 3.6.

Using these surfaces, the best estimates for the rainfall parameters were obtained for each of the 36 basins. These values are given in Table 3.21.

For the 36 basins, the ratio of historical mean to predicted mean, Y_m, is presented in Table 3.22. The results for the ratios of the variances Y_v, are presented in Table 3.23.

The sample moments for Y_m are:

$$E[Y_m] = .975$$

$$V[Y_m] = .247 \qquad (3.30)$$

and the sample moments for Y_v are:

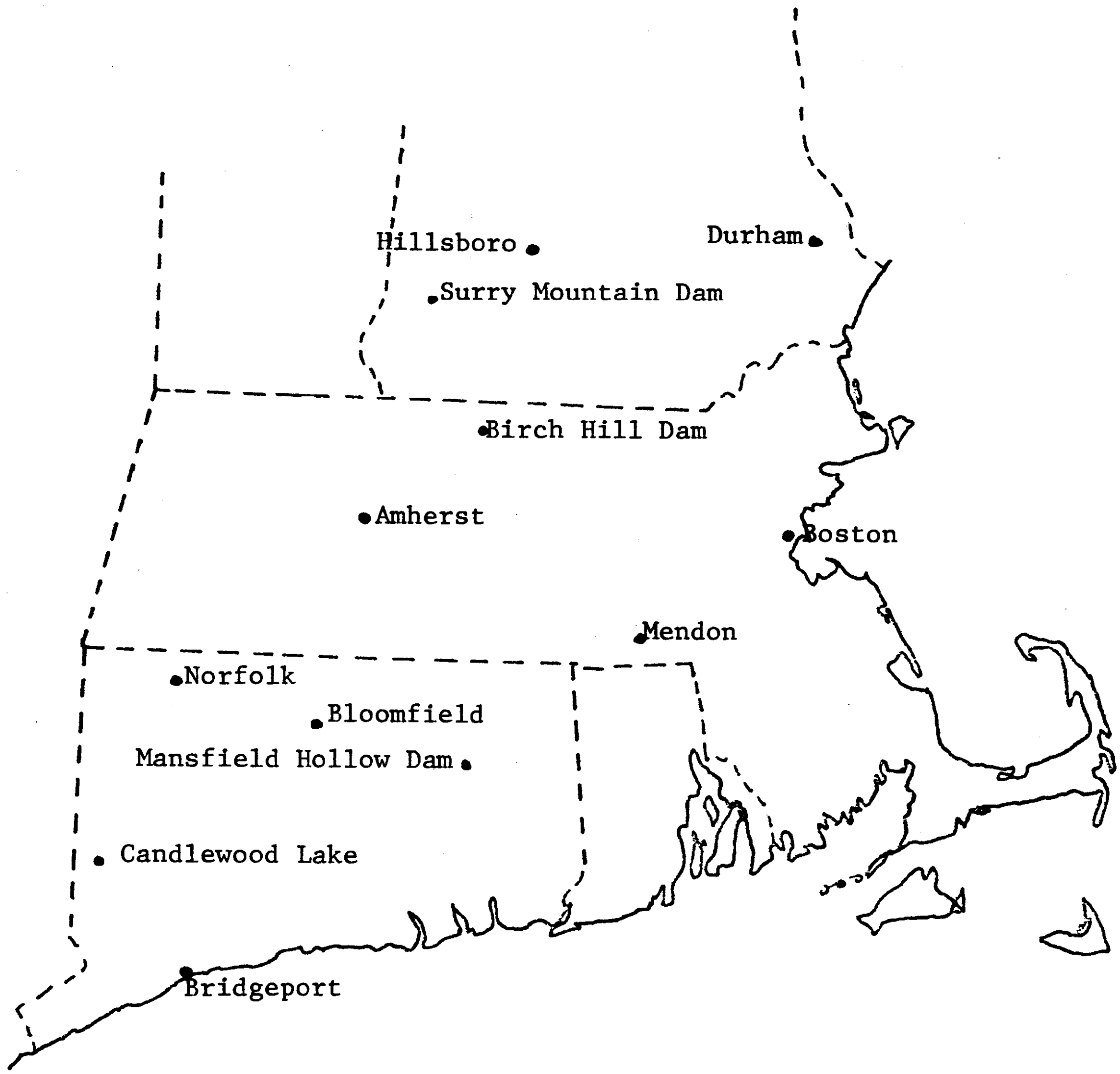


Figure 3.3: Location of Rainfall Gages

Station Name: Surry Mountain Dam, N.H.

Station No. 8539

Month	# Storms/Yrs.	T_S (hrs)		T_R (hrs)		Depth (in.)		Int. (in/hrs)	
		Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
1	226/22	66.89	80.87	7.32	6.37	.232	.333	.0310	.0736
2	204/22	64.22	58.43	7.44	5.91	.260	.319	.0463	.1401
3	234/22	60.91	71.36	6.80	6.10	.244	.304	.0481	.1390
4	244/22	60.40	73.34	6.66	5.94	.255	.348	.0510	.1762
5	259/22	56.00	63.55	5.95	5.85	.262	.343	.0424	.0427
6	239/23	63.70	78.49	4.74	4.88	.282	.359	.0699	.1188
7	233/22	69.10	76.65	3.98	3.48	.271	.345	.0738	.0861
8	209/22	68.54	79.95	4.52	4.77	.342	.473	.0786	.0983
9	171/22	84.24	102.72	5.73	5.27	.367	.562	.0567	.0523
10	157/22	104.92	122.85	6.78	6.18	.357	.525	.0462	.0663
11	233/22	65.62	75.65	7.14	6.33	.342	.457	.0472	.0750
12	225/22	65.08	62.91	7.85	7.49	.282	.399	.0431	.1530

Table 3.9 Rainfall Statistics for Surry Mountain, N.H.

Station Name: Hillsboro 2W, N.H.

Station No. 4062

Month	# Storms/Yrs.	T_s (hrs)		T_R (hrs)		Depth (in.)		Int. (in/hrs)	
		Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
1	152/21	106.55	143.94	8.32	8.51	.395	.498	.0917	.2175
2	156/22	110/23	222.83	7.35	6.76	.416	.468	.1377	.3032
3	180/22	82.88	95.04	7.60	6.32	.389	.464	.0832	.2174
4	196/22	78.24	91.87	7.23	6.53	.3674	.4458	.0663	.1442
5	202/21	70.47	77.39	6.06	5.88	.299	.365	.0621	.1196
6	199/23	72.77	94.34	4.74	4.68	.311	.371	.0877	.1480
7	180/22	85.19	95.92	3.87	3.58	.311	.361	.1052	.1361
8	192/22	84.59	109.80	4.54	4.44	.364	.607	.1025	.2448
9	150/22	100.25	131.09	5.63	4.68	.418	.667	.0733	.0895
10	123/22	126.49	157.13	7.70	7.92	.445	.646	.0638	.1025
11	182/21	87.48	135.70	7.61	6.70	.468	.641	.0695	.1350
12	139/21	93.69	86.62	8.45	7.71	.494	.616	.1094	.2697

Table 3.10 Rainfall Statistics for Hillsboro 2W, N.H.

Station Name: Durham, N.H.

Station No. 2174

Month	# Storms/Yrs.	T_S (hrs)		T_R (hrs)		Depth (in.)		Int. (in/hrs)	
		Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
1	168/21	89.92	106.74	6.85	6.32	.338	.408	.1013	.2407
2	164/21	88.82	120.25	6.42	5.94	.381	.443	.1179	.2249
3	200/22	82.83	126.03	6.78	6.23	.405	.497	.1003	.2110
4	194/21	71.25	89.37	7.40	6.75	.349	.468	.0539	.1210
5	228/23	67.19	87.02	5.41	5.28	.304	.472	.0638	.1123
6	188/22	81.75	99.14	4.41	5.36	.287	.361	.0929	.1846
7	174/21	89.25	115.34	3.91	3.53	.322	.394	.1109	.2045
8	174/21	83.62	109.42	4.53	4.28	.349	.463	.0911	.1133
9	146/21	98.29	138.33	5.97	5.73	.404	.657	.0603	.0561
10	149/22	110.72	126.02	7.63	7.48	.436	.762	.0477	.0295
11	218/22	66.99	80.39	6.94	6.95	.462	.667	.0654	.0726
12	163/21	87.77	88.05	7.10	7.02	.459	.683	.1297	.3127

Table 3.11 Rainfall Statistics for Durham, N.H.

Station Name: Norfolk, Conn

Station No. 5445

Month	# Storms/Yrs.	T_S (hrs)		T_R (hrs)		Depth (in.)		Int. (in/hrs)	
		Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
1	230/22	67.30	71.46	8.64	8.73	.358	.773	.0579	.1627
2	222/22	59.71	57.66	8.04	7.21	.335	.436	.0549	.1449
3	249/22	59.51	60.13	7.34	7.10	.371	.477	.1032	.2727
4	249/21	53.82	58.57	7.57	7.01	.363	.485	.0506	.1020
5	280/22	51.48	60.41	5.65	4.79	.247	.279	.0455	.0442
6	243/23	63.50	71.91	4.80	4.70	.353	.495	.0750	.0822
7	219/22	70.09	72.27	4.14	3.41	.322	.429	.0837	.1023
8	214/22	67.89	77.47	4.65	5.80	.428	1.070	.0837	.1151
9	163/22	97.56	119.67	6.35	6.07	.477	.735	.0770	.1171
10	151/22	96.37	115.10	8.03	9.08	.536	.978	.0633	.0752
11	227/22	71.67	97.35	7.48	7.59	.450	.644	.0895	.2368
12	215/22	64.20	73.50	8.57	6.82	.4640	.8513	.0865	.2628

Table 3.12 Rainfall Statistics for Norfolk, Conn.

Station Name: Manfield Hollow Dam, Conn.

Station No. 4488

Month	# Storms/Yrs.	T_S (hrs)		T_R (hrs)		Depth (in.)		Int. (in/hrs)	
		Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
1	150/18	81.25	78.23	6.89	5.69	.317	.442	.0736	.2384
2	163/19	77.63	85.41	6.58	5.24	.305	.399	.0526	.0930
3	182/10	70.23	69.74	7.30	6.13	.364	.481	.0728	.2414
4	185/19	68.28	86.06	6.75	6.35	.359	.536	.0477	.1042
5	205/19	65.41	81.30	5.60	5.11	.272	.386	.046	.045
6	217/19	60.99	72.82	3.82	3.79	.217	.355	.0585	.0768
7	154/17	71.92	74.07	4.24	4.17	.398	.668	.0845	.1035
8	167/18	79.39	103.03	4.82	5.07	.393	.657	.0818	.0915
9	132/18	98.42	111.43	5.55	5.98	.427	.703	.0669	.0632
10	115/18	103.21	120.83	7.05	7.12	.498	.739	.0656	.0816
11	159/18	80.37	105.35	7.27	6.51	.399	.534	.0487	.0731
12	178/18	71.92	90.49	7.60	6.64	.378	.525	.0462	.0924

Table 3.13 Rainfall Statistics for Mansfield Hollow Dam, Conn.

Station Name: Candlewood Lake, Conn.

Station No. 1093

Month	# Storms/Yrs.	T_S (hrs)		T_R (hrs)		Depth (in.)		Int. (in/hrs)	
		Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
1	208/22	82.67	138.54	7.64	8.48	.3073	.4804	.0604	.2037
2	172/22	82.68	135.12	7.56	7.25	.3589	.4479	.0930	.2411
3	194/22	83.21	98.36	7.40	7.59	.392	.531	.1298	.3905
4	226/22	65.92	88.06	6.84	6.60	.317	.474	.0481	.0956
5	245/22	59.18	83.86	6.48	6.21	.291	.430	.0562	.1674
6	216/23	73.29	100.81	4.90	4.93	.291	.471	.0841	.2034
7	204/22	76.15	98.11	4.66	4.64	.333	.508	.0850	.1838
8	199/22	75.07	95.86	6.07	6.76	.430	.798	.0711	.1036
9	144/21	97.50	118.87	5.55	5.88	.365	.585	.0670	.1109
10	140/21	106.90	134.05	7.26	8.05	.497	1.009	.0559	.0648
11	173/20	85.03	107.00	7.39	7.78	.430	.625	.0729	.1551
12	196/22	85.02	146.62	8.33	7.76	.408	.555	.0658	.1587

Table 3.14 Rainfall Statistics for Candlewood Lake, Conn.

Station Name: Bridgeport, Conn.

Station No. 806

Month	# Storms/Yrs.	T_S (hrs)		T_R (hrs)		Depth (in.)		Int. (in/hrs)	
		Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
1	223/22	66.21	74.73	6.50	6.79	.248	.375	.0295	.0266
2	182/22	71.86	67.56	7.63	6.81	.322	.443	.0322	.0264
3	232/22	65.38	72.96	6.83	6.61	.312	.484	.0344	.0303
4	244/22	62.40	68.87	6.03	6.26	.304	.447	.0405	.0415
5	243/22	60.50	69.30	4.88	4.99	.279	.444	.0510	.0644
6	174/20	77.95	94.28	4.32	4.62	.246	.329	.0534	.0491
7	197/22	77.99	102.34	3.62	4.04	.312	.428	.0911	.1216
8	205/22	75.23	87.66	4.20	4.20	.382	.603	.0832	.0986
9	163/22	96.62	121.24	4.76	4.82	.330	.594	.0628	.0719
10	113/21	116.73	123.10	6.53	7.02	.503	.784	.0702	.0993
11	195/22	83.54	114.47	6.42	6.22	.377	.560	.0474	.0454
12	212/22	73.94	91.16	7.32	6.25	.338	.508	.0354	.0326

Table 3.15 Rainfall Statistics for Bridgeport, Conn.

Station Name: Bloomfield, Conn

Station No. 0634

Month	# Storms/Yrs.	T_S (hrs)		T_R (hrs)		Depth (in.)		Int. (in/hrs)	
		Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
1	115/18	118.39	138.12	8.54	7.64	.466	.569	.1037	.2307
2	103/16	109.89	156.80	8.19	9.13	.398	.412	.1085	.2536
3	115/17	94.72	99.45	8.19	6.09	.473	.516	.0999	.2214
4	150/18	95.62	179.32	7.47	6.41	.394	.492	.0782	.2796
5	156/18	88.29	175.52	6.65	7.61	.301	.365	.0660	.1651
6	139/19	103.22	140.53	4.96	4.72	.350	.487	.0992	.2152
7	117/19	112.18	127.55	4.11	3.90	.342	.379	.1247	.2263
8	140/20	132.66	420.30	6.47	8.27	.416	.871	.0808	.1798
9	103/20	188.75	490.61	5.80	6.32	.512	.774	.1606	.5340
10	86/18	136.06	152.20	7.44	9.30	.581	1.056	.0788	.1153
11	131/19	112.52	156.03	7.47	6.79	.459	.611	.0963	.2354
12	123/20	114.79	204.73	8.06	6.64	.438	.485	.1132	.3031

Table 3.16 Rainfall Statistics for Bloomfield, Conn

Station Name: Mendon, Mass.

Station No. 4667

Month	# Storms/Yrs.	T_S (hrs)		T_R (hrs)		Depth (in.)		Int. (in/hrs)	
		Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
1	152/19	84.58	83.50	8.15	7.16	.420	.535	.0996	.2409
2	158/20	79.11	70.26	6.21	5.01	.408	.496	.1618	.3798
3	167/20	78.96	73.02	6.26	5.87	.429	.520	.1690	.3727
4	188/19	65.99	83.88	6.84	6.41	.376	.536	.0898	.2912
5	173/20	95.72	285.11	6.71	6.53	.329	.378	.0708	.1584
6	199/21	73.53	90.43	4.42	3.64	.259	.317	.0598	.0587
7	143/19	85.09	87.73	4.48	4.53	.362	.454	.0915	.1335
8	170/21	103.06	192.18	4.67	4.78	.433	1.023	.1029	.1874
9	131/19	108.54	217.15	6.26	6.11	.547	.877	.0964	.2839
10	121/21	121.08	140.90	7.20	7.53	.473	.708	.0784	.1196
11	152/21	94.05	121.72	7.31	6.34	.471	.598	.0945	.2427
12	123/20	122.10	135.04	8.45	7.29	.606	.688	.1728	.4628

Table 3.17 Rainfall Statistics for Mendon, Mass.

Station Name: Boston, Mass.

Station No. 0770

Month	# Storms/Yrs.	T_S (hrs)		T_R (hrs)		Depth (in.)		Int. (in/hrs)	
		Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
1	226/22	67.83	79.65	8.61	7.85	.346	.475	.0300	.0266
2	241/22	54.14	54.68	7.38	7.09	.340	.525	.0316	.0316
3	252/22	56.95	66.89	7.49	7.43	.359	.578	.0338	.0290
4	263/22	54.14	65.45	6.54	6.86	.303	.490	.0349	.0330
5	279/23	54.78	66.81	5.48	5.97	.272	.554	.0373	.0392
6	266/23	60.00	72.56	4.06	4.64	.246	.385	.0573	.0690
7	218/22	67.77	78.15	3.62	3.87	.251	.386	.0649	.0863
8	236/22	67.74	83.25	3.80	3.98	.311	.668	.0706	.0903
9	186/21	74.88	90.65	5.13	6.05	.371	.760	.0510	.0562
10	187/22	82.36	99.70	6.22	6.83	.373	.677	.0456	.0483
11	245/22	62.44	78.02	6.62	6.25	.400	.609	.0428	.0428
12	202/22	71.05	73.52	8.49	8.19	.440	.630	.0370	.0364

Table 3.18 Rainfall Statistics for Boston, Mass.

Station Name: Birch Hill Dam, Mass.

Station No. 0666

Month	# Storms/Yrs.	T_S (hrs)		T_R (hrs)		Depth (in.)		Int. (in/hrs)	
		Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
1	223/22	67.52	73.30	7.54	6.65	.276	.384	.0506	.1322
2	189/22	74.47	72.33	7.31	6.98	.304	.382	.0928	.2775
3	216/22	67.14	71.30	7.02	6.55	.297	.378	.0650	.1479
4	242/22	60.27	74.28	6.57	6.38	.276	.371	.0472	.0928
5	235/22	62.11	70.83	6.00	5.32	.288	.332	.0485	.0463
6	224/23	66.31	85.14	4.99	5.34	.314	.470	.0626	.0697
7	200/22	80.23	87.71	3.64	3.07	.277	.344	.0911	.1335
8	202/22	73.26	88.92	4.22	4.66	.318	.586	.0698	.0812
9	167/22	92.05	109.56	5.37	4.79	.362	.518	.0628	.0668
10	141/22	109.44	123.04	6.70	6.53	.394	.517	.0631	.1010
11	219/22	70.81	94.10	7.02	6.79	.360	.483	.0635	.1481
12	203/22	74.23	78.16	7.58	6.79	.337	.434	.0670	.1678

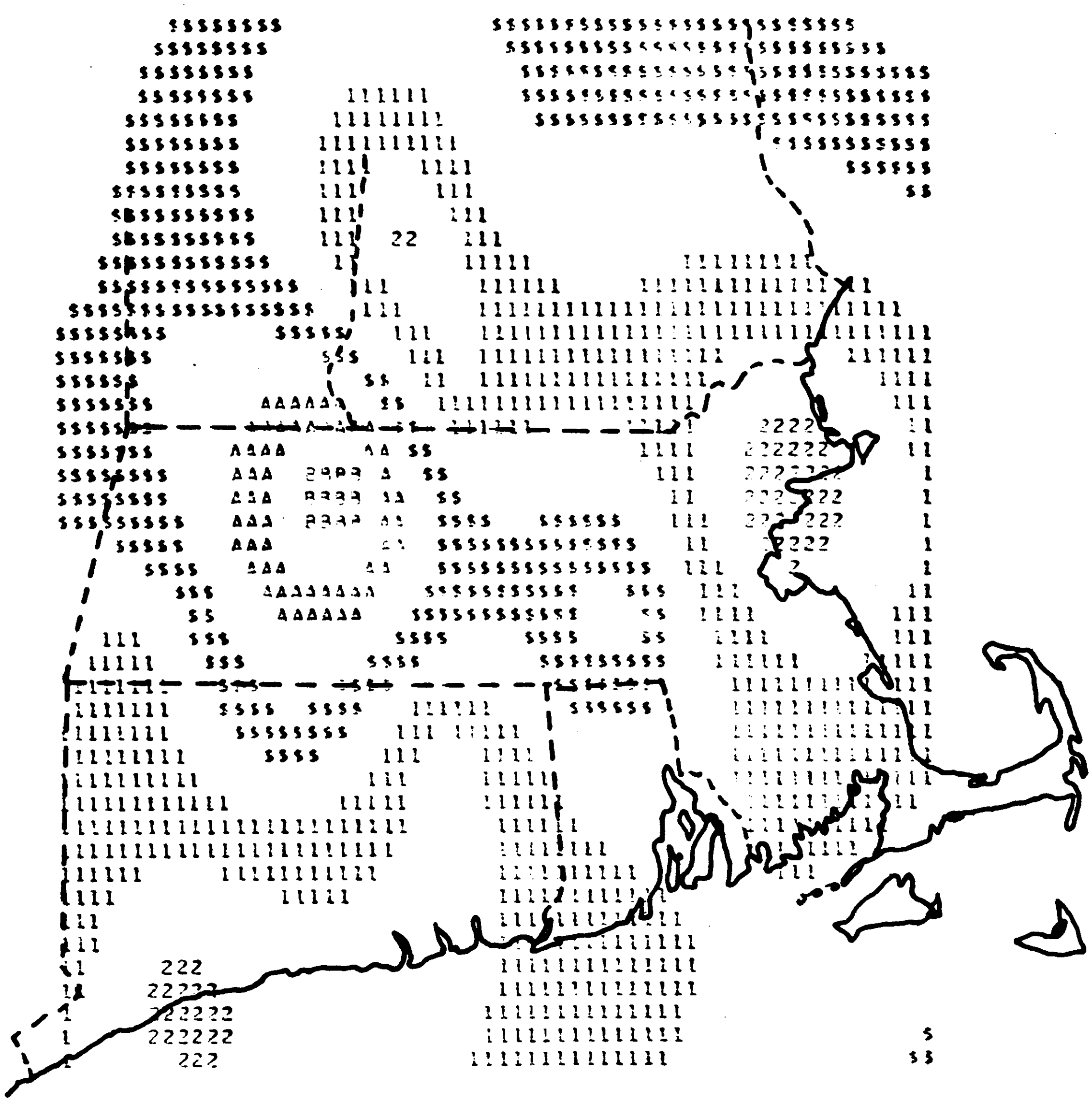
Table 3.19 Rainfall Statistics for Brick Hill Dam, Mass.

Station Name: Amherst, Mass.

Station No. 0120

Month	# Storms/Yrs.	T_S (hrs)		T_R (hrs)		Depth (in.)		Int. (in/hrs)	
		Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
1	144/15	78.50	92.45	4.81	5.31	.298	.389	.1344	.2688
2	123/15	161.26	919.52	5.08	7.32	.305	.365	.1414	.2687
3	149/16	73.49	69.30	5.30	6.70	.351	.435	.1462	.2750
4	174/16	65.49	77.32	5.00	5.71	.315	.407	.1091	.2203
5	199/17	59.24	66.75	4.33	4.67	.287	.332	.0975	.1500
6	191/17	58.16	77.20	3.11	3.29	.294	.474	.1205	.3158
7	151/16	75.33	78.36	2.70	3.17	.314	.377	.1686	.2387
8	159/16	69.74	79.89	3.01	5.33	.398	.803	.1966	.3203
9	137/16	85.53	113.62	4.09	4.69	.398	.524	.2025	.4613
10	127/16	93.09	115.63	4.25	4.78	.403	.558	.1808	.4632
11	152/17	136.19	668.76	5.67	6.09	.406	.530	.1204	.2368
12	123/15	80.40	85.38	5.68	6.63	.370	.490	.1086	.2066

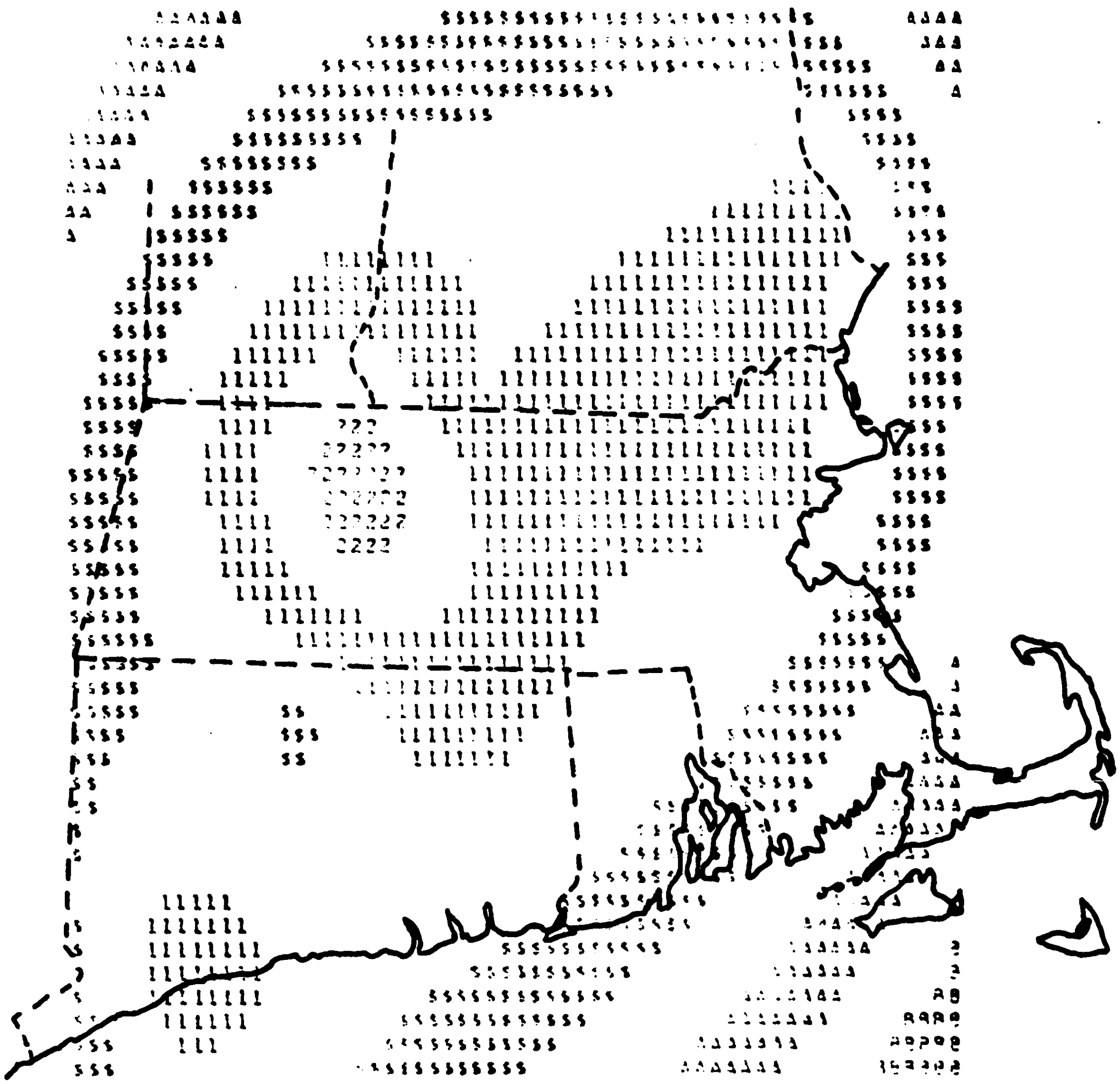
Table 3.20 Rainfall Statistics for Amherst, Mass.



REFERENCE CONTOUR ($) = 0.0979

CONTOUR INTERVAL = 0.0105

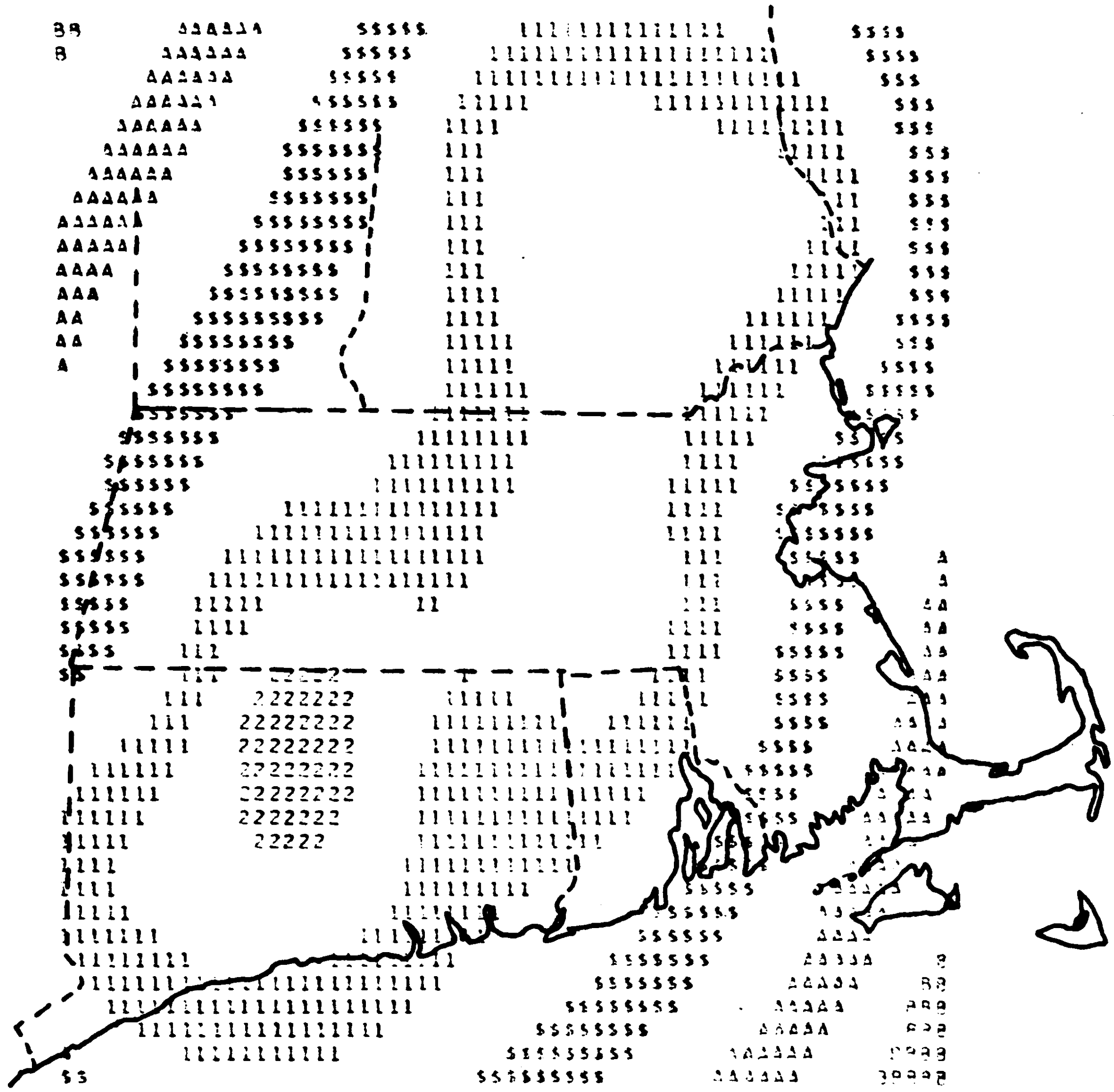
Figure 3.4: Multiquadric Surface for Average Intensity for Southern New England



REFERENCE CONTOUR ($) = 7.3085

CONTOUR INTERVAL = 0.5742

Figure 3.5: Multiquadric Surface for Average Storm Duration, for Southern New England



REFERENCE CONTOUR ($) = 135.1102

CONTOUR INTERVAL = 10.6900

Figure 3.6: Multiquadric Surface for Average Number of Storms per Year, for Southern New England

Basin	β	λ	Θ
117	11.8	.164	101
118	11.9	.167	101
135	12.5	.159	110
137	12.5	.159	110
139	11.5	.154	93
143	11.1	.151	95
151	12.3	.161	95
157	16.7	.164	118
169	20.0	.167	110
171	20.0	.167	110
177	12.5	.159	120
181	11.7	.167	105
195	18.2	.167	110
198	16.4	.164	110
200	16.4	.164	110
204	10.0	.154	100
209	14.3	.161	100
249	16.7	.161	120
251	16.7	.161	125
254	17.2	.161	120
258	18.2	.161	120
264	15.4	.151	115
274	14.3	.169	115
276	14.3	.172	115
285	7.04	.223	115
291	11.1	.182	115
295	9.1	.182	115
307	9.1	.167	108
313	10.0	.147	110
317	10.0	.147	90
322	11.1	.147	90
325	16.7	.167	105
328	19.6	.172	105
332	14.3	.151	120
324	12.5	.151	105
352	10.0	.143	105
187	9.4	.159	95
189	11.1	.167	105

Table 3.21 Best Rainfall Parameters for the 36 Southern New England Basins

Station No.	Predicted Mean	Historical Mean	Ratio Y_m
117	405	312	0.7703
118	5212	2262	0.4340
135	2023	1253	0.6192
137	1451	633	0.4363
139	1822	1361	0.7469
143	4641	2217	0.4777
151	5287	3611	0.6830
157	2291	3384	1.4768
169	400	224	0.5604
171	1939	1112	0.5735
177	1060	356	0.3357
181	1288	481	0.3735
195	1306	835	0.6392
198	2549	3179	1.2471
200	685	1129	1.6489
204	5498	2161	0.3931
209	1984	1538	0.7750
249	5403	5436	1.0062
251	2366	4029	1.7026
254	1853	2101	1.1340
258	821	1813	2.2084
264	902	1257	1.3936
274	344	322	0.9370
276	7961	4271	0.5365
285	2936	2254	0.7678
291	1548	941	0.6080
295	21961	7233	0.3294
307	17600	16890	0.9596
313	7709	8742	1.1340
317	1680	2003	1.1925
322	2318	2970	1.2815
325	1930	3097	1.6050
328	239	495	2.0675
332	1557	2024	1.2996
343	3847	4709	1.2242
353	8837	13478	1.5251

Table 3.22

Ratio of the Means, Best Rainfall Parameters

Station No.	Predicted Variance	Historical Variance	Ratio Y_v
117	19834	23000	1.1596
118	4145262	1238000	0.2987
135	550174	343000	0.6234
137	265847	50000	0.1881
139	428421	2362000	5.5133
143	3097219	1053000	0.3400
151	4694603	2798000	0.5960
157	747257	11238000	15.0390
169	23068	9000	0.3901
171	693327	250000	0.3606
177	148412	59000	0.3975
181	242862	39000	0.1606
195	290080	71000	0.2448
198	983848	20118992	20.4493
200	59379	1153000	19.4178
204	4387153	8418000	1.9188
209	635916	695000	1.0929
249	4390080	6990000	1.5922
251	694179	3568000	5.1399
254	429380	2062000	4.8023
258	275116	961000	3.4931
264	103437	1267000	12.2490
274	13504	64000	4.7394
276	11127653	6264000	0.5629
285	1069503	1599000	1.4951
291	295516	1405000	4.7544
295	84106656	81024992	0.9634
307	46531632	173538000	3.7295
313	8107109	55468992	6.8420
317	381541	4759000	12.4731
322	824797	8911000	10.8039
325	719989	5000000	6.9446
328	7188	87000	12.1041
332	297353	2199000	7.3952
343	2039205	65184992	31.9659
353	10402987	322579968	31.0084

Table 3.23

Ratio of the Variances, Best Rainfall Parameters

$$E[Y_v] = 6.42$$
$$V[Y_v] = 66.74 \tag{3.31}$$

Again, the fraction of area that contributes to direct runoff, FA_r, is adjusted so that the $E[Y_m]$ is unbiased. This value of FA_r is .168. Even with the best estimates for the rainfall parameters, the prediction of the variance is still too low, causing $E[Y_v]$ to be highly biased. The low prediction of the variance implies that the predicted coefficient of variation is too low. The coefficient of variation can be calculated from Equations (3.22) and (3.23) as

$$C_v = \frac{1}{\frac{\sqrt{6}}{\pi}\,\ln(nI_o) + \frac{\sqrt{6}}{\pi}\,\gamma} \tag{3.32}$$

or substituting values for π and γ and simplifying

$$C_v = \frac{1}{.78\,\ln(nI_o) + .45} \tag{3.33}$$

for $I_o = e^{-2\sigma}\sigma^{-\sigma+1}\Gamma(\sigma)$

σ is a function of both rainfall and basin parameters. Its value ranges from .3 to .65 which implies that I_o has values that fall into the range of .7 to .35 with .6 being a typical value. n is the number of excess rainfall events and is estimated from

$$n = \Phi_1 \cdot \Phi_2 \cdot \Theta \tag{3.34}$$

where

Φ_1 is the ratio of annual runoff to point precipitation

Φ_2 is the ratio of direct runoff to annual runoff and

Θ is the number of independent rainfall events.

Eagleson takes as typical values for New England $\Phi_1 = .50$ and $\Phi_2 = .58$, which give just under 30% of the rainfall events as being excess events. Letting $\Theta \simeq 100$ events per year, average for New England, gives n = 30. Substituting into 3.33 gives

$$C_v = \frac{1}{3.1 + \ell n(I_o)} \tag{3.35}$$

Since I_o ranges from .35 to .7, C_v will range from .44 to .35. This range is much lower than the C_v observed in the historical records. The sample statistics for the coefficient of variation for the 36 New England basins are:

$$\begin{aligned} \text{mean} &= .770 \\ \text{variance} &= .126 \end{aligned} \tag{3.36}$$

The predicted moments, obtained from the predicted mean and standard deviation, are:

$$\begin{aligned} \text{mean} &= .369 \\ \text{variance} &= .0015 \end{aligned} \tag{3.37}$$

The poor performance of Eagleson's formula can be partially related back to (3.33) and the assumptions involved in the treatment of the random variables in the model. An important parameter in the model is n, the number of excess rainfall events. n is not known with certainty and thus is a random variable. Eagleson uses an expected value of n from Equation (3.34), but Θ, Φ_1 and Φ_2 are all random variables and the expected value of n can be calculated from

$$
\begin{aligned}
E[n] &= E[\Phi_1] \cdot E[\Phi_2] \cdot E[\Theta] \\
&+ \rho_{\Phi_1\Phi_2} \cdot V[\Phi_1]^{1/2} \cdot V[\Phi_2]^{1/2} \\
&+ \rho_{\Phi_1}\Theta \cdot V[\Phi_1]^{1/2} \cdot V[\Theta]^{1/2} \\
&+ \rho_{\Phi_2}\Theta \cdot V[\Phi_2]^{1/2} \cdot V[\Theta]^{1/2}
\end{aligned} \tag{3.38}
$$

and only equals (3.34) under the assumption that Φ_1, Φ_2 and Θ are all independent. It would be suspected that Φ_1 and Φ_2 would be highly dependent since both are ratios concerning runoff, direct runoff and precipitation.

In the analysis, there are many variables whose values are uncertain but constant. Also, there are variables whose values are uncertain and change from storm to storm - for example, A_r, the area that produces runoff. Essentially, Eagleson handles this uncertainty by inserting the expected value into the function. This procedure implies that the expected value of a function is equal to the function of the expected value, $E[g(\underline{x})] = g(E]\underline{x}])$. This relationship is not true, except as a first order approximation.

There is still the problem that the underlying analysis may not represent the process. The rainfall is modeled by the average excess intensity, $\bar{i}_e$, excess storm duration, tr_e and a rectangular storm interior. The assumption that the probability density function for $\bar{i}_e$ and tr_e are exponential with no seasonal variations does not seem to capture the observed behavior. Also, the runoff process may not accurately represent the behavior of the New England basins. The assumption that the time of concentration of the catchment, t_e, is less than the time of concentration of the streams, t_s, must be seriously questioned, especially for the smaller New England basins.

It is felt that Eagleson's approach is an important conceptual contribution to the dynamics of flood frequencies, but, for general application in obtaining prior information, the results are limited.

Eagleson's procedure may still be used with other regional information to provide prior information on the mean and variance. If it is assumed that there is a regional coefficient of variation and Eagleson's prediction is used for the mean, then an estimate of the moments for the variance can be found. The relationship between the standard deviation σ, mean m, and coefficient variation C_v, is

$$\sigma^2 = C_v^2 \cdot m^2 \qquad (3.39)$$

From the first-order analysis, estimates for the mean and variance of σ^2 are

$$E[\sigma^2] \doteq C_v^{\,2} \cdot m^2 + C_v^{\,2}\, V[m] + m^2 \cdot V[C_v] + + 4 \cdot C_v \cdot m \cdot \rho_{C_v m} \cdot V[m]^{1/2} \cdot V[C_v]^{1/2} \tag{3.40}$$

$$V[\sigma^2] \doteq 4C_v^{\,4}\, m^2\, V[m] + 4C_v^{\,2} \cdot m^4\, V[C_v] + 8C_v^{\,3}\, M^3\, \rho_{C_v m}\, V[m]^{1/2} \cdot V[C_v]^{1/2} \tag{3.41}$$

For the 36 New England basins, the mean and variance for Cu were given in Equation (3.36). The sample correlation coefficient between C_v and m for the 36 basins is .31. The moments of the mean are found for the 'ungaged' basin by applying the predicted mean and the moments of the ratio of the means.

This was done for the Blackstone River near Woonsocket, R.I. The predicted mean was 5600 cfs, which implies the moments

$$E[\mu] = 5575 \text{ cfs} \qquad V[\mu] = 7369600 \text{ cfs}^2 \tag{3.42}$$

Applying (3.36), (3.40), (3.41) and (3.42) gives

$$E[\sigma^2] = 3.184 \times 10^7 \text{ cfs}^{\,2} \qquad V[\sigma^2] = 7.9975 \times 10^{14} \text{ cfs}^4 \tag{3.43}$$

Thus, the prior information on the mean and variance of the annual series for the Blackstone River is given by (3.42) and (3.43) and implies an equivalent prior sample size n' of 4 where

$$n' \quad \frac{E[\sigma^2]}{V[\mu]} \tag{3.44}$$

This is about the same amount of information as was found in the regression equation.

3.4.4 Summary

The analytical procedures of Eagleson were unable on their own to provide prior information on the mean and variance of annual floods. This lack of information can be attributed to the numerous assumptions that must be made if an analytical solution is to be obtained. The analytical procedures of Eagleson's seem to give a large bias to the variance for the example basins studied, which, in turn, implies a high amount of risk in applying the prior information. When this problem is encountered, a possible procedure to obtain prior information is the application of Eagleson's approach in conjunction with a regional estimation of the coefficient of variation. This latter method provides information equivalent to about 4 years of data.

3.5 Conclusion

This chapter studies two sources of prior information that researchers have used to estimate the occurrances of floods - the multivariate regression and an analytical flood frequency formula (after Eagleson, 1972).

The multivariate regression is an empirical technique that fits a linear function to data from near-by or similar basins. This procedure will out perform many empirical flood formulae found in the literature, because regression is a minimum variance procedure.

Eagleson's analytical flood frequency formula predicts the occurrance of flood magnitudes by considering the overland flow dynamics of excess rainfall and the pdf of the magnitude of the rainfall. For the region studied, this procedure did not provide any useful information. When the analytical formula is used in conjunction with a regional coefficient of variation, limited prior information is obtained.

For the region studied, the empirical and analytical procedures provided some prior information. This information is equivalent to 4 to 7 years of data, and may be significant when only very short historical records exist. More research should be done in the area of estimating prior information. One possibility is the use of rainfall-runoff simulation models. Often rainfall data has a longer record than streamflow data, and extensive research is needed to determine if such models can be used to extend the streamflow record.

CHAPTER 4

Bayesian Distribution Theory

4.1 Introduction

The analysis of estimating flood exceedance probabilities, the probability that a flood, with a magnitude greater than q, will occur during some time interval, has always been a source of controversy in hydrology. The procedure traditionally followed has been to:

1. Observe a historical record of flood events
2. Pick a generating process or probability density function that seems "reasonable"
3. Estimate the parameters of this underlying process from the historical record, and
4. Make inferences about the occurrence of future flood events.

This procedure will be refered to as the estimation of floods by the application of distribution theory. When historical records are not available, inferences about the occurrence of floods are made using procedures such as those described and analyzed in Chapter 3.

Much of the controversy about the distribution approach has centered on which underlying process or probability density function (pdf) is the appropriate one to use and what is the 'best' approach to estimating the parameters of the chosen distribution. In Chapter 2, it was argued that when parameters are uncertain and thus random variables, point estimation will underestimate the uncertainty in the pdf of the flood events. Only through the application of Bayesian distribution theory, is it possible to fully account for parameter uncertainty. The analysis leads to the Bayesian pdf. This Bayesian distribution of the magnitude of flood events, $\tilde{f}(q)$, is, from Equation (2.6),

$$\tilde{f}(q) = \int_{\underline{\Theta}} f(q|\underline{\Theta}) \cdot f''(\underline{\Theta})d\underline{\Theta} \tag{4.1}$$

where

$f(q)$ is the Bayesian pdf of q

$f(q|\underline{\Theta})$ is the 'modelled' pdf of q which is conditional upon the set of uncertain parameters $\underline{\Theta}$

$f''(\underline{\Theta})$ is the (posterior) pdf of the parameter set $\underline{\Theta}$

The traditional approach to flood analysis uses the historical record to find point estimates for the parameters of $\underline{\Theta}$. Bayesian analysis uses the historical record in conjunction with other sources of information, as discussed in Chapter 3, to find a pdf for the parameter set $\underline{\Theta}$. The pdf for $\underline{\Theta}$ is called the prior pdf if it is evaluated prior to obtaining more data, and it is called the posterior pdf if it is evaluated after obtaining the data.

The posterior pdf for $\underline{\Theta}$ is found through the application of Bayes theorem and is, from Equation (2.1)

$$f''(\underline{\Theta}) = f''(\underline{\Theta}|\underline{Q}, I_o) \propto L(\underline{Q}|\underline{\Theta}) \cdot f'(\underline{\Theta}) \tag{4.2}$$

where

$f''(\underline{\Theta}) = f''(\underline{\Theta}|\underline{Q},I_o)$ is the posterior PDF of $\underline{\Theta}$ conditional upon the current observed sample $\underline{Q}$ and initial information I_o

$L(\underline{Q}|\underline{\Theta})$ is the likelihood function of the sample, $\underline{Q}$

$f'(\underline{\Theta}) = f'(\underline{\Theta}|I_o)$ is the prior PDF of the parameter set $\underline{\Theta}$ conditional upon the initial information, I_o.

The evaluation of $f''(\underline{\Theta})$ does not depend upon the functional form of $f'(\underline{\Theta})$. If $f''(\underline{\Theta})$ is restricted to a particular form defined as the natural conjugate of $L(\underline{Q}|\underline{\Theta})$, then the posterior pdf of $\underline{\Theta}$, $f''(\underline{\Theta})$, will be of the same functional form as $f'(\underline{\Theta})$. By restricting prior pdf to natural conjugate forms, it is possible to evaluate both $f''(\underline{\Theta})$ and $\tilde{f}(q)$ analytically. This is an extremely attractive consideration since this analysis is concerned with the methodology of Bayesian procedures. The extensive numerical analysis needed to solve for $\tilde{f}(q)$, using priors which are not natural conjugate forms, is not considered in this thesis.

Four alternative models are analyzed in this chapter. Three of the models consider the annual series of flood events which is made up of the largest flood event in each year. The three models are analyzed, conditional upon the assumption that the annual series was generated by 1) a normal process, 2) a log-normal process or 3) a gamma-1 process.

The fourth model considers the partial duration series of flood peaks above a 'base' level. This model is defined as an exceedance model. The base level is chosen so that the occurrence of events greater than the base value can be described by a Poisson process and the magnitude of events larger than the base level can be described by an exponential pdf.

This chapter will not address directly the controversy of which pdf best describes the generating process of flood events. This is left to Chapter 7 when model selection is considered. This chapter will analyze, in a Bayesian framework, alternative probability models of the occurrence of flood magnitudes. The flood exceedance probability from each model will be calculated for the Blackstone River at Woonsocket, Rhode Island and compared. Chapter 6 uses each of the probability models in a decision problem where the decision act is concerned with the design capacity of a local flood control system consisting of channelization, dikes and flood walls. The decision example is for Woonsocket, Rhode Island and will utilize the results of the probability models developed in this chapter.

4.2 Bayesian Inference from a Normal Process

4.2.1 Prior and Posterior Distribution for Mean μ and Precision h for the Normal Process

Assume that the series of annual floods comes from a normal process with unknown mean μ, and variance σ^2. Define the precision, h, as being $1/\sigma^2$ and let $\underline{Q}$ be the record of the observed annual floods.

The 'modelled' pdf of the annual flood q is distributed $N(\mu,h)$ which is

$$f(q|\mu,h) \propto h^{1/2} \exp[\frac{h}{2}(q-\mu)^2] \quad (4.3)$$

The likelihood function of the set of independent observations $\underset{\sim}{Q}$ is defined as

$$L[\underset{\sim}{Q}|\mu,h] = \prod_{\text{all } q_i} f(q|\mu,h) \quad (4.4)$$

which, for the normal process, can be shown to be (Raiffa and Schlaifer, 1961)

$$L[\underset{\sim}{Q}|h,\mu] \propto h^{1/2} \cdot \exp[-\frac{h}{2} n(m-\mu)^2] \cdot \exp[-\frac{h}{2} \nu v] \cdot h^{\;/2} \quad (4.5)$$

where

n = number of observations in the sample $\underset{\sim}{Q}$

$m = \frac{1}{n} \sum_{i=1}^{n} q_i$

$v = \frac{1}{n-1} \sum_{i=1}^{n} (q_i-m)^2 \quad (\equiv 0 \text{ for } n=1)$

$\nu = n-1$

It can be observed that the likelihood function is composed of two parts. The first part is the kernel of a normal density function and the second part is the kernel of a gamma-2 density function. Thus, the natural conjugate will be a normal-gamma distribution, $f'_{N\gamma}(\mu,h|m',v',n',\nu')$, of the form

$$f'_{N\gamma}(\mu,h|m',v',n',\nu') \propto h^{1/2} \cdot \exp[-\tfrac{1}{2}\, hn'(\mu-m')^2 \cdot$$
$$\exp[-\tfrac{1}{2}\, h\nu'v'] \cdot h^{\nu'/2-1} \qquad (4.6)$$

where the definitions of m', v', and n' follow from those of Equation (4.5). The distributions for the marginals of $f'(\mu,h)$ can easily be found. For the mean, μ,

$$f'_s(\mu|m', \nu'/n', \nu') \propto [\nu' + (\mu-m')^2\, n'/v']^{-(\nu'+1)/2} \qquad (4.7)$$

which is Student, with moments

$$E[\mu] = m'$$
$$v[\mu] = \frac{v'}{n'}\,\frac{\nu'}{(\nu'-2)}\ ; \qquad \nu' > 2 \qquad (4.8)$$

And for the precision, h,

$$f'_{\gamma 2}(h|v', \nu') \propto \exp[-\tfrac{1}{2}\, h\nu'v']h^{\nu'/2-1} \qquad (4.9)$$

which is gamma-2. Often prior information is available on the variance, $\sigma^2 = 1/h$.

From (4.9), the distribution on σ^2 can be shown to be

$$f'_{I\gamma 1}(\sigma^2|\tfrac{1}{2}\,\nu', \tfrac{1}{2}\,\nu'v') \propto \exp[-\tfrac{1}{2}\,\sigma^2\nu'v'] \cdot (\sigma^2)^{\nu'/2+1} \qquad (4.10)$$

which is distributed inverted gamma-1 and has moments

$$E[\sigma^2] = \frac{1/2\nu'v'}{\frac{1}{2}\nu' - 1}$$

$$v[\sigma^2] = \frac{(1/2\nu'v')^2}{(\frac{1}{2}\nu' - 1)^2(\frac{1}{2}\nu' - 2)} \; ; \quad \nu' \geq 4 \qquad (4.11)$$

Information from the prior using the restrictive pdf assessment procedure will be the first two moments - the mean, μ, and the variance, σ^2. By solving Equations (4.8) and (4.11) the parameters of the prior can be determined.

By comparing Equations (4.8) and (4.11), it is important to notice that

1. $m' = E[\mu]$, the prior mean
2. $n' = \frac{E[\sigma^2]}{V[\mu]}$, the prior equivalent sample size in terms of the mean
3. $\nu' = 2 \cdot \frac{E^2[\sigma^2]}{V[\sigma^2]} + 4$
4. $v' = V[\mu] \cdot \frac{n'(\nu'-2)}{\nu'}$

   or $= E[\sigma^2] \cdot \frac{\nu'-2}{\nu'}$

Steps 1 through 4 can be easily carried out to find the parameters of the normal-gamma prior pdf. There is no redundancy in the statistics n' and ν' for the prior pdf, as there exists in the likelihood function. There, $\nu = n-1$, but for the prior ν' the relationship between ν' and n' need not follow the relationship $\nu' = n' - 1$

(Raiffa and Schlaiffer, 1961).

The posterior distribution of (μ,h) can be found by applying Equation (4.2) and will be Normal-Gamma, $f''_{N\gamma}(\mu,h|m'',v'', n'', \nu'')$, of the same form as (4.6) with the following parameters, (Raiffa and Schlaifer, 1961):

$$
\begin{aligned}
n'' &= n' + n \\
m'' &= \frac{1}{n''}(n' \cdot m' + n \cdot m) \\
\nu'' &= \nu' + \nu + 1 \\
v'' &= \frac{1}{\nu''}\left[(\nu'v' + n'm'^2) + (\nu v + nm^2) - n'' m''^2\right]
\end{aligned}
\qquad (4.12)
$$

where n, m, v, ν are the sample statistics defined in Equation (4.5) and n', m', ν', v' are the prior statistics.

4.2.2 Bayesian Distribution of the Annual Maximum Flood Discharge from a Normal Process

The Bayesian distribution of the annual maximum flood, q, is found by applying Equation (4.1). If the posterior pdf of the parameters μ,h is Normal-Gamma, as defined by Equation (4.6), with the parameters as given in (4.12) and if the modelled process is Normal with mean μ, precision h, then it can show, as in Appendix A, that the Bayesian distribution of q is Student. That is

$$\tilde{f}_S(\mu|m'',\frac{n''}{n''+1} \cdot v'',\nu'') = \iint_{\mu h} f_N(q|\mu,h) \cdot f_{N\gamma}(\mu,h|m'',v'',n'',\nu'') \cdot d\mu dh \qquad (4.13)$$

The form of the Student is

$$\tilde{f}_s(q|m'', v''/r, \nu'') = \frac{1}{B(\frac{1}{2}, \frac{1}{2}\nu'')} \cdot [1 + \frac{(q-m'')^2}{v''\nu''} r]^{-(\nu''+1)/2} \cdot [\frac{\nu''v''}{r}]^{-1/2} \quad (4.14)$$

where $B(\frac{1}{2}, \frac{1}{2}\nu'')$ is the Beta function defined as

$$\frac{\Gamma(1/2) \cdot \Gamma(1/2\nu'')}{\Gamma(1/2\nu'' + 1/2)}$$

$$r = n''/(n'' + 1)$$

The first two moments of $\tilde{f}_s(q)$ are:

$$E[q] = m''$$
$$v[q] = v'' \cdot (\frac{\nu''}{\nu''-2})(\frac{n''+1}{n''}) \quad (4.15)$$

Inferences about a Normal process with mean and variance unknown will be made from a Student pdf which fully accounts for the parameter uncertainty. For the same data, the Student pdf is similar in form to a normal pdf but has a larger variance. This is shown in Section 4.6.2 where Bayesian and classical distribution procedures are compared.

4.3 Bayesian Distribution of Annual Flood Discharges from a Log-Normal Process

In the previous section it was assumed that the distribution of the annual maximum flood event came from an independent normal process. For many rivers, sample information would seem to indicate that a more appropriate model of the underlying process would have a probability distribution which takes on values greater than zero only for positive flood discharges, $q > 0$, and that the distribution would be positively skewed (long upper tail). Such a model is an independent log-normal process.

Given that $z = \ln q$ is normally distributed with mean μ and variance σ^2, then q has, by definition, a log-normal distribution. It is well known that q has a mean of

$$\eta = \exp[\mu + \sigma^2/2]. \tag{4.16}$$

The model, where hydrologic variables follow a log-normal distribution, has been widely used in hydrologic analysis. The Bayesian pdf of the log-normal process can be found analytically by applying derived distribution theory to the results from the normal process. It can also be found by finding the posterior pdf of η from the joint pdf of $\ln \eta$ and , then integrating over $f''(\eta)$ numerically. This is the approach of Equation (4.1) to find the Bayesian pdf of the flood discharges. This latter procedure is given in Appendix B.

The approach in this section is to use the normal analysis results and then to apply transforms to obtain the results for the log-normal analysis. Using the results from the normal process analysis leads to analytical results for the log-normal process. This feature is very desirable.

If two random variables, say x and y, are related by a monotonic function, x = g(y), then their density functions are related by

$$f(y) = f(x) \cdot \left|\frac{dx}{dy}\right|$$
$$= f[g(y)] \cdot \left|\frac{dg(y)}{dy}\right| \qquad (4.17)$$

From Equation (4.14), the Bayesian distribution of stream flows from the normal process, with unknown mean and variance, is Student of the form

$$f_s(z|m'', v''/r, v'') = \frac{1}{B(1/2,\ 1/2\nu''} \cdot [1 + \frac{r(z-m'')^2}{v''\nu''}]^{-(\nu''+1)/2} \cdot [\frac{\nu''v''}{r}]^{-1/2} \qquad (4.18)$$

where $r = n''/n''+1$ and all other parameters are as defined in (4.14). From Equation (4.17) and the relationship between z and q; namely $z = \ln q$, the Bayesian pdf for q can be shown to be:

$$f(q|m'', v''/r, \nu'' = \frac{1}{q} \cdot \frac{1}{B(1/2,\ 1/2\nu'')} \cdot [1 + \frac{r(\ln q-m'')^2}{v''\cdot\nu''}]^{-(\nu''+1)/2} \cdot [\frac{r}{\nu''v''}]^{1/2} \qquad (4.19)$$

This pdf has been defined in the literature as the 'log-student' density function. The mean of the log-student does not exist (Zellner, 1971; Kaufman, 1972) which results in a number of interesting consequences.

In finite act decision problems with acts whose values are linear in q, the expected utility of every act will be infinite. This restriction is not crucial since realistic value functions in hydrologic decision problems are rarely linear in q. In the analysis of the log-normal process by numerical procedures, presented in Appendix B, the mean of the posterior pdf for the mean of q, $f''(\eta)$ does not exist. Thus, optimal Bayes point estimators from the often used quadratic loss function, $\ell(\eta,\hat{\eta}) = (\eta - \hat{\eta})^2$; symmetric linear, $|\eta - \hat{\eta}|$; or asymmetric linear will result in infinite losses for every choice of $\hat{\eta}$. This latter point does not bother the analysis for the Bayesian pdf, since $f''(\eta)$ is a proper density function and the integration over $f''(\eta)$ can be carried out.

4.4 Bayesian Distribution of Annual Flood Discharge from a Gamma-1 Process

The previous section considered the case where the magnitude of the annual flood series was distributed according to a log-normal pdf. Many rivers exhibit characteristics which would suggest that a probability model which has a long upper tail (positively skewed) would be an appropriate model. This is why the log-normal model has been widely used; however, it is often the case that when logs of the observed sample are taken, the resulting data is not symmetrical about the mean (no skew) as is predicted. If the logs of the flood series are negatively skewed, a more appropriate probability model of the flood discharge, q, may be to assume that they were generated from a gamma-1 process.

The gamma-1 pdf closely resembles that of the log-normal distribution. Both are uni-modal and positively skewed. The gamma-1 has a shorter tail so that the probability of observing an extreme event of the same magnitude will be higher for floods generated by a log-normal process. Table 4.1 compares the return period for extreme events from the gamma-1 process with those from the log-normal process. Both processes have a mean of 3000 cfs and a standard deviation of 2121 cfs.

If the logs were positively skewed, then it might be more appropriate to assume that the logs of the annual flood series are generated by a gamma-1 process; or that the flood discharges are generated by a log-gamma-1. A log-Pearson III pdf is a type of log-gamma. It is also interesting to realize that if the logs of a flood discharge sample are negatively skewed and they are fitted to a log-Pearson III using traditionally recommended procedures, then there exists, 1) a maximum flood that will occur, 2) a finite probability of observing a flood discharge that is less than or equal to zero. The Bayesian analysis of a log-gamma will not be analyzed here; instead, the analysis of the gamma-1 will be considered.

The pdf of flood discharge, q, generated from a gamma-1 process is of the form:

$$f(q|a,r) = \exp(- a\cdot q)\cdot \frac{q^{r-1} a^{r}}{\Gamma(r)} \qquad (4.20)$$

The Bayesian analysis for the gamma-1 process with both parameters unknown required numerical procedures due to the complex

Q	Log-Normal T	Gamma-1 T
500	1.00636	1.04665
1000	1.08646	1.1685
1500	1.28256	1.3589
2000	1.59948	1.62547
2500	2.05146	1.98482
3000	2.66477	2.46205
3500	3.47604	3.09217
4000	4.53203	1.92271
4500	5.89035	5.01768
5000	7.62076	6.46298
5500	9.80688	8.37377
6000	12.5481	10.9045
6500	15.9622	14.2624
7000	20.1874	18.7256
7500	25.3859	24.6672
8000	31.7468	32.5889
8500	39.4897	43.1637
9000	48.8698	57.2945
9500	60.1804	76.1894
10000	73.7596	101.467
10500	89.9959	135.274
11000	109.332	180.474
11500	132.277	240.81
12000	159.403	321.181
12500	191.368	427.837
13000	228.915	568.719
13500	272.88	753.558
14000	324.222	993.558
14500	384.006	1301.37
15000	453.414	1689.55
15500	533.83	2168.72
16000	626.67	2745.86
16500	733.655	3419.73
17000	856.592	4177.59
17500	997.694	4984.32
18000	1159.13	5821.38
18500	1343.47	6636.56
19000	1554.02	7403.89
19500	1793.2	8089.3
20000	2064.63	8683.86

Table 4.1

Return Periods for Gamma-1 and Log-Normal Process likelihood function. The likelihood function for a sample q of n (years) in length is:

$$L(a,r|q) = \prod_{i=1}^{n} [\exp(-a \cdot q_i) \cdot q_i^{r-1} \frac{a^r}{\Gamma(r)}] \tag{4.21}$$

To find the posterior pdf f"(a,r) requires the likelihood function to be rescaled numerically by the prior probability density function f'(a,r). The Bayesian pdf of the flood discharges, $\tilde{f}(q)$, is calculated by numerically integrating over the product of the posterior pdf f"(a,r) and the modelled density function for q, $f_{\gamma 1}(q|a,r)$. The numerical integration is over all values of a and r and follows from Equation 4.1. Section 4.6.4 presents the numerical values of f"(a,r) and $\tilde{f}(q)$ for the Blackstone River.

4.5 Bayesian Inference from an Exponential-Poisson Exceedance Model

The previous models of the distribution of flood events considered the annual series, that is, a series consisting of the largest flood in each year. Consider now a series composed of all independent flood peaks, often called the partial duration series, and of these flood peaks consider only those flood events above a level Q_b. These flood peaks will be assumed to be independent. While the pdf of the whole series may not be known, Q_b will be chosen large enough so that the probability distribution of flood events greater than Q_b will be assumed to be exponentially distributed, that is:

$$f_Q(q) = \alpha \exp[- \alpha(q-Q_b)] \tag{4.22}$$

for $q \geq Q_b$

Let $z = (q - Q_b)$ be the exceedance discharge so that Equation (4.24 becomes

$$f_Z(z) = \alpha e^{-\alpha z} \tag{4.23}$$

for $z \geq 0$

It is easily shown that the probability of an exceedance discharge being greater than z is

$$P_z = e^{-\alpha z} \tag{4.24}$$

If it is assumed that the time between independent flood peaks larger than the base Q_b is exponentially distributed with parameter ν, then the occurrence of exceedance flows will be governed by a Poisson process with an average arrival rate ν. It is then seen that the occurrence of floods above some exceedance level z is also Poisson, but with an average arrival rate $\nu \cdot P_z$.

The probability that in time t we have n_z exceedances (n exceedances above a flood exceedance level z) is just

$$P_N(N = n_z) = \frac{(\nu P_z)^{n_z}}{n_z!} \exp[-(\nu P_z) \cdot t] \tag{4.25}$$

But if z is such that no exceedances occur in time t, then $P_N(N=0)$ is the cumulative density function for z, F (z).

Substituting P_z into (4.25) for $n_z=0$ gives

$$F_Z(z) = \exp[-\nu t e^{-\alpha z}] \qquad (4.26)$$

for $z \geq 0$.

The probability that $z = 0$ is the probability that q is less than the base value Q_b.

If z is large so that the probability of exceeding it is small and the arrival rate of such events is small, then (4.28) can be closely approximated by:

$$F_Z(z) \simeq 1 - \nu t \cdot \exp[-\alpha z] \qquad (4.27)$$

Equation (4.26) or (4.27) represents the underlying probabilistic model of the exceedance floods. It is of a fairly general form since the upper tails of many distributions may be represented as exponential. The proposed model has been used in its classical formulation for extreme discharges by Shane and Lynn (1964) and Todorouic and Zelenhasic (1970) and for rainfall events by Grayman and Eagleson (1971).

4.5.1 Prior and Posterior PDF for Parameters α and ν

The posterior pdf for a parameter may be found by the application of Equation (4.2). It is assumed that α and ν are independent.

In the underlying model, the time between exceedance floods is assumed to be exponentially distributed with parameter ν. The likelihood function for ν will be

$$L[\nu \mid \underline{Z}(n,T)] \propto \nu^n e^{-T\nu} \qquad (4.28)$$

where $\underline{Z}(n,T)$ is the record of exceedance floods in which there occurred n exceedances in T years of record. Equation (4.28) is the kernel of a gamma-1 probability distribution. Thus the natural conjugate is also gamma-1 with parameters u' and s'; that is:

$$f'(\nu|u',s') \propto \nu^{u'} \exp[-\, s'\cdot\nu] \tag{4.29}$$

and the posterior pdf for ν is:

$$f''(\nu|u'',s'') \propto \nu^{u''} \exp[-\, s''\cdot\nu] \tag{4.30}$$

where

$u'' = u' + n$

$s'' = s' + T$

The probability density function for the magnitude of floods exceeding the base Q_b is assumed exponential with parameter α. The likelihood function will be:

$$L[\alpha|\underline{Z}] \propto \alpha^{n} \exp[-\, \alpha\sum_{i}^{n} z_i] \tag{4.31}$$

where

n is the number of exceedances

Z is a vector of exceedance discharges.

Like ν, the average arrival rate parameter, α has a likelihood function which has a gamma-1 kernel with parameters n and $\sum_i^n z_i$. The natural conjugate prior is gamma-1 and if it has parameters v', ℓ', then the posterior pdf of α will be gamma-1

$$f''(\alpha|v'',\ell'') \propto \alpha^{v''} \exp[-\alpha\cdot\ell''] \quad (4.32)$$

where

$$v'' = v' + n$$

$$\ell'' = \ell' + \sum^{n} z_i$$

4.5.2 Bayesian Distribution of Flood Events for the Exceedance Model

The Bayesian pdf f (z) follows from Equation (4.1), thus

$$\tilde{f}_Z(z) = \int_\alpha \int_\nu f_Z(z|\alpha,\nu)\cdot f''_{\gamma 1}(\alpha|v'',\ell'')\cdot f''_{\gamma 1}(\nu|u'',s'')\cdot d\nu\cdot d\alpha \quad (4.33)$$

where

$f_Z(z|\alpha,\nu) = \alpha\nu t\cdot\exp[-\alpha z]$, following from Equation (4.29)

The integration of (4.33) is given in Appendix C and is

$$\tilde{f}_Z(z) = \bar{\alpha}\bar{\nu}t[1 + \frac{\bar{\alpha}z}{v''+1}]^{-(v''+2)} \quad (4.34)$$

In a similar manner the exceedance probability, $G_Z(z) = 1 - F_Z(z)$, can be calculated to be:

$$G_Z(z) = \bar{\nu}t[1 + \frac{\bar{\alpha}z}{v''+1}]^{-(v''+1)} \quad (4.35)$$

The exceedance model is formulated to consider the distribution of extreme events. It is these events which are of interest to the decision maker and, yet, by the very nature of the problem, there are few observations to provide information about parameters. This scarcity of data does not imply, a priori, that the exceedance model does a poorer job of representing extreme events. In fact, since it was "designed" for extreme events, it may represent them very well.

4.6 Inference of Flood Discharges for the Blackstone River at Woonsocket, Rhode Island

4.6.1 Introduction

In estimating exceedance probabilities of flood discharges using distribution theory, there are two main areas of controversy - that of choosing the underlying process which best represents the occurrence of floods and that of estimating the uncertain parameters of the models. The first part of this chapter analyzed in detail the effects of parameter uncertainty. Bayesian distribution theory provides an approach to consider the whole pdf of uncertain parameters. This leads to the calculation of the Bayesian pdf of flood discharges, which are free of uncertain parameters. The Bayesian analysis considered four underlying models, which were the normal process, the log-normal process, the gamma-1 process and an exponential Poisson process.

This section applies these four models to the Blackstone River at Woonsocket, Rhode Island. With each model, inferences can be made concerning the probability that a flood greater than some magnitude, q, will occur. This probability is the exceedance probability and will be written G(q) and is equal to 1 - F(q) where F(q) is the cumulative density function.

Decisions governing flood designs are concerned with the occurrence of extreme events which are out on the tails of the probability density function. Each model has a different tail, which will lead to different exceedance probabilities and to different optimal designs. To test conclusively which model best represents a sample would require sample lengths far in excess of those normally observed in hydrology. Furthermore, all tests are weighted where the observations are - that is, around the moddle of the distribution. However, flood designs are affected by the tails of the distribution where there are few, if any, observations.

This section gives the flood frequency curve when each of the four previous models is assumed to be the true model. The information available to each model will be the same - the historical data and a prior obtained from a regional regression. Table 4.2 gives the basin characteristics, Table 4.3 gives the historical record of annual flood peaks from 1929 to 1965, and Figure 4.1 presents a sample histogram of the annual flood peaks.

The sample statistics are:

mean = 6372 cfs

standard deviation = 5206 cfs

coefficient of variation = .817

coefficient of skew = 3.76

The sample statistics, and especially the higher moments, are greatly affected by the extreme flood of 32900 cfs which occurred in 1955.

Area of basin = 416 square miles

Area of lakes, ponds and reservoirs = 12.5 square miles

Main stream length = 42.6 miles

Average channel slope = 11.5 ft. per mile

Average tributary slope = 41.5 ft. per mile

Average land slope = 307 ft. per mile

Mean altitude of basin = 495 feet

Datum of gage = 107.42 feet

Table 4.2

Basin Characteristics for the Blackstone River, at Woonsocket, Rhode Island

Year	Flood Discharge (cfs)
1929	4570
1930	1970
1931	8220
1932	4530
1933	5780
1934	6560
1935	7500
1936	15000
1937	6340
1938	15100
1939	3840
1940	5860
1941	4480
1942	5330
1943	5310
1944	3830
1945	3410
1946	3830
1947	3150
1948	5810
1949	2030
1950	3620
1951	4920
1952	4090
1953	5570
1954	9400
1955	32900
1956	8710
1957	3850
1958	4970
1959	5398
1960	4780
1961	4020
1962	5790
1963	4510
1964	5520
1965	5300

Table 4.3

Historical Record of Annual Flood Peaks,
Blackstone River, Woonsocket, Rhode Island

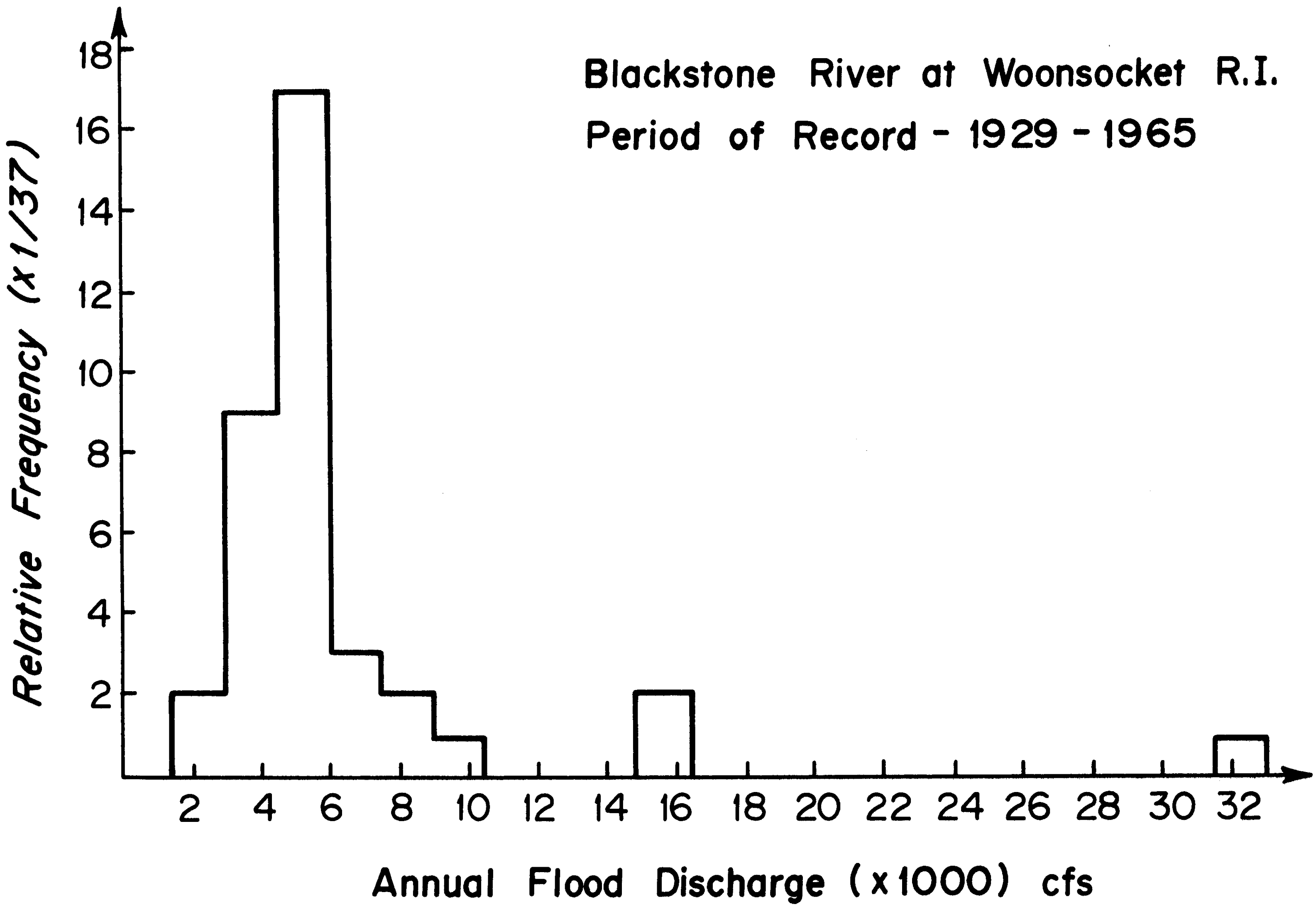


Figure 4.1: Sample Histogram of Annual Flood Peaks, Blackstone River, at Woonsocket, R.I.

4.6.2 Inferences from the Normal Process Reference Section 4.2

The Normal model, with mean and precision unknown, was analyzed with a Normal-Gamma prior in Section 4.2. The prior parameters were obtained from regression models of the mean and variance of the series of annual floods using 36 New England basins. This regression model was discussed in Chapter 3.

The prior parameters are:

$n' = 7$ years

$m' = 4042$ cfs

$\nu' = 6$ years

$v' = 9.22 \times 10^6 \text{ cfs}^2$

The sample parameters are:

$n = 37$ years

$m = 6372$ cfs

$\nu = 36$ years

$v = 27.1 \times 10^6 \text{ cfs}^2$

Thus giving the following posterior parameters (see Section 4.2 for the procedure to go from prior and sample parameters to posterior parameters).

$n'' = 44$ years

$m'' = 6001$ cfs

$\nu'' = 43$ years

$v'' = 24.7 \times 10^6 \text{ cfs}^2$

Table 4.4 gives the value of the probability density function, cumulative density, the exceedence probability and the return period for flood discharges up to 33000 cfs. Above this level the exceedence probability is virtually zero.

Figure 4.2 compares the three flood frequency curves that are obtained from using the prior information, using the sample data and using both prior and sample information. It is clear from Figure 4.2 that the posterior pdf is a weighted average of the prior information and sample information.

4.6.3 Inferences from the Log-Normal Process Reference Section 4.3

The log-normal model, with unknown mean and variance, was analyzed in Section 4.3. The prior parameters were obtained from regressions that are similar to the ones used for the analysis of the normal process. The prior parameters are

$n' = 4$ years

$m' = 6.4$ log cfs

$\nu' = 3$ years

$v' = .22$ log cfs^2

The sample parameters are

$n = 37$ years

$m = 8.6$ log cfs

$\nu = 36$ years

$v = .262$ log cfs^2

and the posterior parameters are

NORMAL MODEL

DISCHARGE	DENSITY	CUMULATIVE	EXCEEDANCE CUMULATIVE	RETURN PERIOD
0.	0.3850659E-04	0.11831117	0.88168883	1.134
1000.	0.4782701E-04	0.16150272	0.83849728	1.193
2000.	0.5719899E-04	0.21411514	0.78588486	1.272
3000.	0.6580737E-04	0.27580243	0.72419757	1.381
4000.	0.7277883E-04	0.34536219	0.65463781	1.528
5000.	0.7733043E-04	0.42074931	0.57925069	1.726
6000.	0.7891467E-04	0.49924076	0.50075924	1.997
7000.	0.7733694E-04	0.57773560	0.42226440	2.368
8000.	0.7279099E-04	0.65313214	0.34686786	2.883
9000.	0.6582242E-04	0.72270590	0.27729410	3.606
10000.	0.5721672E-04	0.78441048	0.21558952	4.638
11000.	0.4784561E-04	0.83704150	0.16295850	6.137
12000.	0.3852493E-04	0.88025194	0.11974806	8.351
13000.	0.2990321E-04	0.91443348	0.08556652	11.687
14000.	0.2240346E-04	0.94051749	0.05948251	16.812
15000.	0.1622367E-04	0.95974457	0.04025543	24.841
16000.	0.1137345E-04	0.97345459	0.02654541	37.671
17000.	0.7731092E-05	0.98292619	0.01707381	58.569
18000.	0.5104363E-05	0.98927647	0.01072353	93.253
19000.	0.3278945E-05	0.99341542	0.00658458	151.870
20000.	0.2053092E-05	0.99604219	0.00395781	252.665
21000.	0.1255275E-05	0.99766850	0.00233150	428.909
22000.	0.7507468E-06	0.99865264	0.00134736	742.190
23000.	0.4399847E-06	0.99923545	0.00076455	1307.961
24000.	0.2531173E-06	0.99957383	0.00042617	2346.464
25000.	0.1431806E-06	0.99976665	0.00023335	4285.367
26000.	0.7976729E-07	0.99987477	0.00012523	7985.348
27000.	0.4383581E-07	0.99993432	0.00006568	15224.332
28000.	0.2379809E-07	0.99996674	0.00003326	30066.695
29000.	0.1278154E-07	0.99998379	0.00001621	61680.937
30000.	0.6800430E-08	0.99999267	0.00000733	136400.125
31000.	0.3588903E-08	0.99999714	0.00000286	349525.312
32000.	0.1880952E-08	0.99999911	0.00000089	1118481.00

Table 4.4 Values of Probability Density, Cumulative Density, Exceedance Probability and Return Periods of Flood Discharges from the Blackstone River, for the Bayesian Normal Process

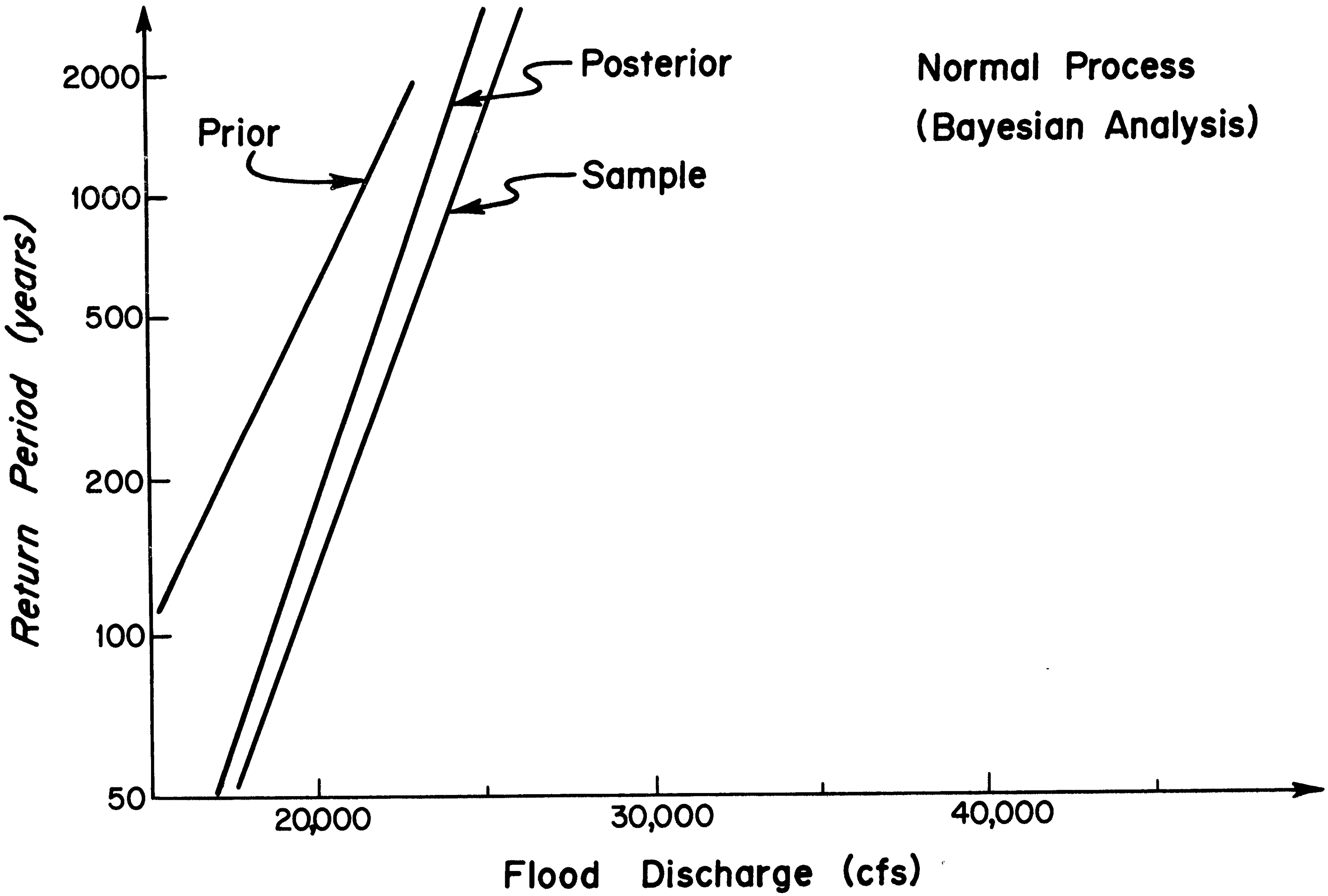


Figure 4.2: Bayesian Frequency Curves for the Normal Process

$n'' = 41$ years

$m'' = 8.39$ log cfs

$\nu'' = 40$ years

$v'' = .689$ log cfs^2

The values for the density function, cumulative density function, exceedance probability and teturn period for discharges up to 60,000 cfs are given in Table 4.5.

The flood frequency curves obtained from using the prior information, using the sample information and using both prior and sample information are given in Figure 4.3. Figure 4.4 compares the flood frequency curves obtained from the Bayesian analysis using both prior and sample information to the flood frequency curve from the classical analysis which uses maximum likelihood point estimators to estimate the uncertain parameters. It is interesting to note that the affect of accounting for the uncertainty in the parameters leads to higher flood discharges for the same exceedance probability (return period). In general, when the same information is used the Bayesian appraoch will always lead to higher discharges for the same return period; but, when both prior and sample data are used in the Bayesian analysis and only sample information is used in the classical analysis, no general statement can be made with respect to the relative positions of the two frequency curves.

4.6.4 Inferences from a Gamma-1 Process - Reference Section 4.4

The gamma-1 process for the generation of annual floods was analyzed in Section 4. Due to the complex form of the likelihood

LOG-NORMAL MODEL

DISCHARGE	DENSITY	CUMULATIVE	EXCEEDANCE CUMULATIVE	RETURN PERIOD
0.	0.0	0.0	1.00000000	1.000
1000.	0.8308068E-05	0.00161082	0.99838918	1.002
2000.	0.7360293E-04	0.03811926	0.96188074	1.040
3000.	0.1435299E-03	0.14985418	0.85014582	1.176
4000.	0.1651569E-03	0.30826312	0.69173688	1.446
5000.	0.1499817E-03	0.46784210	0.53215790	1.879
6000.	0.1209856E-03	0.60375577	0.39624423	2.524
7000.	0.9172902E-04	0.70983219	0.29016781	3.446
8000.	0.6733890E-04	0.78888947	0.21111053	4.737
9000.	0.4866807E-04	0.84643847	0.15356153	6.512
10000.	0.3496923E-04	0.88789058	0.11210942	8.920
11000.	0.2512582E-04	0.91766334	0.08233666	12.145
12000.	0.1811706E-04	0.93908608	0.06091392	16.417
13000.	0.1313748E-04	0.95457226	0.04542774	22.013
14000.	0.9593125E-05	0.96583807	0.03416193	29.272
15000.	0.7058537E-05	0.97409379	0.02590621	38.601
16000.	0.5235277E-05	0.98019093	0.01980907	50.482
17000.	0.3914517E-05	0.98473060	0.01526940	65.490
18000.	0.2950555E-05	0.98813790	0.01186210	84.302
19000.	0.2241591E-05	0.99071574	0.00928426	107.709
20000.	0.1716140E-05	0.99268103	0.00731897	136.631
21000.	0.1323679E-05	0.99419105	0.00580895	172.148
22000.	0.1028317E-05	0.99535972	0.00464028	215.504
23000.	0.8044110E-06	0.99627048	0.00372952	268.131
24000.	0.6334412E-06	0.99698490	0.00301510	331.664
25000.	0.5019891E-06	0.99754924	0.00245076	408.036
26000.	0.4002530E-06	0.99799764	0.00200236	499.411
27000.	0.3210027E-06	0.99835593	0.00164407	608.245
28000.	0.2588836E-06	0.99864405	0.00135595	737.492
29000.	0.2099067E-06	0.99887693	0.00112307	890.416
30000.	0.1710712E-06	0.99906611	0.00093389	1070.795
31000.	0.1401068E-06	0.99922055	0.00077945	1282.956
32000.	0.1152873E-06	0.99934733	0.00065267	1532.166
33000.	0.9529475E-07	0.99945170	0.00054830	1823.809
34000.	0.7911035E-07	0.99953800	0.00046200	2164.523
35000.	0.6594712E-07	0.99960977	0.00039023	2562.581
36000.	0.5519530E-07	0.99966961	0.00033039	3026.739
37000.	0.4637253E-07	0.99971968	0.00028032	3567.343
38000.	0.3910412E-07	0.99976170	0.00023830	4196.398
39000.	0.3309206E-07	0.99979717	0.00020283	4930.125
40000.	0.2810007E-07	0.99982709	0.00017291	5783.250
41000.	0.2393911E-07	0.99985248	0.00014752	6778.672
42000.	0.2045905E-07	0.99987406	0.00012594	7939.996
43000.	0.1753844E-07	0.99989235	0.00010765	9289.707
44000.	0.1507870E-07	0.99990803	0.00009197	10873.113
45000.	0.1300085E-07	0.99992150	0.00007850	12738.961
46000.	0.1124011E-07	0.99993294	0.00006706	14913.078
47000.	0.9743545E-08	0.99994284	0.00005716	17494.488
48000.	0.8467815E-08	0.99995136	0.00004864	20560.312
49000.	0.7377302E-08	0.99995863	0.00004137	24174.660
50000.	0.6442548E-08	0.99996489	0.00003511	28484.234
51000.	0.5639347E-08	0.99997038	0.00002962	33756.973
52000.	0.4947264E-08	0.99997514	0.00002486	40233.129
53000.	0.4349566E-08	0.99997908	0.00002092	47798.336
54000.	0.3832138E-08	0.99998266	0.00001734	57653.660
55000.	0.3383166E-08	0.99998569	0.00001431	69905.062
56000.	0.2992686E-08	0.99998808	0.00001192	83886.062
57000.	0.2652361E-08	0.99999046	0.00000954	104857.562
58000.	0.2355121E-08	0.99999273	0.00000727	137518.125
59000.	0.2094944E-08	0.99999392	0.00000608	164482.500
60000.	0.1866806E-08	0.99999511	0.00000489	204600.187

Table 4.5 Values of Probability Density, Cumulative Density Exceedance Probability and Return Periods of Flood Discharges from the Blackstone River, for the Bayesian Log-Normal Process

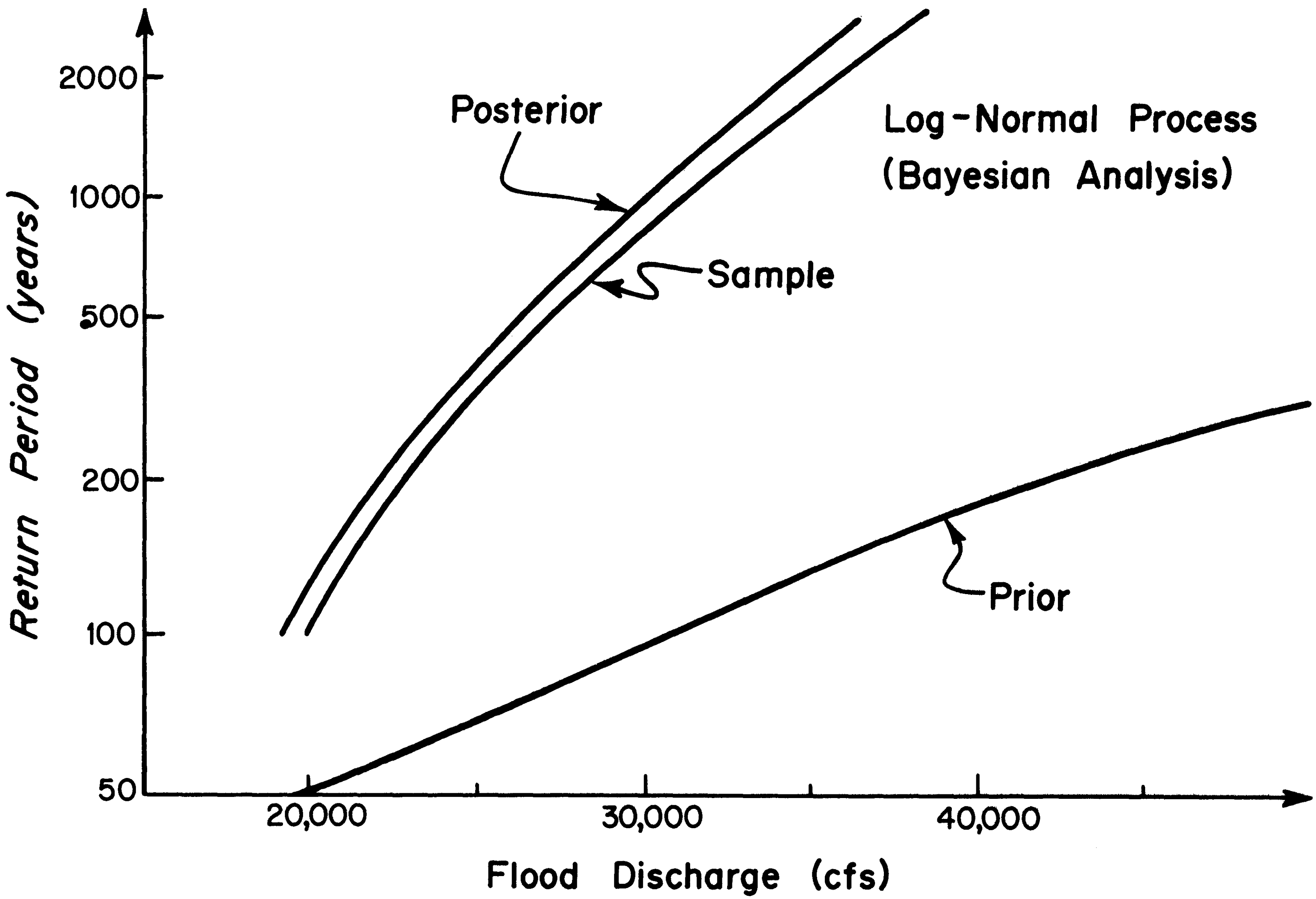


Figure 4.3: Bayesian Frequency Curves for the Log-Normal Process

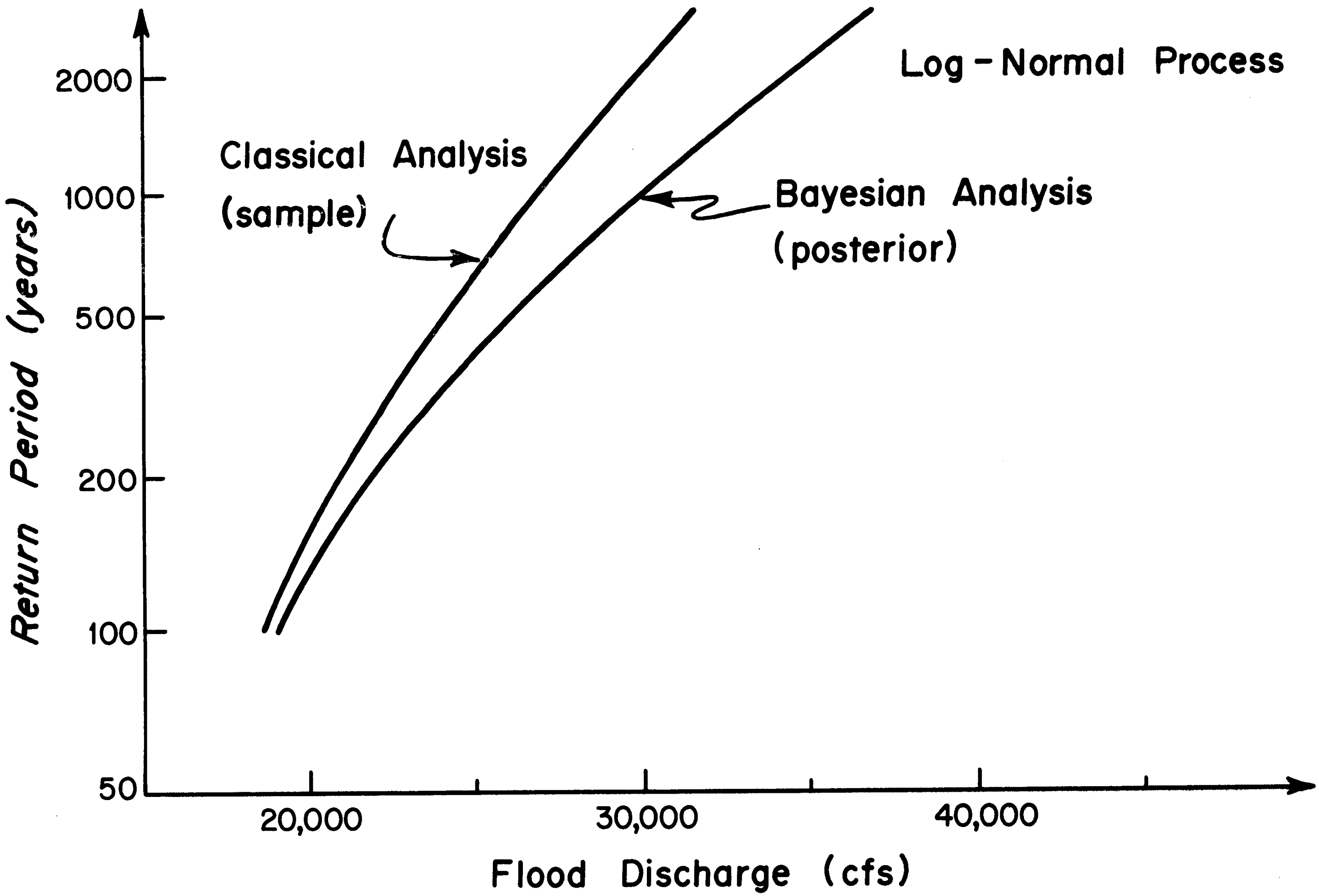


Figure 4.4: Comparison of Classical and Bayesian Frequency Curves for the Log-Normal Process

function, numerical procedures were required to find the posterior pdf for the uncertain parameters and to find the Bayesian pdf for the annual flood series.

The prior pdf's for f'(a) and f'(r) were of the family of gamma-1 functions. The prior parameters were estimated from two regional regressions - one for the mean annual flood and one for the coefficient of variation. The regional regression used the same 36 New England basins described in Chapter 3.

The prior pdf for the scale parameter, a is:

$$f'(a) \propto \exp(-\ 7677 \cdot a) \cdot a^{3} \tag{4.36}$$

and for the location parameter, r,

$$f'(r) \propto \exp(-\ .922 \cdot r) \cdot r^{2} \tag{4.37}$$

The posterior pdf is obtained using a numerical procedure to solve Equation 4.2 and is shown graphically in Figure 4.5.

The Bayesian pdf is evaluated numerically, and the values for the density function, cumulative, exceedance probability and return period for flood discharges up to 40,000 cfs are given in Table 4.6.

Figure 4.6 compares the flood frequency curves from the Bayesian analysis using just the prior information, just the sample information and both prior and sample. Figure 4.7 compares the flood frequency curves from the Bayesian analysis using both prior and sample information to the flood frequency curve from the classical analysis which estimated the parameters by maximum likelihood point

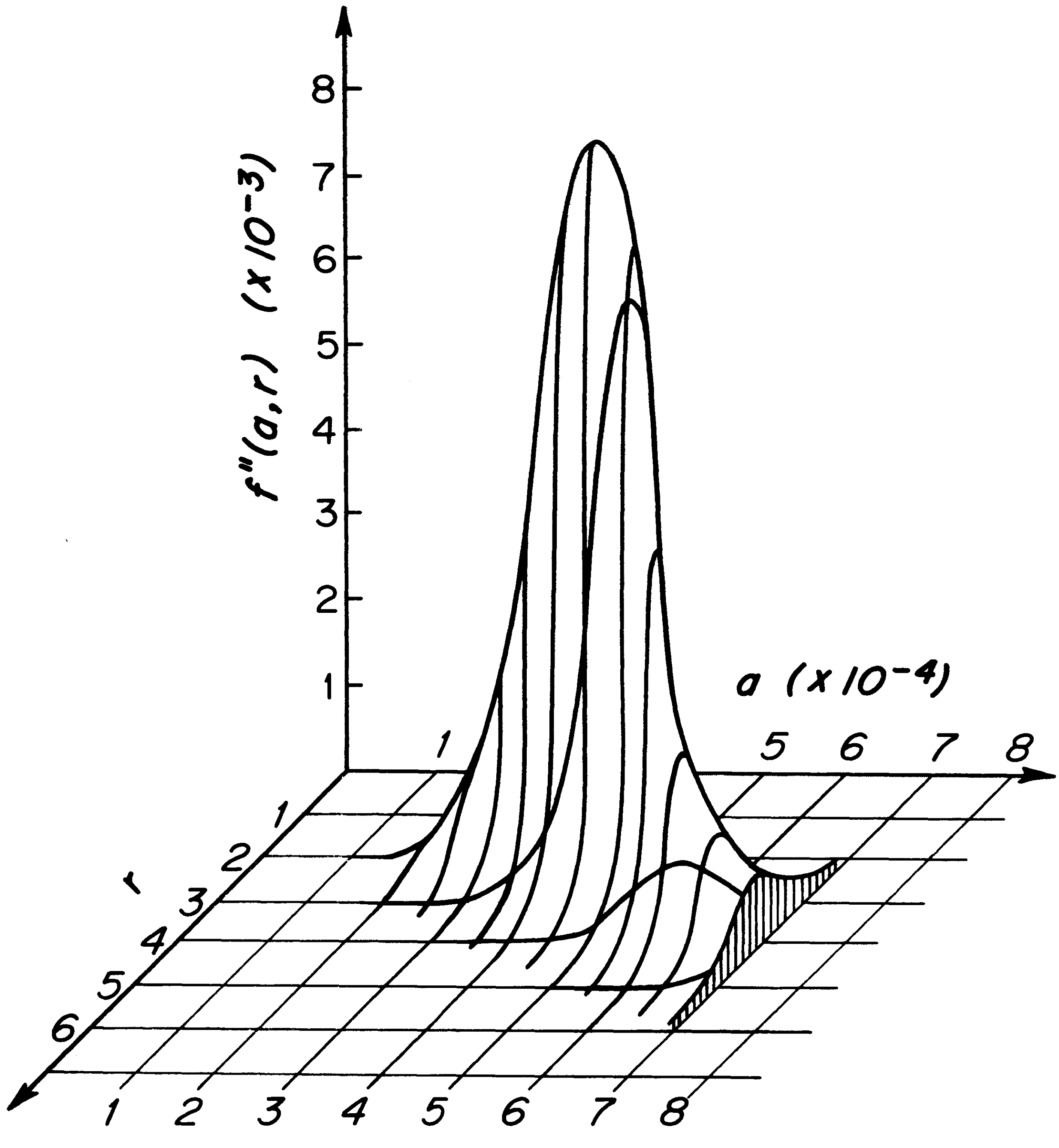


Figure 4.5: Posterior Density Function for the Parameters of the Gamma-1 Model

GAMMA MODEL

DISCHARGE	DENSITY	CUMULATIVE	EXCEEDANCE CUMULATIVE	RETURN PERIOD
0.	0.0	0.0	1.00000000	1.000
1000.	0.4030846E-04	0.01768712	0.98231292	1.018
2000.	0.8592839E-04	0.08174777	0.91825223	1.089
3000.	0.1145047E-03	0.18356329	0.81643671	1.225
4000.	0.1241931E-03	0.30427301	0.69572699	1.437
5000.	0.1197980E-03	0.42715394	0.57284606	1.746
6000.	0.1071468E-03	0.54107523	0.45892477	2.179
7000.	0.9094822E-04	0.64025855	0.35974145	2.780
8000.	0.7432996E-04	0.72284335	0.27715665	3.608
9000.	0.5905701E-04	0.78938586	0.21061414	4.748
10000.	0.4592385E-04	0.84169036	0.15830964	6.317
11000.	0.3512437E-04	0.88202971	0.11797029	8.477
12000.	0.2652040E-04	0.91268647	0.08731353	11.453
13000.	0.1982355E-04	0.93571877	0.06428123	15.557
14000.	0.1470193E-04	0.95286840	0.04713160	21.217
15000.	0.1083728E-04	0.96554905	0.03445095	29.027
16000.	0.7951317E-05	0.97487485	0.02512515	39.801
17000.	0.5813040E-05	0.98170501	0.01829499	54.660
18000.	0.4238663E-05	0.98669177	0.01330823	75.141
19000.	0.3084921E-05	0.99032438	0.00967562	103.353
20000.	0.2242480E-05	0.99296659	0.00703341	142.179
21000.	0.1628938E-05	0.99488646	0.00511354	195.559
22000.	0.1182944E-05	0.99628067	0.00371933	268.865
23000.	0.8591364E-06	0.99729311	0.00270689	369.428
24000.	0.6242353E-06	0.99802852	0.00197148	507.232
25000.	0.4538448E-06	0.99856287	0.00143713	695.832
26000.	0.3302520E-06	0.99895144	0.00104856	953.684
27000.	0.2405686E-06	0.99923420	0.00076580	1305.823
28000.	0.1754497E-06	0.99944037	0.00055963	1786.901
29000.	0.1281267E-06	0.99959058	0.00040942	2442.454
30000.	0.9370183E-07	0.99970025	0.00029975	3336.094
31000.	0.6862979E-07	0.99978060	0.00021940	4557.785
32000.	0.5034644E-07	0.99983943	0.00016057	6227.621
33000.	0.3699512E-07	0.99988240	0.00011760	8503.402
34000.	0.2722997E-07	0.99991393	0.00008607	11618.570
35000.	0.2007706E-07	0.99993718	0.00006282	15917.660
36000.	0.1482906E-07	0.99995428	0.00004572	21873.812
37000.	0.1097230E-07	0.99996680	0.00003320	30120.672
38000.	0.8133167E-08	0.99997598	0.00002402	41630.809
39000.	0.6039496E-08	0.99998266	0.00001734	57653.660
40000.	0.4492875E-08	0.99998766	0.00001234	81049.312

Table 4.6 Values of Probability Density, Cumulative Density, Exceedance Probability and Return Periods of Flood Discharges from the Blackstone River, for the Gamma-1 Process

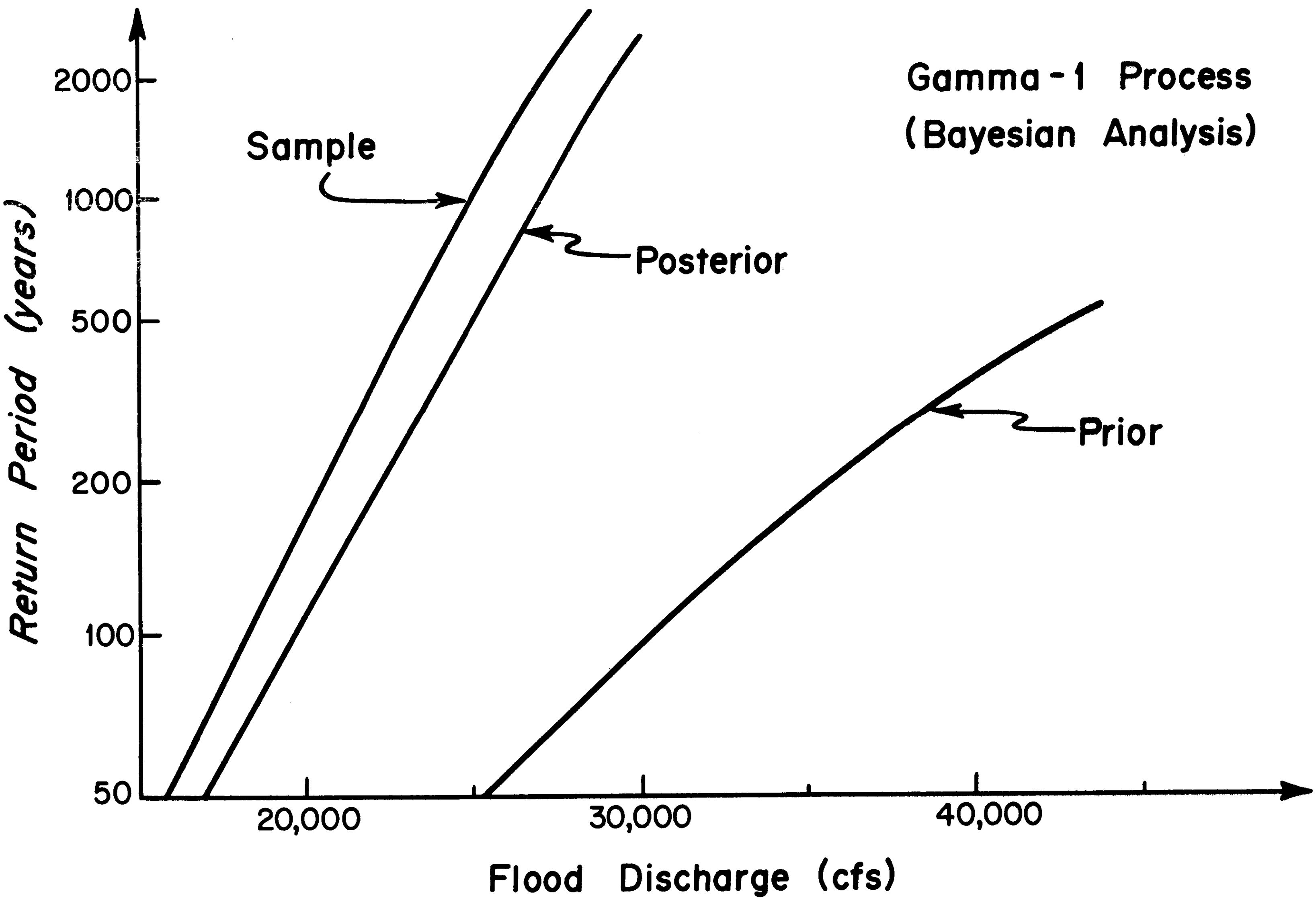


Figure 4.6 Bayesian Frequency Curves for the Gamma-1 Process

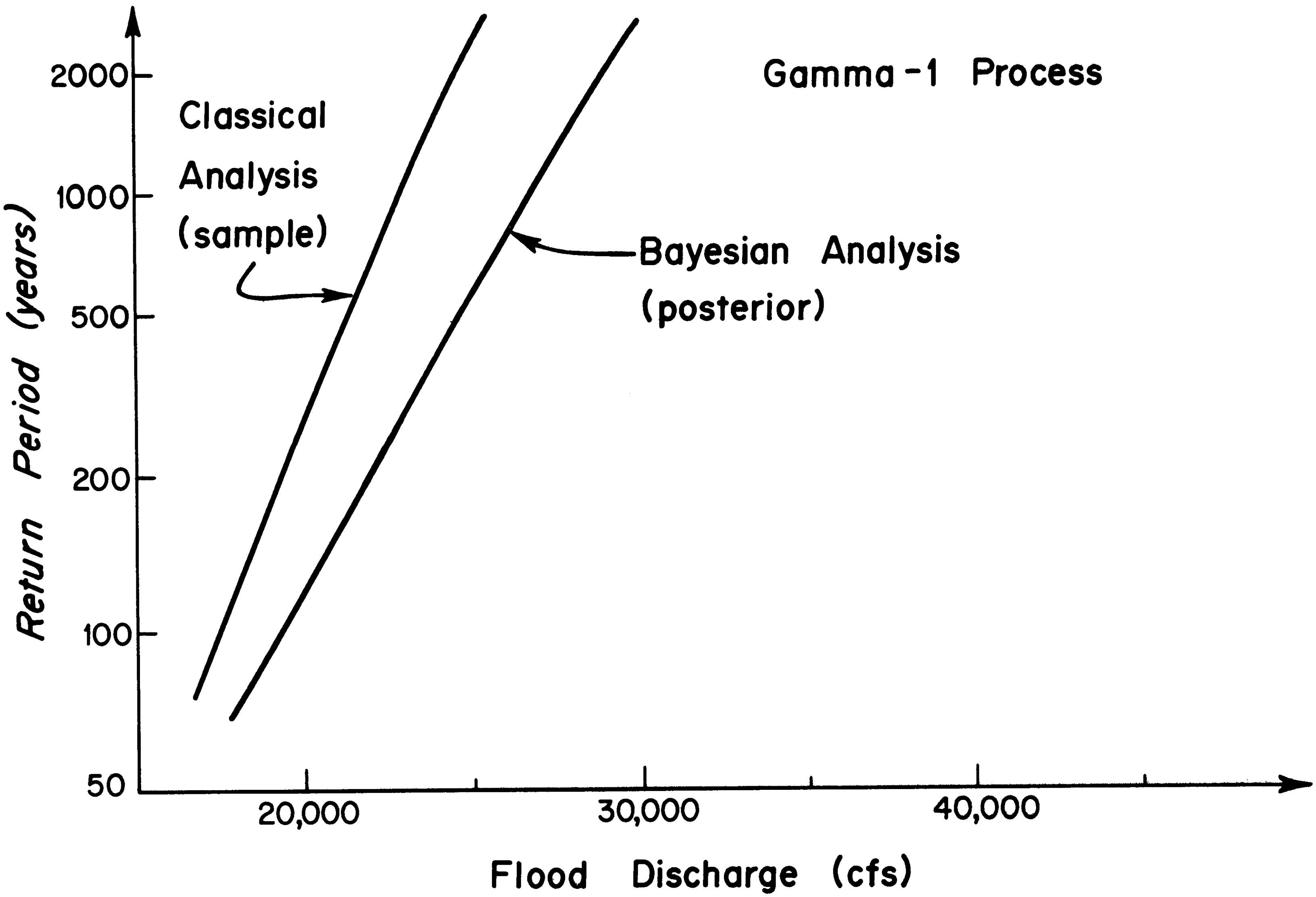


Figure 4.7: Comparisons of Classical and Bayesian Frequency Curves for the Gamma-Process

estimation.

4.6.5 Inferences from the Exponential-Poisson Process - Reference Section 4.5

The exceedance model considers flood discharges greater that 8500 cfs. It is assumed that these floods have a pdf that is exponential and that the pdf of the time between floods exceeding 8500 is also exponential (Poisson arrivals). This model is analyzed fully in Section 4.5.

The prior distribution on ν, the average arrival rate of floods larger than 8500 cfs, was obtained through a subjective assessment. The subjective assessment could have been obtained by an engineer saying simply "based upon my experience in the area, my best estimate of ν is .1 and there is a 50-50 chance that ν could be plus or minus .030 of .1". The implication of this statement is that the standard deviation is about .044. If this is accepted for the example, then the prior pdf of ν is

$$f(\nu|u',s') \propto \nu^{u'} \exp(-\ s'\cdot\nu)$$

where

$$u' = 5$$

$$s' = 50$$

Appendix D gives a procedure for obtaining $f'(\nu)$ based upon the uncertainty in plotting positions and the subjective assessment of equivalent record lengths.

The sample parameters are u = 5, s = 37 which together with the prior parameters gives a gamma-1 posterior pdf with parameters u" = 10, s" = 87. The posterior average arrival rate, $\bar{\nu}$, is .115 events per year.

The prior pdf for the event magnitude parameter, α, is estimated using a regression equation on the mean exceedance flood (exceedance floods being those annual floods larger than 8500 cfs). The prior pdf for is

$$f(\alpha|v',\ell') \propto \alpha^{v'} \exp(-\ell'\cdot\alpha)$$

where

$v' = 2$ events

$\ell' = 10850$ cfs

Sample values for the parameters are v = 5, ℓ = 38610 which results in a gamma-1 posterior density function with parameters v" = 7, ℓ" = 49468.

Table 4.7 gives the density function, cumulative density, exceedance probability and return period for flood discharges from 8500 cfs to 60,000 cfs. It is interesting to note that only once in approximately 9 years will a flood greater than 8500 cfs occur.

Figure 4.8 compares the flood frequency curves obtained from just using prior information, from just using sample information and fron using both prior and sample information.

Figure 4.9 compares the flood frequency curves obtained from the Bayesian analysis, using both prior and sample information, to

E X C E E D A N C E   M O D E L

DISCHARGE	DENSITY	CUMULATIVE	EXCEEDANCE CUMULATIVE	RETURN PERIOD
8500.	0.1627249E-04	0.88459999	0.11500001	8.696
9000.	0.1501463E-04	0.89281702	0.10718298	9.330
10000.	0.1281365E-04	0.90669835	0.09330165	10.718
11000.	0.1096905E-04	0.91856271	0.08143729	12.279
12000.	0.9417840E-05	0.92873394	0.07126606	14.032
13000.	0.8109097E-05	0.93747896	0.06252104	15.995
14000.	0.7001378E-05	0.94501919	0.05498081	18.188
15000.	0.6061044E-05	0.95153767	0.04846233	20.635
16000.	0.5260425E-05	0.95718771	0.04281229	23.358
17000.	0.4576827E-05	0.96209735	0.03790265	26.383
18000.	0.3991531E-05	0.96637422	0.03362578	29.739
19000.	0.3489128E-05	0.97010815	0.02989185	33.454
20000.	0.3056752E-05	0.97337574	0.02662426	37.560
21000.	0.2683726E-05	0.97624141	0.02375859	42.090
22000.	0.2361136E-05	0.97875994	0.02124006	47.081
23000.	0.2081506E-05	0.98097801	0.01902199	52.571
24000.	0.1838595E-05	0.98293525	0.01706475	58.600
25000.	0.1627113E-05	0.98466563	0.01533437	65.213
26000.	0.1442605E-05	0.98619843	0.01380157	72.455
27000.	0.1281304E-05	0.98755854	0.01244146	80.376
28000.	0.1140005E-05	0.98876774	0.01123226	89.029
29000.	0.1016001E-05	0.98984438	0.01015562	98.468
30000.	0.9069669E-06	0.99080467	0.00919533	108.751
31000.	0.8109203E-06	0.99166262	0.00833738	119.942
32000.	0.7261615E-06	0.99243033	0.00756967	132.106
33000.	0.6512440E-06	0.99311823	0.00688177	145.311
34000.	0.5849097E-06	0.99373567	0.00626433	159.634
35000.	0.5260809E-06	0.99429053	0.00570947	175.148
36000.	0.4738247E-06	0.99479002	0.00520998	191.939
37000.	0.4273343E-06	0.99524015	0.00475985	210.091
38000.	0.3859149E-06	0.99564636	0.00435364	229.693
39000.	0.3489572E-06	0.99601346	0.00398654	250.844
40000.	0.3159338E-06	0.99634558	0.00365442	273.641
41000.	0.2863849E-06	0.99664646	0.00335354	298.193
42000.	0.2599079E-06	0.99691939	0.00308061	324.611
43000.	0.2361541E-06	0.99716717	0.00283283	353.004
44000.	0.2148145E-06	0.99739248	0.00260752	383.505
45000.	0.1956202E-06	0.99759752	0.00240248	416.236
46000.	0.1783338E-06	0.99778438	0.00221562	451.340
47000.	0.1627465E-06	0.99795479	0.00204521	488.946
48000.	0.1486754E-06	0.99811035	0.00188965	529.199
49000.	0.1359585E-06	0.99825257	0.00174743	572.269
50000.	0.1244524E-06	0.99838269	0.00161731	618.310
51000.	0.1140297E-06	0.99850184	0.00149816	667.484
52000.	0.1045791E-06	0.99861109	0.00138891	719.990
53000.	0.9600075E-07	0.99871129	0.00128871	775.969
54000.	0.8820575E-07	0.99880332	0.00119668	835.644
55000.	0.8111562E-07	0.99888796	0.00111204	899.245
56000.	0.7466008E-07	0.99896580	0.00103420	966.931
57000.	0.6877696E-07	0.99903744	0.00096256	1038.901
58000.	0.6341031E-07	0.99910349	0.00089651	1115.432
59000.	0.5851018E-07	0.99916440	0.00083560	1196.748
60000.	0.5403194E-07	0.99922067	0.00077933	1283.152

Table 4.7 Values of Probability Density, Cumulative Density, Exceedance Probability and Return Periods of Flood Discharges from the Blackstone River, for the Exceedance Process

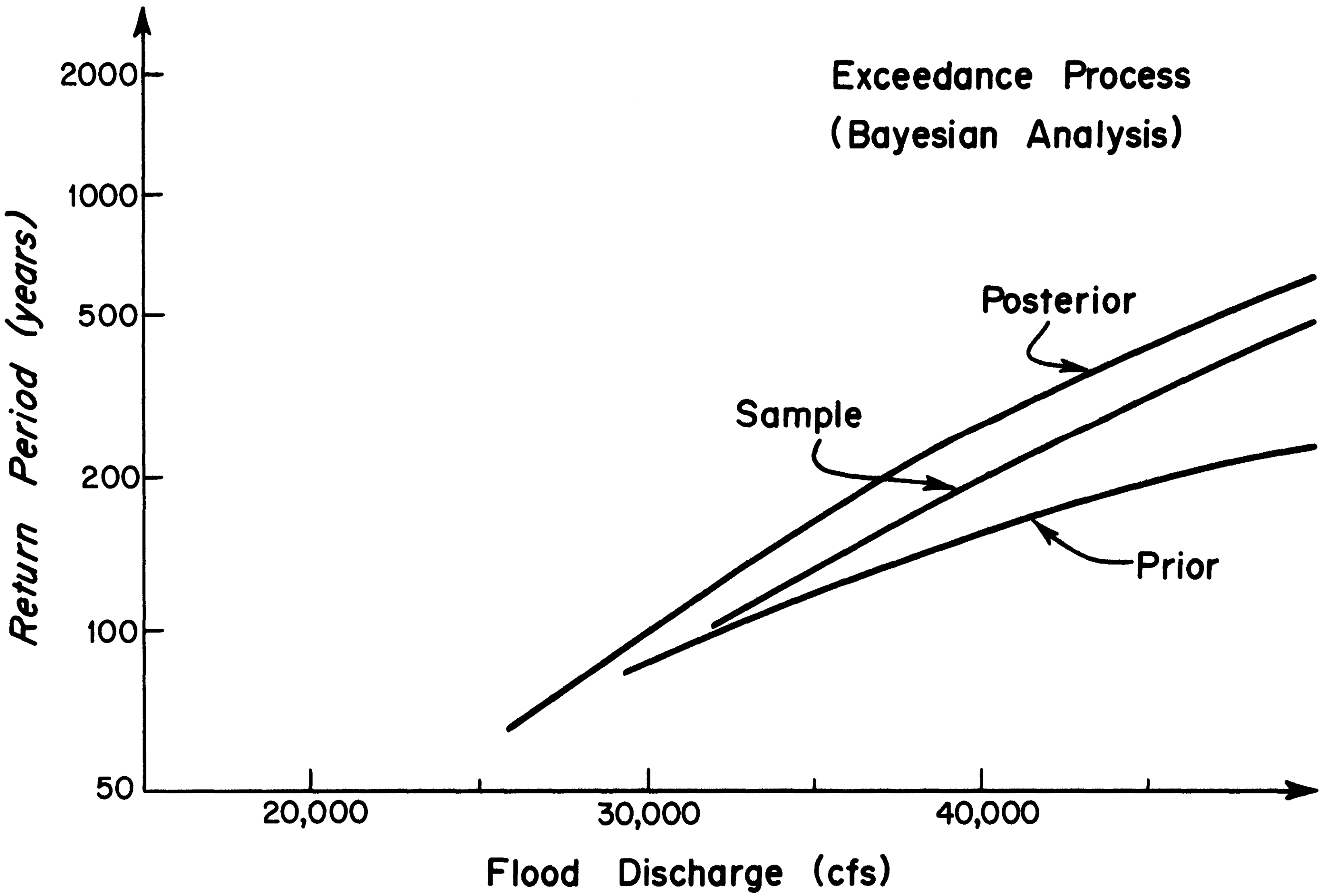


Figure 4.8: Bayesian Frequency Curve for the Exceedance Process

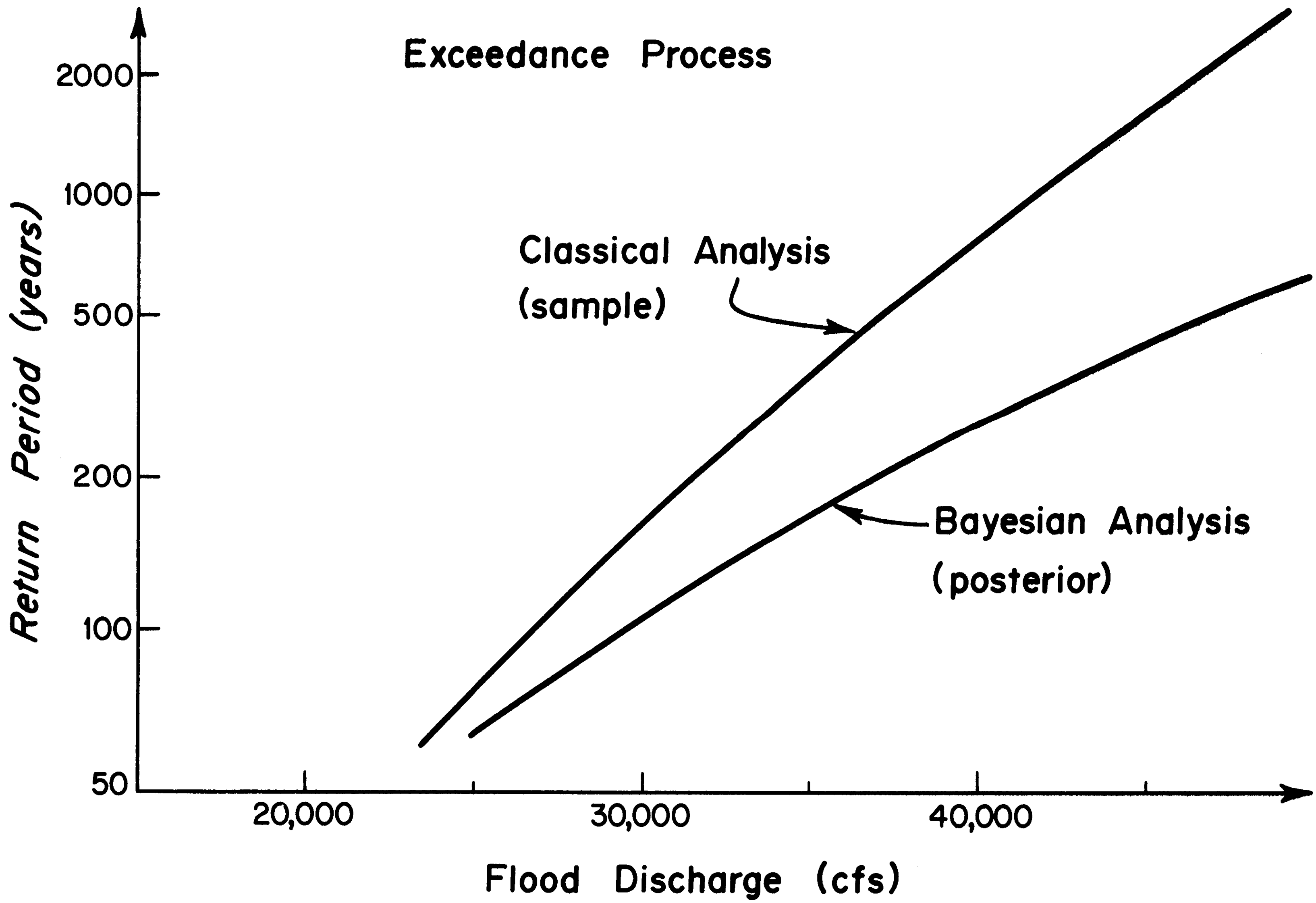


Figure 4.9 Comparison of Classical and Bayesian Frequency Curves for the Exceedance Process

the flood frequency curve obtained from a classical analysis using point estimators. As in the other models, the inclusion of parameter uncertainty leads to higher flood discharges for the same exceedance probability.

4.6.6 Summary

This section calculated the flood frequency curves for the Blackstone River at Woonsocket, Rhode Island using four different models of the underlying process. Bayesian procedures were used to find the pdf for the flood discharges when parameter uncertainty was considered. The resulting distributions were used to find the exceedance probability for flood discharges. The flood frequency curves from the four models were compared in Figure 4.10. In all four cases the Bayesian analysis, using both prior and sample information, was used. The flood frequency curves for the models were quite different, especially for large return periods.

4.7 Conclusion

This chapter recognized that when flood frequency analysis is performed using distribution theory, a number of issues are raised. The first issue is how to handle uncertain parameters. Bayesian distribution theory allows us to estimate probability density functions for uncertain parameters, using various sources of data, and then to obtain the Bayesian pdf for the flood discharges that will be 'free' of the uncertain parameters. Equations 4.1 and 4.2 set out the analytical procedures. The second issue raised is which probabilistic

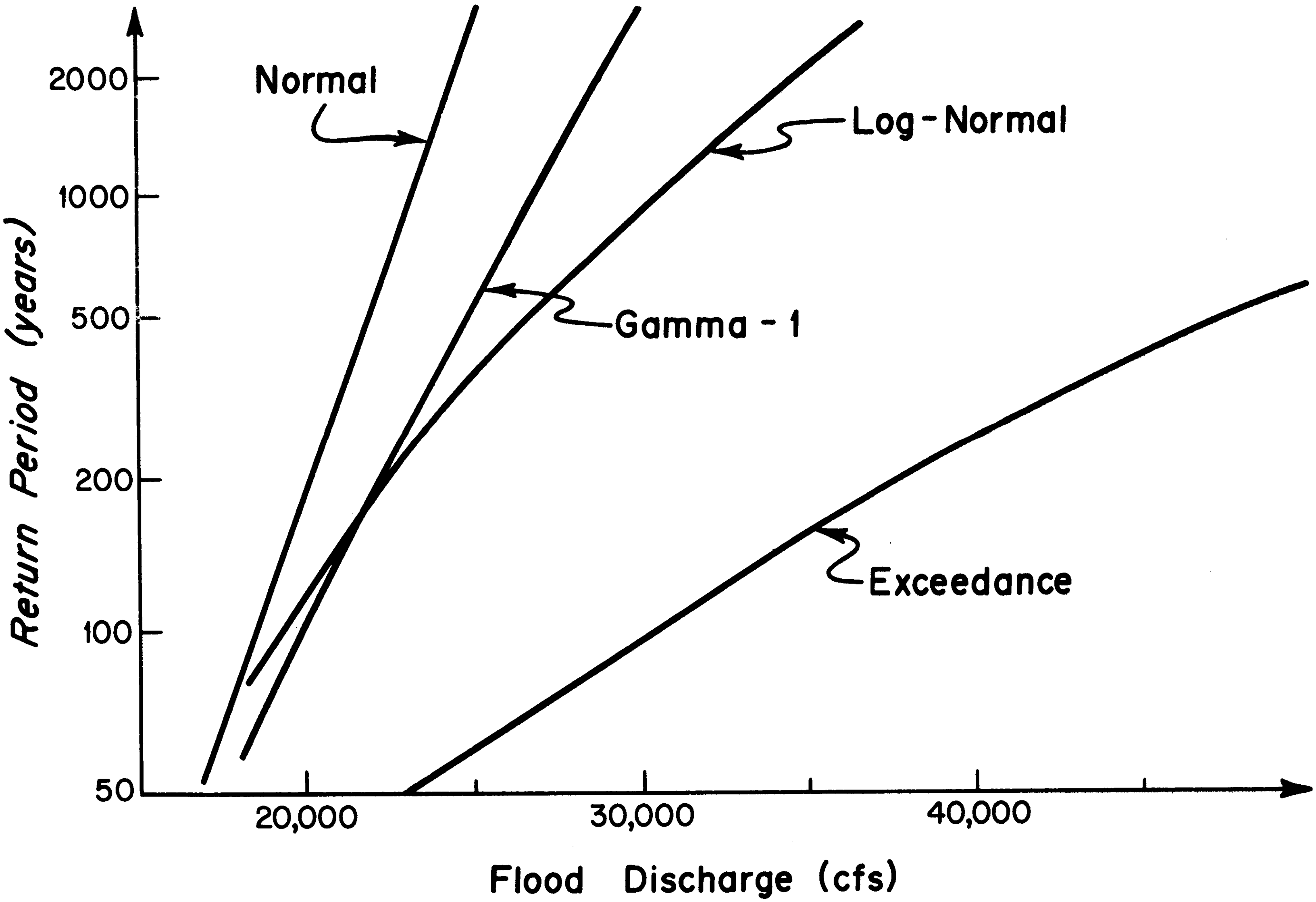


Figure 4.10 Comparison of Bayesian Frequency Curves for the Normal, Log-Normal, Gamma-1 and Exceedance Processes

model best represents the underlying process. Four models were analyzed in detail and applied to the Blackstone River at Woonsocket, Rhode Island. The model gave quite different results, but generally the gamma-1, log-normal, and to a lesser extent, the normal models fell closer together than the exponential-Poisson model that only considers floods in the tail of the distribution of all flood events. Which model is the most appropriate model has not yet been addressed and will be left to Chapter 7.

One of the underlying assumptions in the estimation of flood exceedance probabilities by distribution theory is that the probability of the occurrence of a flood of a certain magnitude will remain constant. In many river basins, this assumption does not hold, due to urbanization of the watershed or structural changes to the river channel. In these cases, hydrologists have recently begun to apply rainfall-runoff simulation models to find the flood frequency curves. Such an analysis contains uncertain parameters which can also be treated as random variables. There are the parameters of the rainfall pdf which can be analyzed using the theory of this chapter, and there are uncertain parameters in the modelling of the overland flow or runoff. These parameters can be considered as random variables and analyzed in a Bayesian framework. This analysis is carried out in the next chapter.

Chapter 5

Bayesian Analysis of Rainfall Runoff Modelling

5.1 Introduction

Chapter 4 considers the uncertainty in flood frequency analysis when distribution theory is applied. This uncertainty is centered upon the parameters of the underlying probability density function or model that is used to represent the occurrence of flood events. The uncertainty as to model choice is discussed by comparing the results obtained from the alternative models, but the analysis of model uncertainty is left to Chapter 7.

The analysis of flood frequency by applying distribution theory uses as inputs both prior information and historical flood records. It has been shown that prior information based upon a regional regression or an analytical flood frequency analysis provides relatively little information. It is the historical record that provides most of the information upon which inferences are drawn.

The analysis of flood frequency using distribution theory has the basic assumption that the probability of a flood of a given magnitude is constant and does not change with time. Thus, basins which change physically with time, due to changes in the river itself, through channelization for example, or due to urbanization of the watershed, can not be analyzed effectively by the distribution theory procedures of flood frequency analysis.

This problem has been recognized and some procedures have been applied to estimate the frequency curves. The most successful methods are those that analyze the rainfall as a stochastic process and then estimate the flood discharge by modelling the physical process of overland flow of the excess rainfall. This has been done analytically by Eagleson (1972), and through simulation by Leclerc and Schaake (1972), Ott and Linsley (1972), and others. Such frequency analyses have often been criticized (Ibbitt, 1972) on the basis that the deterministic catchment model has parameters which are unknown with certainty and whose values seem to be determined through 'intuition' and best guesses. Chapter 4 analyzed the uncertainty in the parameters of the probability distributions of floods by considering the parameters as random variables and applying Bayesian statistics. The resulting probability distributions of floods reflected the uncertainty in their parameters. In an analogous manner, the uncertainty in the flood frequency curve, due to uncertainty in the rainfall-runoff simulation modelling, may be analyzed. The uncertain parameters, whether they are in the pdf of the rainfall model or in the deterministic runoff model, may be regarded as random variables. The procedures of Bayesian statistics can then be applied.

While this chapter is aimed at simulation modelling, the vehicle for the analysis will be Eagleson's (1972), analytical derivation. Eagleson's derivation is used in the analysis, and the extension to computer simulation modelling is straightforward.

5.2 General Theory of Derived Flood Frequency Analysis

Flood Frequency analysis aims at finding the probability that a flood will have a discharge less than or equal to some value q_m. This probability is defined as the cumulative density function (CDF) evaluated at q_m and written as $F(q_m)$.

Consider the case when all parameters are known with certainty. The modelling procedure for $F(q_m)$ can be considered as a simple urn problem. A random sample is drawn from an urn which yields the values of the elements of $\underline{\theta}$, a vector that describes the rainfall event. In this analysis, the vector $\underline{\theta}$ will contain two elements, the average intensity, $\bar{i}$, and the storm duration, t_r. With the values of rainfall intensity and storm duration, the overland flow modelling predicts (perfectly) the resulting peak discharge. This sampling for the rainfall values is done for every storm; thus, the stochastic process of the flood discharges is a function of the stochastic process of the rainfall events and the determinstic runoff modelling.

It has been shown by Eagleson (1972) that there exists in the $\bar{i} - t_r$ plane a line of constant peak discharges, q_m, such that all combinations of $\bar{i}$ and t_r to the southwest of this boundary produce discharges less than q_m. This is shown in Figure 5.1. The probability of observing particular values of $\bar{i}$, t_r is given by their joint probability density function, $f(\bar{i},t_r)$. Finding the cumulative density function for the peak discharge from a rainfall event is equivalent to finding the cumulative density function for the

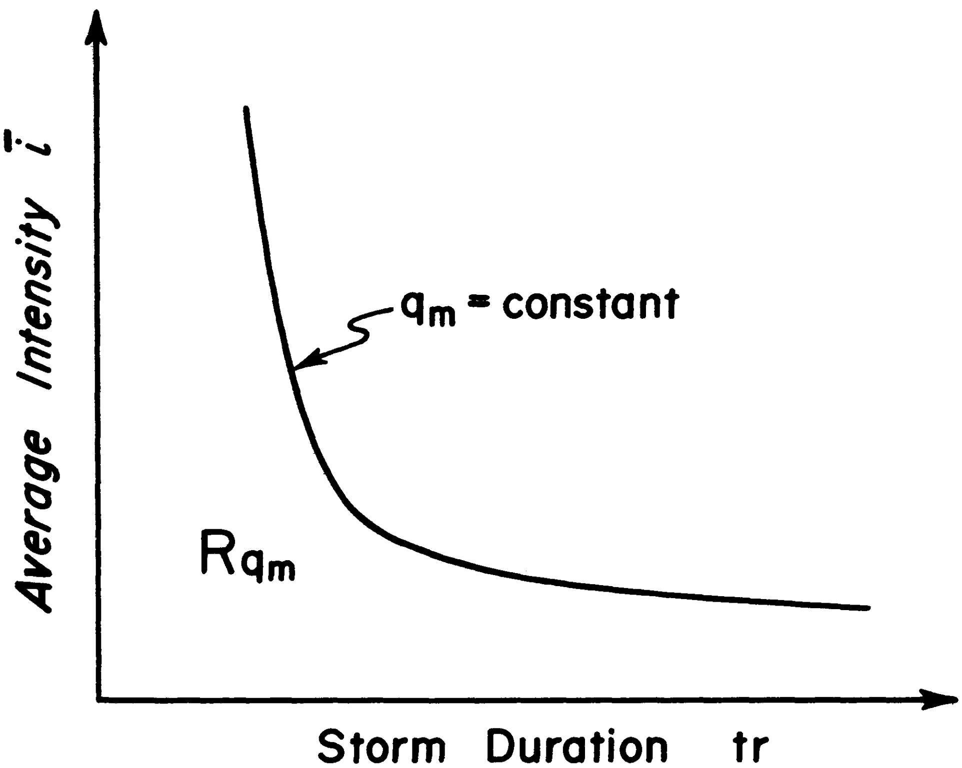


Figure 5.1 $\bar{i}$, t_r Plane Showing Peak Discharge

rainfall parameters, $\bar{i}$ and t_r, that produce the peak discharge q_m. This is the problem of finding the volume under the joint density function of $\bar{i}$, t_r for the region Rq_m. This region has boundaries $\bar{i} = 0$, $t_r = 0$, and q_m = constant. The volume under $f(\bar{i}, t_r)$ for this region is found by solving the integration

$$F(q_m) = \int_{Rq_m} f(\bar{i},t_r)\, d\bar{i}\, dt_r \qquad (5.1)$$

The resulting volume is shown in Figure 5.2. The boundary q_m = constant is located by the modelling of the runoff, either by computer simulation or by analytical techniques. The shape and location of the boundary depends upon

1. The shape of the rainfall event,
2. the modelling of the catchment response, (overland flow) to the rainfall,
3. the values of the parameters in the catchment model.

Traditionally, the assessment of $F(q_m)$ has been to pick a storm pattern, choose a runoff model and set the parameters with the 'best' available estimates. Such a procedure does not account for the uncertainty in the region Rq_m due to parameter uncertainty.

Now consider the case where the parameters are unknown and can be treated as random variables. Such uncertain parameters can be divided into two categories. The first category consists of those parameters that are fixed but unknown. A 'true' value is thought to exist and, through more data, better information may be obtained.

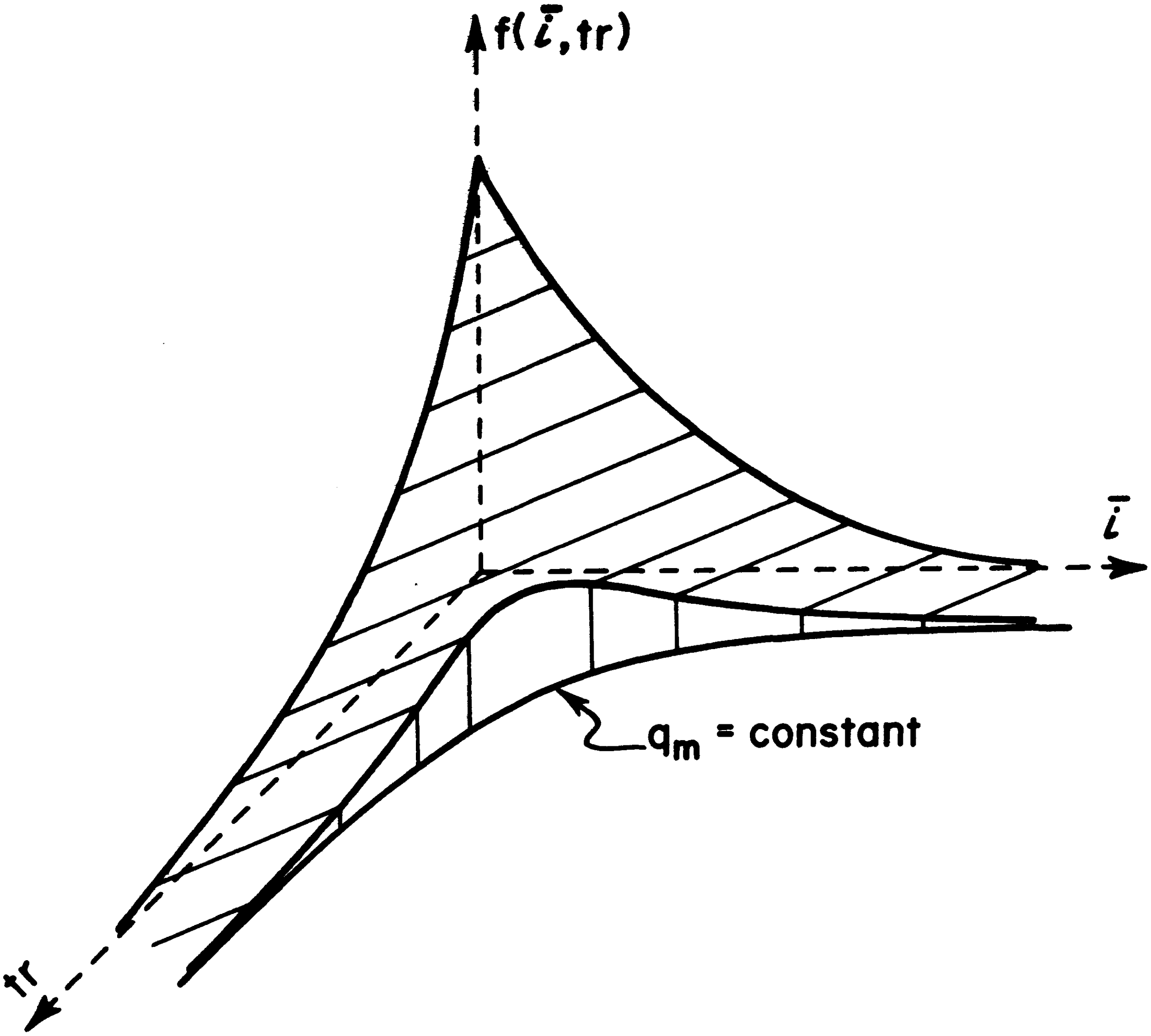


Figure 5.2 Graphical Representation of $F(q_m)$.

Such variables would be the parameters of the rainfall pdf of intensity and duration or parameters of the runoff modelling, such as stream length or slope. The second category of uncertain parameters are those parameters that vary from rainfall event to rainfall event. Such a parameter would be infiltration. Let infiltration be modelled as a constant water loss, ϕ , over the rainfall event. Then the value of ϕ can be viewed as a stochastic process along with the rainfall event, and these two processes join together to generate peak discharges.

Again, handling these uncertain parameters can be viewed as an urn sampling problem. The difference between the two types of uncertain parameters is important because it governs at what point "sampling" is done. Assume for the moment that the only uncertain parameters are those that vary from rainfall event to rainfall event and that the water loss ϕ is the only uncertain parameter. Then the sampling would be to choose from one urn a value of the rainfall intensity and storm duration set. From a second urn, a value for the water loss is obtained, which, combined with the runoff model and the rainfall values, produces the flood peak. The cumulative for the flood peak that accounts for the uncertainty in ϕ can be calculated by

$$\tilde{F}(q_m) = \int_{\phi} f(\phi)\, d\phi \cdot \int_{Rq_m|\phi} f(\bar{i},t_r)\, d\bar{i}\, dt_r \qquad (5.2)$$

where $f(\phi)$ is the density function for the water loss

$Rq_m|\phi$ is the region in the $\bar{i} - t_r$ plane where the flood peak is less than or equal to q_m. This region is conditional upon ϕ .

The cumulative $\tilde{F}(q_m)$ will be called the Bayesian cumulative of q_m and is the expected value of the cumulative, taking parameter uncertainty into account.

When there exist parameters that are fixed but uncertain, Equation (5.2) is followed, but conditional upon the uncertain parameters. Then, at the end, the cumulative is weighed by the probability density function for the fixed but uncertain parameters. For example, assume that the rainfall pdf has two parameters, ξ and λ , which are unknown. Since it is assumed that the pdf is fixed but uncertain, the parameter uncertainty is introduced at the end. If the cumulative of Q_{max} is desired, where Q_{max} is the largest of n events and where the events are independent random occurrences, then $\tilde{F}_{Q_{max}}$ is found from

$$\tilde{F}_{Q_{max}} = \int_{\xi,\lambda} \tilde{F}^n (q_m | \xi,\lambda) \cdot f(\xi,\lambda) \, d\xi \, d\lambda \qquad (5.3)$$

where $\tilde{F}(q_m|\xi,\lambda)$ is the cumulative of the flood peak q_m conditional upon the parameters ξ and λ and found from

$$\tilde{F}(q_m|\xi,\lambda) = \int_{\phi} f(\phi) \, d\phi \int_{Rq_m|\phi} f(\bar{i}, t|\xi,\lambda) \, d\bar{i} \, dt_r$$

$f(\xi,\lambda)$ is the probability density function for the fixed but uncertain rainfall parameters.

The analysis of the rainfall distribution in a Bayesian framework within the rainfall runoff analysis must be done at the end. The effect of parameter uncertainty is to introduce uncertainty as to the location of the boundary q_m = constant. The fixed but unknown parameters can be viewed as an uncertainty in the boundary due to a lack of information. The parameters that vary from event to event cause shifting in the boundary due to the interaction of stochastic processes.

There are also two density functions of interest that can be evaluated. The first is the marginal distribution of the exceedance probability at q_m. The exceedance probability, $G(q)$, is the probability of observing a flood greater than q_m. The marginal distribution of the exceedance probability, conditional upon the flood level q_m, will be written as $f[G(q_m)]$. The second marginal distribution of interest is the probability density function on the flood discharges, conditional upon an exceedance probability level; it will be written as $f[q|G(q)]$. The two density functions are displayed in Figure 5.3. These density functions are useful in performing sensitivity analysis on $G(q_m)$ and q_m due to the uncertainty in Rq_m.

They may play a larger role if, in a decision problem, the utility function for the decision set $\underline{A}$ depended upon the exceedance probability of the design discharge q_d. Under these conditions, the expected utility of a decision act, a_i, from the set $\underline{A}$, is given by

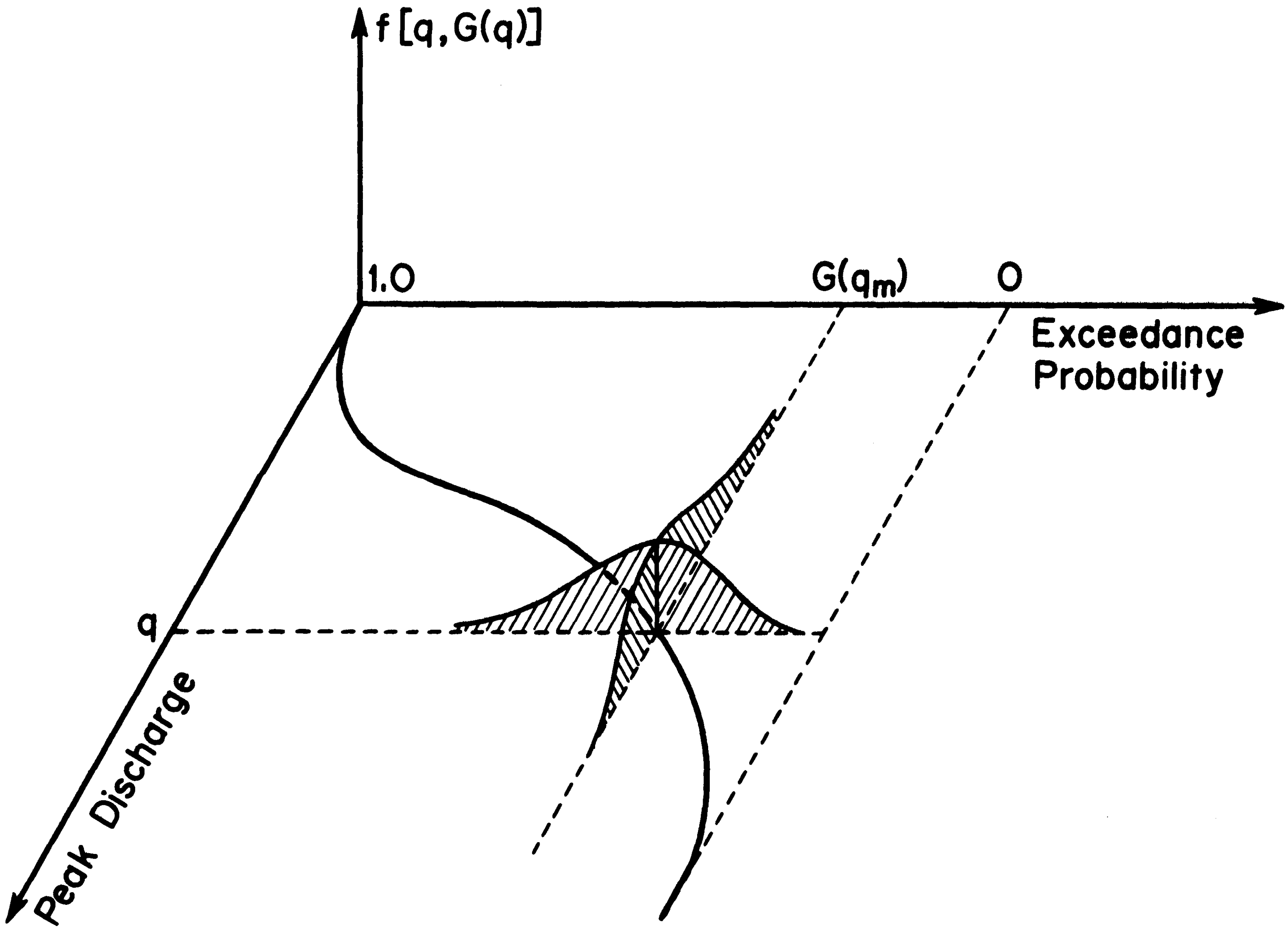


Figure 5.3 Graphical Representation in the Joint pdf for q and G(q) in the q, G(q) Plane

$$E[u(a_i)] = \int u[a_i, G(q_d)] \cdot f[G(q_d)]\, dG(q_d) \qquad (5.4)$$

The evaluation of (5.4) requires the density function $f[G(q_d)]$.

5.3 Derivation of the Bayesian Flood Frequency Curve.

This section presents the analytical derivation of the marginal probability density functions for the exceedance probability, conditional upon a flood magnitude, $f[G(q_m)]$, and the marginal probability density function of the flood discharges, conditional upon the exceedance probability level, $f[q|G(q)]$. To fully focus upon the methodological aspects of the analysis and to permit analytical derivation of the required equations, the following assumptions will be employed:

1. All parameters will be known with certainty, except ϕ, the temporally and spatially averaged water loss rate of the rainfall event.
2. The rainfall event has a rectangular interior pattern.
3. Following Eagleson (1972), the joint probability density function for the average rainfall intensity $\bar{i}$ and storm duration t_r is of the form

$$f(\bar{i},t_r) = \frac{\lambda\beta}{K} \exp\,[-\lambda\bar{i} - \frac{\beta}{K} t_r] \qquad (5.5)$$

where K is a factor to reduce point rainstorm depths to areal averages for events of common probability. λ and β are parameters of the point rainfall density function. All rainfall parameters are assumed known with certainty.

4. The response of the catchment to a rainfall event will follow Eagleson (1972). Eagleson analytically derived the peak discharge from a catchment by applying kinematic wave theory under the assumptions that the catchment can be modelled by an idealized flow plane and that the time of concentration of the stream is larger than the time of concentration for the catchment. Eagleson's catchment response will be used to define the boundary q_m = constant.

The extension to a simulation model is straightforward. The model will define lines of constant peak discharges in the $\bar{i} - t_r$ plane for given values of ϕ . The volume under the $f(\bar{i},t_r)$ surface, for the region Rq_m, can be found either by analytical procedures or by numerical procedures, depending upon the form $f(\bar{i},t_r)$ and the representation of the boundary of constant peak discharge.

Eagleson approximates the boundary q_m = constant by a function of the form

$$g(i) = B/i^m \tag{5.6}$$

taking $m = 1/2$

$$\text{where} \quad B = 2.97\left(\frac{A_r}{\alpha_c L_s}\right)^{1/2}\left(1 - \frac{655\ \alpha_s^{4/3}\ A_r^2}{\alpha_c L_s\ q_m}\right)^{1/2} \tag{5.7}$$

A_r is area contributing to direct runoff α_c and α_s are parameters of the catchment.

L_s is the stream length

$\bar{i} = i_e - q_m/645\ A_r$; i_e being the average excess rainfall intensity

For storm durations greater than the sum of the times of concentration for the catchment and the stream

$$q_m = 645\ A_r\ i_e \qquad (5.7)$$

The analysis here considers all rainfall events whereas Eagleson only considered events that produced direct runoff (excess rainfall events).

To find the cumulative for the peak discharge, $\tilde{F}(q_m)$ Equation (5.2) is applied. The inner integration is over the rainfall probability density function. The limits of integration cover the region Rq_m, which is a function of uncertain water loss parameters, ϕ. In fact, the region Rq_m in the $\bar{i} - t_r$ plane now becomes a volume in the $\bar{i} - t_r - \phi$ space, and the integration for $\tilde{F}(q_m)$ is done first for Rq_m, conditional upon ϕ. The integration over ϕ is then performed. Figure 5.4 shows the constant boundary in the $\bar{i} - t_r - \phi$ space and the volume, Rq_m, where the discharge is less than or equal to q_m.

The integration of Equation (5.2), over the rainfall pdf, yields $\tilde{F}(q_m|\phi)$, and is evaluated by

$$\tilde{F}(q_m|\phi) = \int_{Rq_m|\phi} f(\bar{i},t_r)\ d\bar{i}\ dt_r \qquad (5.8)$$

The region $Rq_m|\phi$ can be broken into two areas. The first

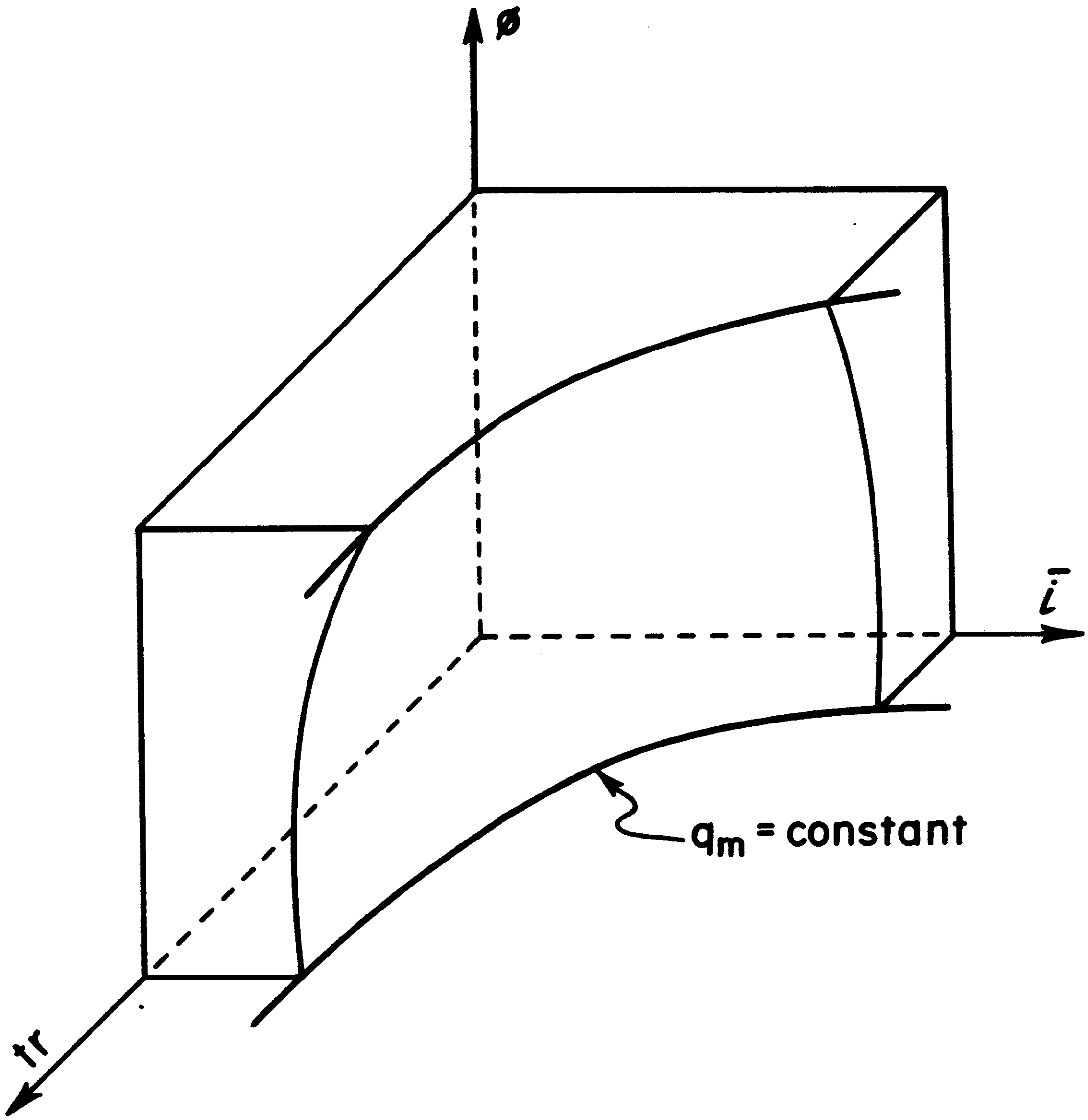


Figure 5.4: $\bar{i}$, t_r ϕ Space Showing Peak Discharge

has the boundaries

$$t_r = 0$$

$$t_r = \infty$$

$$\bar{i} = 0$$

$$\bar{i} = \frac{q_m}{645\ A_r} + \phi$$

The solution to (5.8) for these limits of integration will be represented by I_1. The solution for the following limits of integration will be represented by I_2. These limits are

$$\bar{i} = \frac{q_m}{645\ A_r} + \phi$$

$$\bar{i} = \infty$$

$$t_r = g(i_o)$$

where $g(i_o)$ is a function of the form similar to (5.6). The two areas of integration are shown in Figure 5.5 and are similar to the two regions Eagleson used to solve his function.

$$I_1 = \int_o^{\infty} dt_r \int_o^{q_m/645\ A_r+\phi} \frac{\beta\lambda}{K} \exp\ (-\ \lambda t_r - \frac{\beta}{K}\ \bar{i})\ d\bar{i}$$

$$= 1 - \exp\ (-\beta q_m/645\ KA_r - \beta\phi/K) \qquad (5.9)$$

$$I_2 = \int_{q_m/645\ A_r+\ \phi}^{\infty} d\bar{i} \int_o^{g(\bar{i})} \frac{\beta\lambda}{K}\ (-\lambda t_r - \frac{\beta}{K}\ \bar{i})\ dt_r \qquad (5.10)$$

where $t_r = g(i_o)$. (5.11)

Letting $i_o = \bar{i} - (q_m/645\ A_r + \phi)$ (5.12)

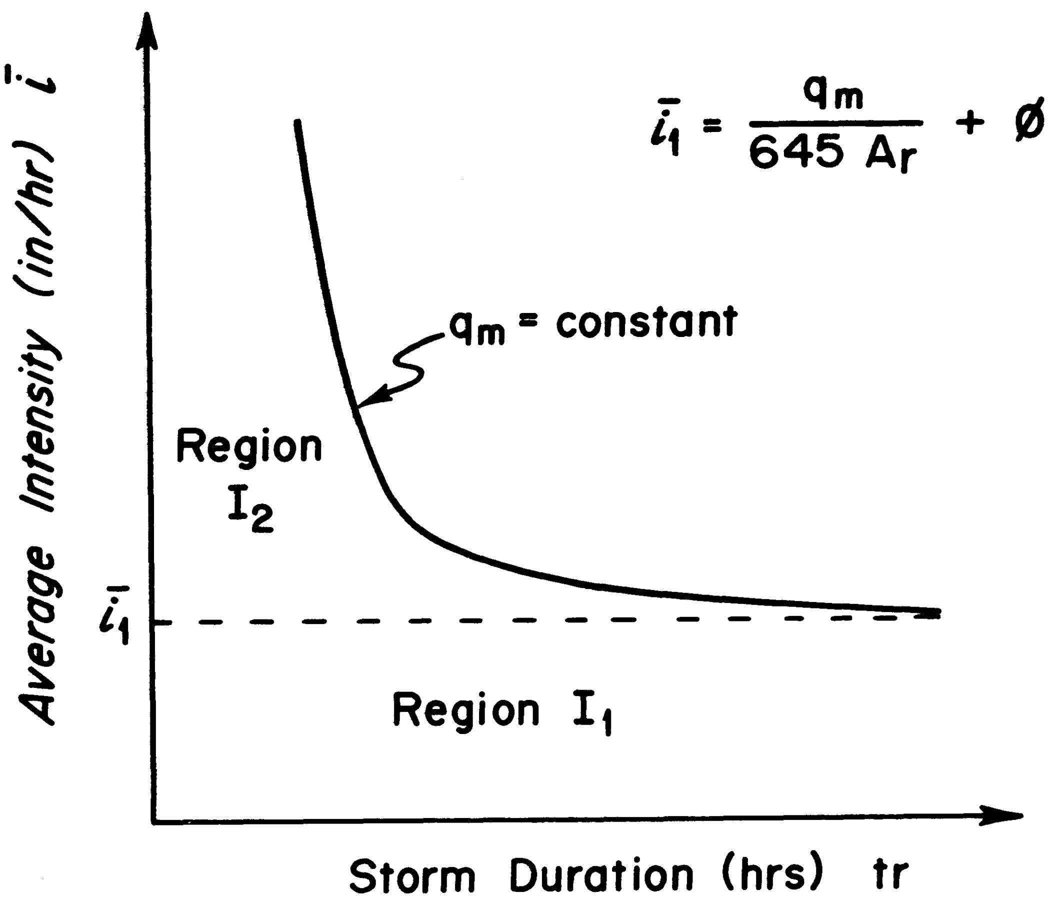


Figure 5.5: Regions of Integration in the $\bar{i}$, t_r Plane

Then, (5.10), becomes,

$$I_2 = \int_o^\infty di_o \int_o^{g(i_o)} \frac{\beta\lambda}{K} \exp\,[-\,t_r - \frac{\beta}{K}\,(i_o + q_m/645\,A_r + \phi)]dt_r \qquad (5.13)$$

which can be simplified to

$$I_2 = \exp(-\frac{\beta}{K}\,\frac{q_m}{645\,A_r} - \frac{\beta}{K}\,\phi) \cdot \left[1 - \frac{\beta}{K}\int_o^\infty \exp\,[-\frac{\beta i_o}{K} - \lambda\,g(i_o)]\,di_o\right]$$

(5.14)

when $g(i_o)$ is of the form of (5.6) then (5.14) integrates to

$$I_2 = \exp\,(-\frac{\beta q_m}{645\,KA_r} - \frac{\beta}{K}\,\phi) \cdot (1 - I_o) \qquad (5.15)$$

where $I_o \cong e^{-\sigma/m}\,\sigma^{-\sigma+1}\,\Gamma(\sigma)$

$$\sigma = \left[2.21\,\frac{\beta\lambda^2 A_r}{K\alpha_c\,L_s}\,(1 - \frac{655\alpha_s^{4/3}\,A_r^{\,2}}{\alpha_c\,L_s^{\,3} q_m^{1/3}})^{1/3}\right]^{1/2}$$

Thus

$$F(q_m) = 1 - I_o \cdot \exp\,(-\frac{\beta q_m}{645\,A_r K} - \frac{\beta}{K}\,\phi\,) \qquad (5.16)$$

When considering the cumulative density function for q_m, conditional upon an excise rainfall event occurring, then (5.16) reduces to Eagleson's expression.

Often, decision makers are interested in the flood exceedance probability, $G(q_m) = 1 - F(q_m)$. Then, from (5.16), $G(q_m)$ is

$$G(q_m) = I_o \cdot \exp\,(-\frac{\beta q_m}{K\,645\,A_r} - \frac{\beta}{K}\,\phi\,) \qquad (5.17)$$

Equation (5.12) provides a relationship between the exceedance probability for a given flood peak, q_m, and the water loss parameter, ϕ. If two random variables are functionally related (for example $y = g(x)$) and if the function is monotonic and continuous, then the following relationships hold

$$E[y^n] = \int_x g^n(x) \cdot f(x)\, dx \tag{5.18}$$

$$f(y) = f(x) \cdot \left| \frac{dx}{dy} \right|$$

These relationships provide a procedure to obtain the marginal probability density function as well as the moments for the exceedance probability $G(q_m)$, given the peak discharge, and for the peak discharge, q, conditional upon the exceedance probability. These marginal density functions reflect the uncertainty in ϕ.

The form of these distributions depends upon the probability density function for ϕ, $f(\phi)$. Three forms will be examined. These are: $f(\phi)$ as a uniform pdf, a gamma-1 pdf, and an exponential. The latter is really a special case of the gamma-1.

5.3.1 Water Loss ϕ, uniformly distributed.

Let $f(\phi)$ be represented by a uniform probability density function between ϕ_o and ϕ^o,

$$f(\phi) = \frac{1}{(\phi^{\circ}-\phi_o)} \qquad \phi_o \leq \phi \leq \phi^o \tag{5.19}$$

$$= \text{otherwise}$$

and let $y = G(q_m)$. Then the Jacobian from (5.16) is

$$\left| \frac{dy}{d\phi} \right| = \frac{C\beta}{K} e^{-\phi\beta/K} \tag{5.20}$$

where $C = I_o \cdot \exp\left(\frac{-\beta q_m}{K\ 645\ A_r}\right)$

$$f(y) = \frac{1}{(\phi^\circ - \phi_o)} \frac{K}{\beta y} \quad \text{for } C \cdot e^{-\phi^\circ\beta/K} \le y \le C \cdot e^{-\phi_o\beta/K}$$

$$= 0 \qquad \text{otherwise} \tag{5.21}$$

The first two moments are

$$E[y] = \frac{CK}{\beta(\phi^\circ - \phi_o)} \left[e^{-\phi_o\beta/K} - e^{-\phi^\circ\beta/K} \right] \tag{5.22}$$

$$E[y^2] = \frac{C^2 K}{2\beta(\phi^\circ - \phi_o)} \left[e^{-2\phi_o\beta/K} - e^{-2\phi^\circ\beta/K} \right] \tag{5.23}$$

The decision maker is not only interested in the distribution of the exceedance probability at a particular flood discharge level, but, given an exceedance probability, he is also interested in the distribution of the flood discharges. This marginal probability density function can be found from (5.16) and (5.18). Due to the complex nature of the discharge in (5.16), analytical derivation is only possible if the following assumption is valid: for a particular basin, I_o is constant over the range of flood discharges that are of interest. Table 5.1 shows that this assumption is a reasonable one, then, the

Table 5.1

Discharge (cfs)	σ	I_o
100	.60199	.36384
1000	.6190	.3496
5000	.6249	.34465
10,000	.62661	.3432

(For catchment and rainfall parameters as given in Table 5.2.)

Table 5.1 Values of I_o for Various Peak Discharges

Jacobian, $|dq/d\phi|$ is, from Equation (5.16),

$$\left| \frac{dq}{d\phi} \right| = 645 \, A_r \tag{5.24}$$

The limits on q, for the derived distribution, may be obtained by rewriting Equation (5.16) as

$$\phi = \frac{K}{\beta} \ln \left(\frac{I_o}{y} \right) - \frac{q_m}{645 \, A_r} \tag{5.25}$$

For y [= $G(q_m)$] a constant and for no water loss ($\phi = 0$) q_m is a maximum and equal to

$$q_m = 645 \, A_r \frac{K}{\beta} \ln \left(\frac{I_o}{y} \right) \tag{5.26}$$

As the water loss increases, the discharge from the rainfall event must decrease until, at some value of ϕ, ϕ_m, there is no excess rainfall and no runoff. This value is:

$$\phi_m = \frac{K}{\beta} \ln \left(\frac{I_o}{y} \right) \tag{5.27}$$

The probability that $q_m = 0$ is the probability that ϕ is greater than or equal to ϕ_m. The spike for $f_Q(q=0)$ can be calculated by

$$f_Q(q = 0) = P(\phi \geq \phi_m) = \int_{\phi_m}^{\infty} f(\phi) \, d\phi \tag{5.28}$$

and the density function for q, $q > 0$, will be the derived density function from (5.18) with limits

$$0 \leq q \leq 645\, A_r \frac{K}{\beta} \ln\left(\frac{I_o}{y} \right) \tag{5.29}$$

With (5.18), (5.19), and (5.24) the distribution f(q) is

$$f(q) = \frac{1}{(\phi^\circ - \phi_o) \cdot 645\, A_r} \tag{5.30}$$

and has limits

$$645\, A_r \left[\frac{K}{\beta} \ln\left(\frac{I_o}{y} \right) - \phi^\circ \right] \leq q \leq 645\, A_r \left[\frac{K}{\beta} \ln\left(\frac{I_o}{y} \right) - \phi_o \right] \tag{5.31}$$

if $\phi^\circ < \frac{K}{\beta} \ln\left(\frac{I_o}{y} \right)$

If $\phi^\circ > \frac{K}{\beta} \ln\left(\frac{I_o}{y} \right)$ and $\phi_o < \frac{K}{\beta} \ln\left(\frac{I_o}{y} \right)$ then

the limits are

$$0 \leq q \leq 645\, A_r \left[\frac{K}{\beta} \ln\left(\frac{I_o}{y} \right) - \phi_o \right] \tag{5.32}$$

for $f(q|q > 0)$. The spike at $q = 0$ may be found from (5.28) or from integrating (5.30) between the limits

$$0 \leq q \leq 645\, A_r \left[\phi^\circ - \frac{K}{\beta} \ln\left(\frac{I_o}{y} \right) \right] \tag{5.33}$$

The first two moments of f(q) are

$$E[q] = 645\, A_r \left[A - \frac{(\phi^\circ + \phi_o)}{2} \right] \tag{5.34}$$

with the constraint of $E[q] \geq 0$ and where

$$A = \frac{K}{\beta} \ln\left(\frac{I_o}{y} \right)$$

$$E[q^2] = (645\ A_r)^2 \cdot \ [A^2 - A(\phi^\circ + \phi_o) + \frac{1}{3}\ (\phi^\circ + \phi_o) \ - \frac{1}{3}\ \phi^\circ \phi_o] \tag{5.35}$$

again with $A = \frac{K}{\beta}\ \ell n \left(\frac{I_o}{y} \right)$

5.3.2 Water Loss ϕ, Gamma-1 Distributed.

Let ϕ be distributed with a probability density function of the form gamma-1, that is

$$f(\phi) = e^{-\alpha\phi}\ \phi^{r-1}\ \alpha^r\ \Gamma\ (r) \tag{5.36}$$

Using the same definitions for y and C as in the uniform pdf analysis and using the Jacobian as given in (5.20), then (5.18) gives

$$f(y) = A^r\ y^{A-1}\ C^{-A}\ [\ell n\ (\frac{C}{y})]^{r-1}\ /\ \Gamma(r) \tag{5.37}$$

where $A = K\alpha\ /\ \beta$ $\qquad 0 \leq y \leq 1$

The first two moments of y are

$$E[y] = C \left(\frac{\alpha}{\alpha + \frac{K}{\beta}} \right)^r \tag{5.38}$$

$$E[y^2] = C^2 \cdot \left(\frac{\alpha}{\alpha + \frac{2K}{\beta}} \right)^r \tag{5.39}$$

For the distribution of q for a given exceedance level $G(q_m)$, again the approximation that $I_o \cong$ constant must be made.

The Jacobian from (5.16) is as given in Equation (5.24) and with Equations (5.18) and (5.36)

$$f(q|q>0) = \frac{1}{645 A_r} \cdot \exp[-\alpha(A-q/645 A_r)] \cdot (A-q/645 A_r)^{r-1} \cdot \alpha^r/\Gamma(r)$$

$$0<q< 645 A_r \frac{K}{\beta} \ln\left(\frac{I_o}{y}\right) \quad (5.32)$$

where $A = \frac{K}{\beta} \ln\left(\frac{I_o}{y}\right)$

$f(q)$ has moments

$$E[q] = 645 A_r(A - r/\alpha) \quad (5.40)$$

$$E[q^2] = (645 A_r)^2 \cdot [A^2 - 2Ar/\alpha + r(r+1)/\alpha^2] \quad (5.41)$$

where $A = \frac{K}{\beta} \ln\left(\frac{I_o}{y}\right)$

5.3.3 Water Loss, ϕ, Exponentially Distributed.

Let ϕ be distributed exponentially. Then $f(\phi)$ is of the form,

$$f(\phi) = \alpha e^{-\alpha\phi} \quad (5.42)$$

which is a special case of the gamma-1 distribution when $r = 1$.

The marginal density function for the exceedance probability, with a peak discharge q_m and marginal density function for the discharge q at an exceedance level $G(q)$, may be found by the application of (5.16), (5.18) and (5.42). The marginals may also be found by taking the results from the gamma-1 analysis.

The results for the exceedance probability $(y \doteq G(q_m))$ are:

$$f(y) = Ay^{A-1} C^{-A} \qquad 0<y<1 \quad (5.43)$$

where $A = K\alpha/\beta$

$$E[y] = C\left(\frac{\alpha}{\alpha+\beta/K}\right) \tag{5.44}$$

$$E[y^2] = C\left(\frac{\alpha}{\alpha+2\beta/K}\right) \tag{5.45}$$

And for the discharge q, conditional upon q being greater than or equal to 0, the results are

$$f(q|q \geq 0) = \frac{\alpha}{645\ A_r} \exp\left(\frac{\alpha q}{645\ A_r}\right) \cdot \left(\frac{I_o}{y}\right)^{-\alpha K/\beta} \tag{5.46}$$

for $0 \leq q \leq 645\ A_r \frac{K}{\beta} \ln\left(\frac{I_o}{y}\right)$

$$E[q] = 645\ A_r\ [A - 1/\alpha] \tag{5.47}$$

where $A = \frac{K}{\beta} \ln\left(\frac{I_o}{y}\right)$

$$E[q^2] = (645\ A_r)^2[A^2 - 2A/\alpha + 2/\alpha^2] \tag{5.48}$$

where $A = \frac{K}{\beta} \ln\left(\frac{I_o}{y}\right)$

5.4 Recurrence Interval

The exceedance probability for the occurrence of flood events, $G(q_m)$, has been evaluated with the total series of independent rainfall events. Often hydrologists are interested in the exceedance probability of a flood peak as that peak relates to a partial duration series. When the number of flood events in this partial duration series equals N, the number of years of record, then the exceedance probability, for this particular partial duration series, can be found in the following manner, (Eagleson, 1972).

Consider a record of N years, which contains, on the average, θ rainfall events per year. There will be θN flood events, some of which will have a maximum discharge equal to 0, due to no excess rainfall. The r^{th} most severe event of the complete series will have an exceedance probability of

$$G(q_{mr}) = \frac{r}{\theta N+1} \tag{5.49}$$

Now consider the annual exceedance series which is composed of the N largest flood events from the set of θN. The exceedance probability of q_{mr}, from the annual exceedance series, is

$$P[q_m > q_{mr}] = \frac{r}{N+1} = \frac{1}{T_e} \tag{5.50}$$

where T_e is the recurrence interval measured in years. For $r \leq N$ (5.49) and (5.50) can be combined to give

$$\frac{1}{T_e} = \theta \cdot G(q_m) \tag{5.51}$$

assuming $N \gg 1$.

Equation (5.51) is used in the next section to compare the flood return periods obtained by the different modelling assumptions of the water loss parameter ϕ.

5.5 Example Application

The results of Section 5.3 and 5.4 can be used to determine the effect of uncertainty in the water loss parameter, ϕ, upon the flood frequency curve. The expected frequency curve for a hypothetical catchment, with parameters as given in Table 5.2, will be determined for the three different probability modelling assumptions of ϕ . An indication of the variance in the process will be obtained by plotting the expected exceedance probability curve, $E[G(q_m)]$, with the expected exceedance probability curve plus and minus one standard deviation. These curves will be from the annual exceedance series, that is, a partial duration series of a length equal to the number of years of record. It should be visualized that there exists a surface in the $G(q_m)- q_m$ plane. This surface represents the joint probability density function. The three curves, $E[G(q_m)]$, $E[G(q_m)] + \sigma$, $E[G(q_m)] -\sigma$ represent three contours. For comparison, the frequency curve from the analysis which assumes ϕ is deterministic is also presented. In this analysis, the value of ϕ chosen is the mean value of $f(\phi)$.

Figure 5.6 and 5.7 are for the case where the water loss is uniformly distributed with means $\overline{\Phi}$ equal to .03 in/hr and .05 in/hr respectively. Figure 5.8 and 5.9 are for the case where $f(\phi)$ is exponential with means of .03 in/hr and .05 in/hr respectively. Figures 5.10 and 5.11 are for $f(\phi)$ gamma-1 distributed with a mean, $\overline{\Phi}$, equal to .03 in/hr and coefficient of variation equal to .577 and .41 respectively. Figures 5.12, 5.13, 5.14, and 5.15 are for $f(\phi)$ gamma-1 distributed with mean, $\overline{\Phi}$, equal to .05 in/hr and coefficient of variation equal to .577, .477, .316, and .10 respectively.

The implications of the uncertainty in the frequency curve is

Table 5.2

Catchment and Rainfall Parameters

A_c	=	100 sq.mi.
A_r	=	$A_c / 3 = 33.333$ sq.mi.
L_s	=	$(3.\ A_c)^{1/2} = 17.32$ mi.
α_c	=	10 sec^{-1}
α_s	=	.1 sec^{-1}
β	=	30 hr/in.
λ	=	.13 hr^{-1}
K	=	.95 ($K = 1 - \exp[-1.1\ \lambda^{-1/4}] + \exp[-1.1\ \lambda^{-1/4} - .01\ A_r]$, Eagleson, 1972)
θ	=	109. events per year.

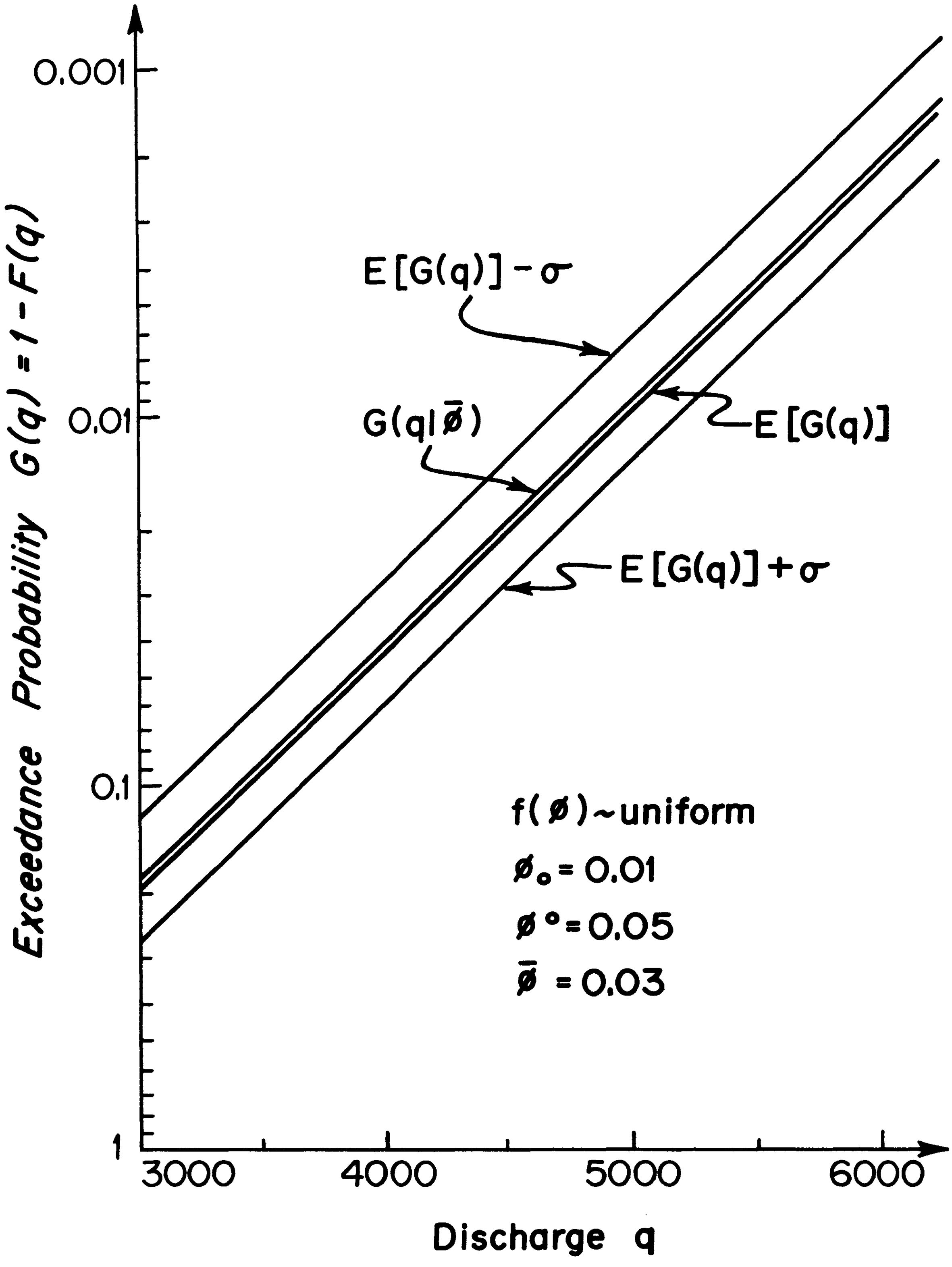


Figure 5.6: Frequency Curves for $f(\phi)$, Uniform with $\bar{\phi}$ = .03 in/hr

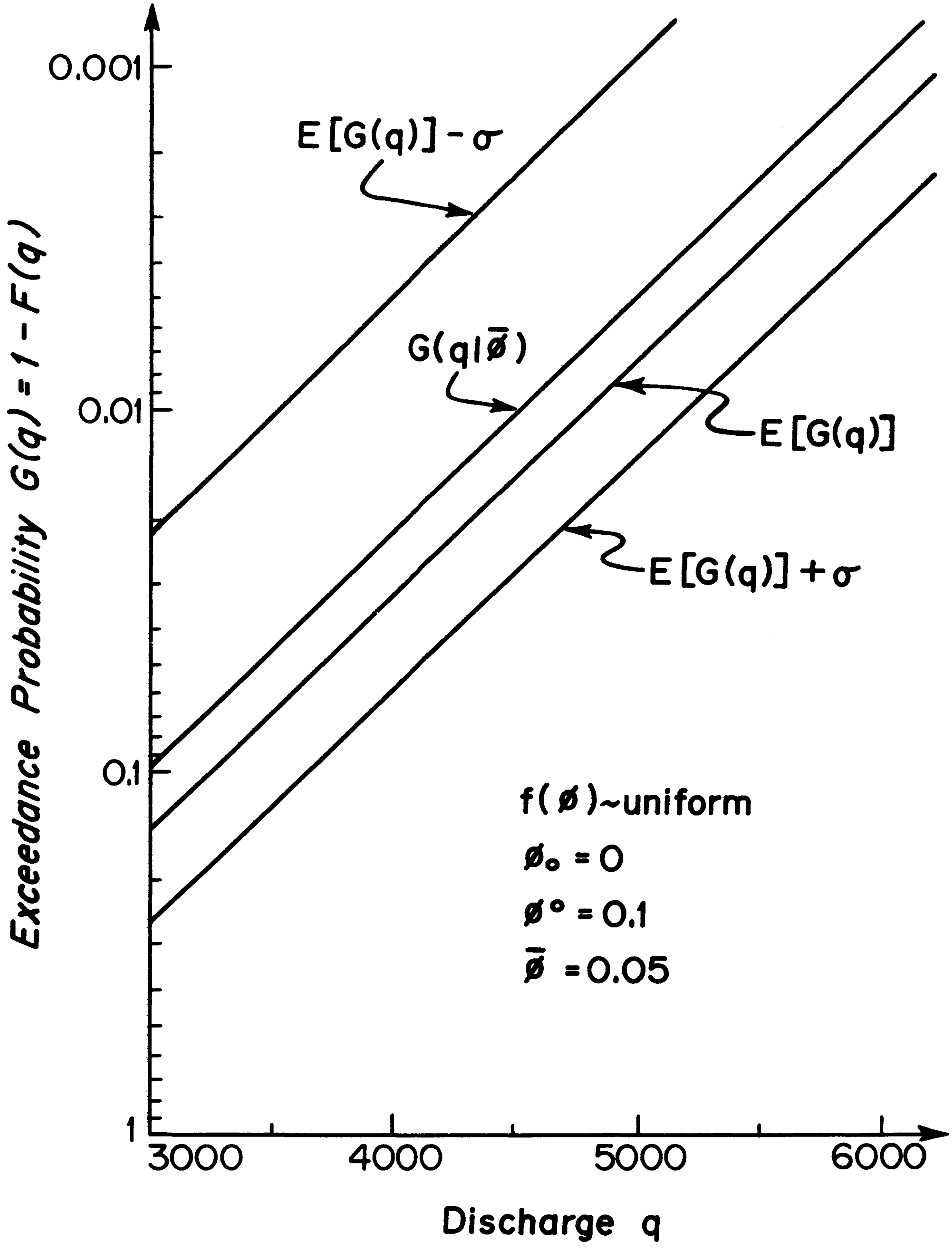


Figure 5.7: Frequency Curves for $f(\phi)$, Uniform with $\bar{\phi}$ = .05 in/hr

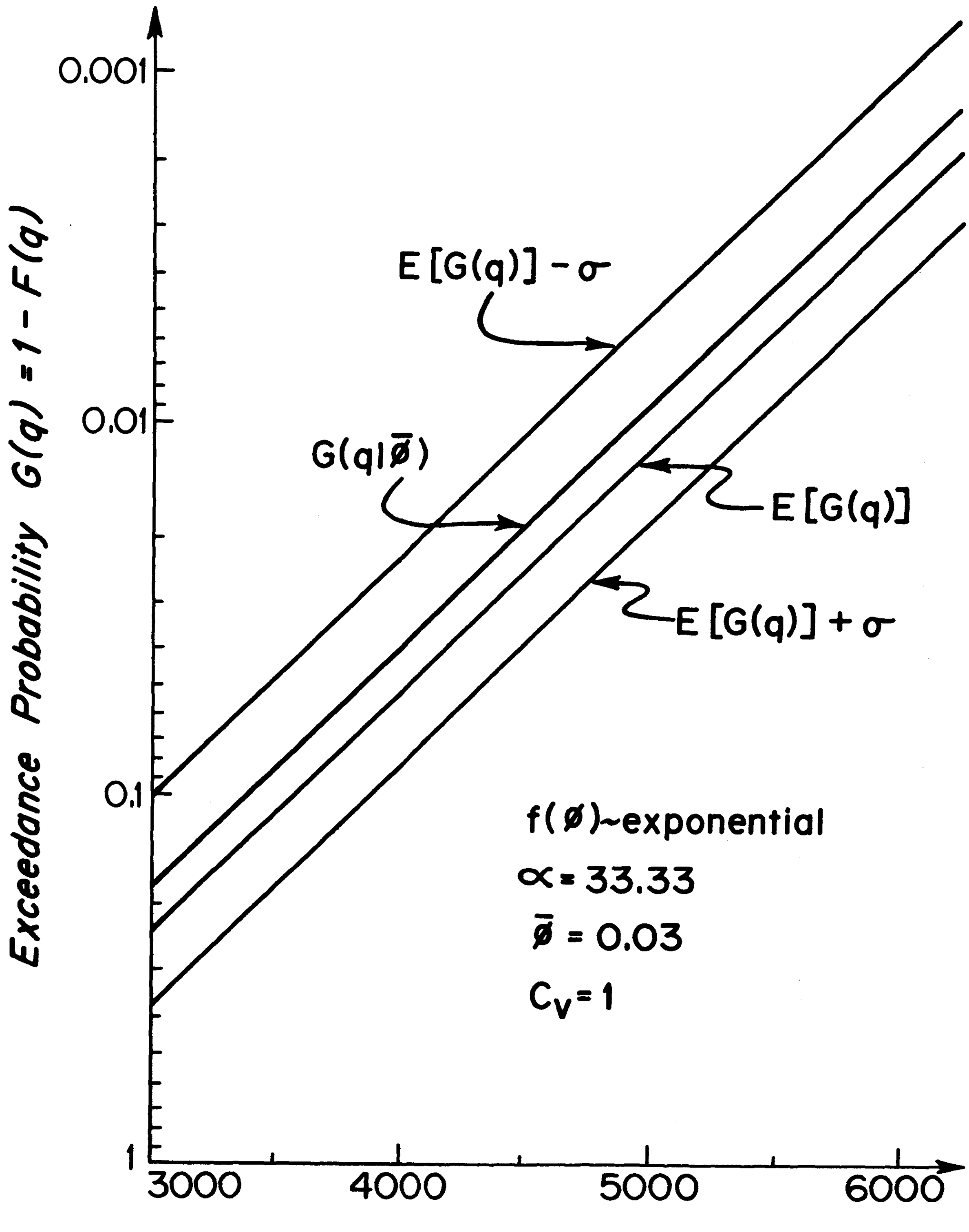


Figure 5.8: Frequency Curves for $f(\phi)$, Exponential with $\bar{\phi} = .03$

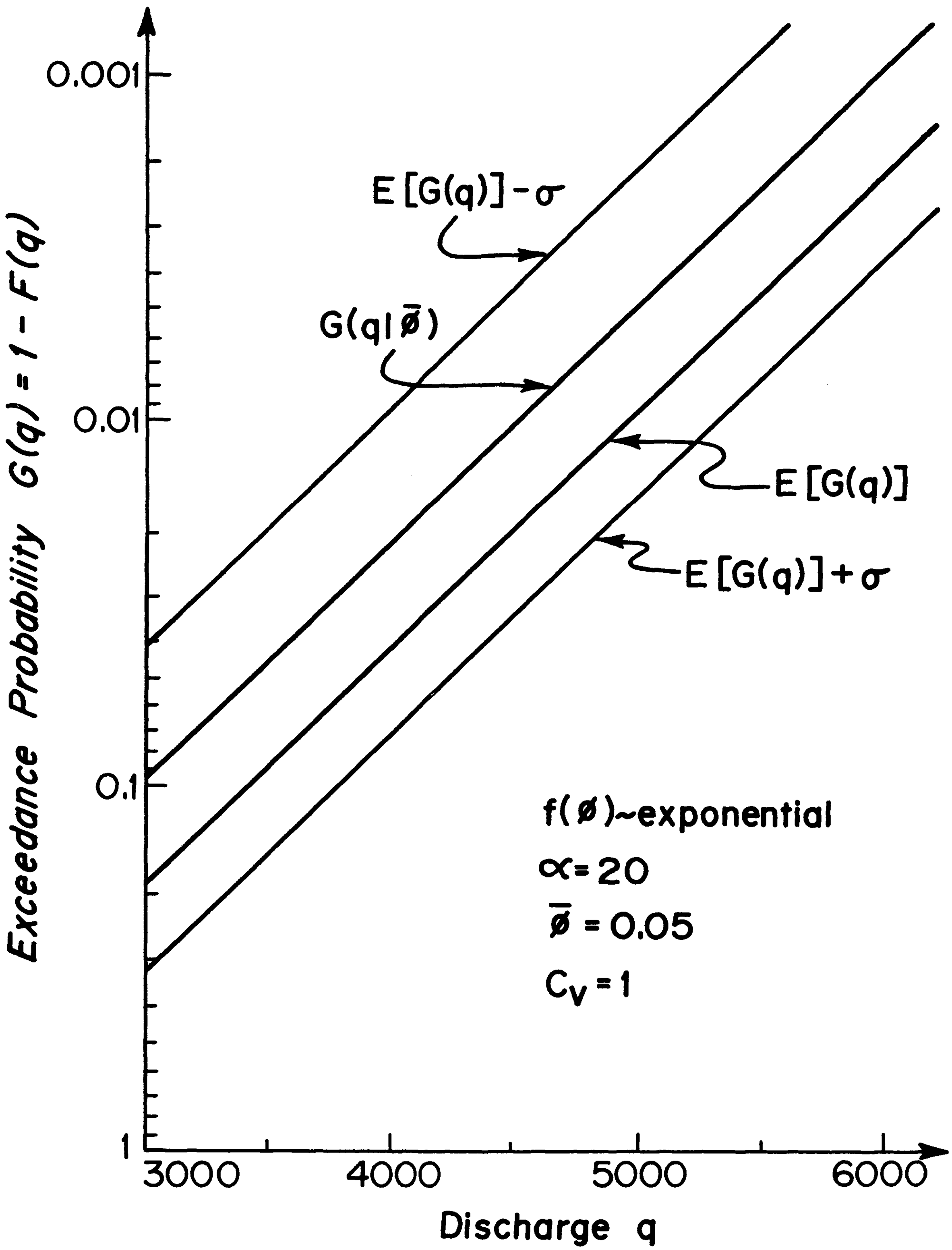


Figure 5.9: Frequency Curves for $f(\phi)$, Exponential with $\bar{\phi} = .05$

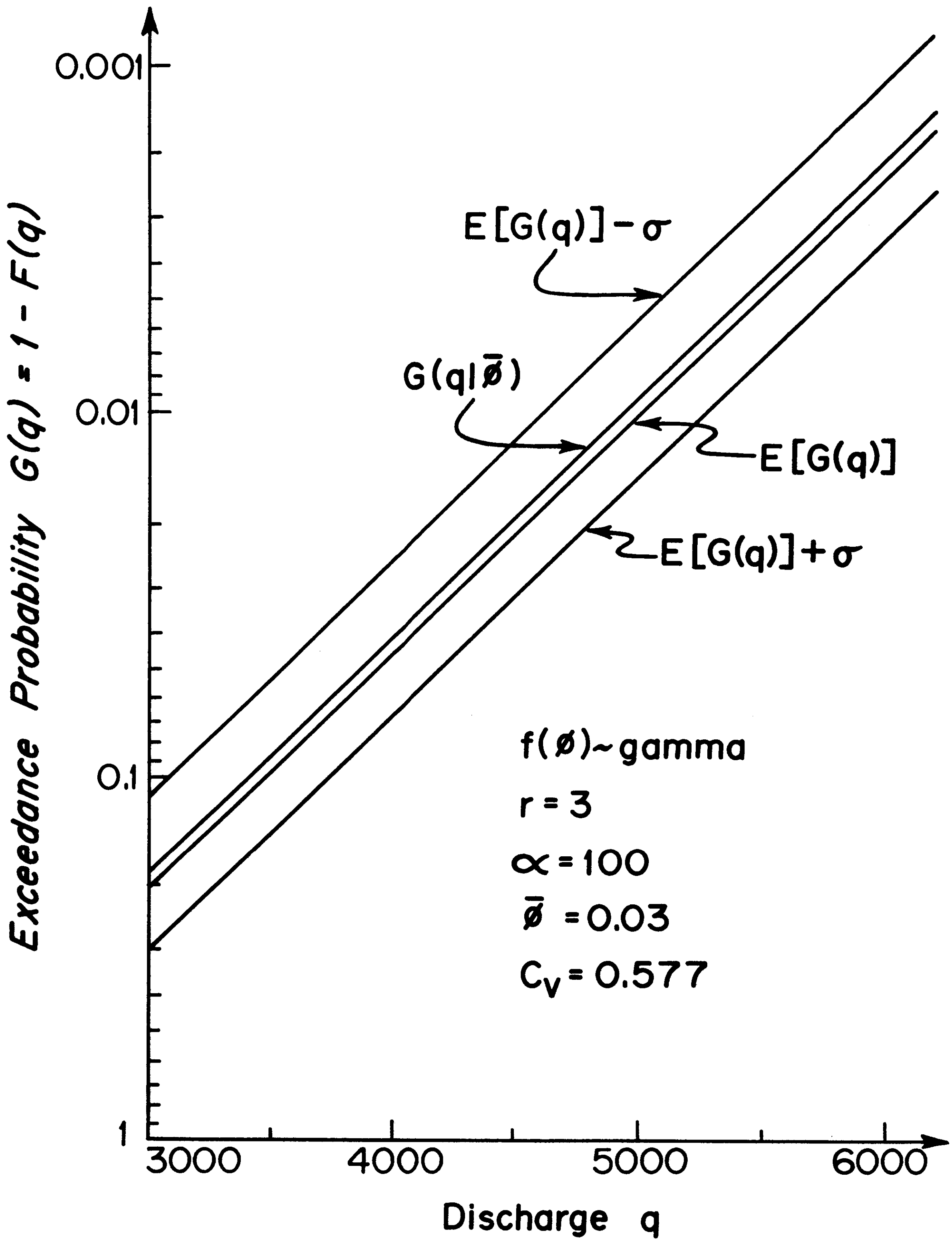


Figure 5.10: Frequency Curves for $f(\phi)$, Gamma-1 with $\bar{\phi}$ = .03 and C_v = .577

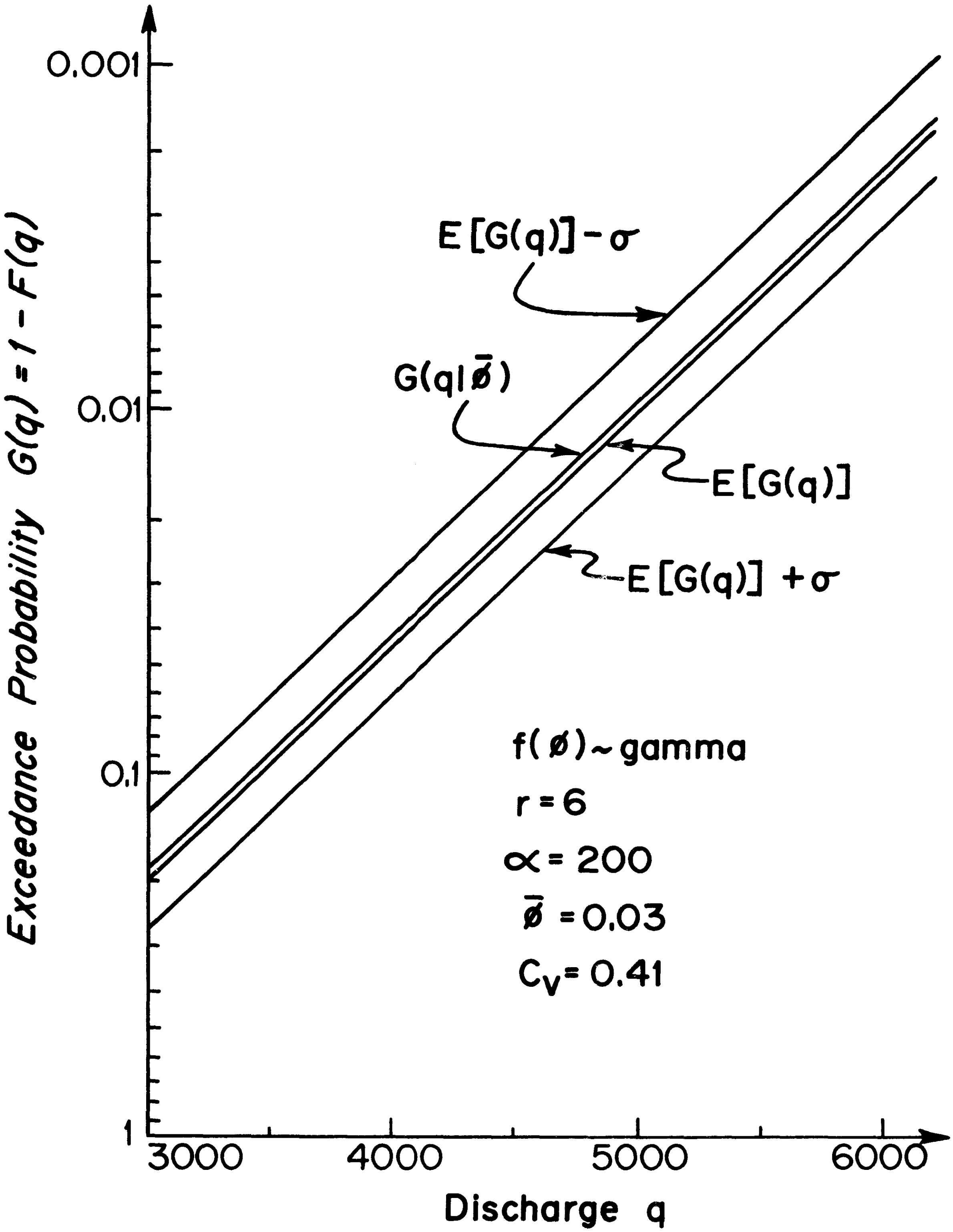


Figure 5.11: Frequency Curves for $f(\phi)$, Gamma-1 with $\bar{\phi}$ = .05 and C_v = .577

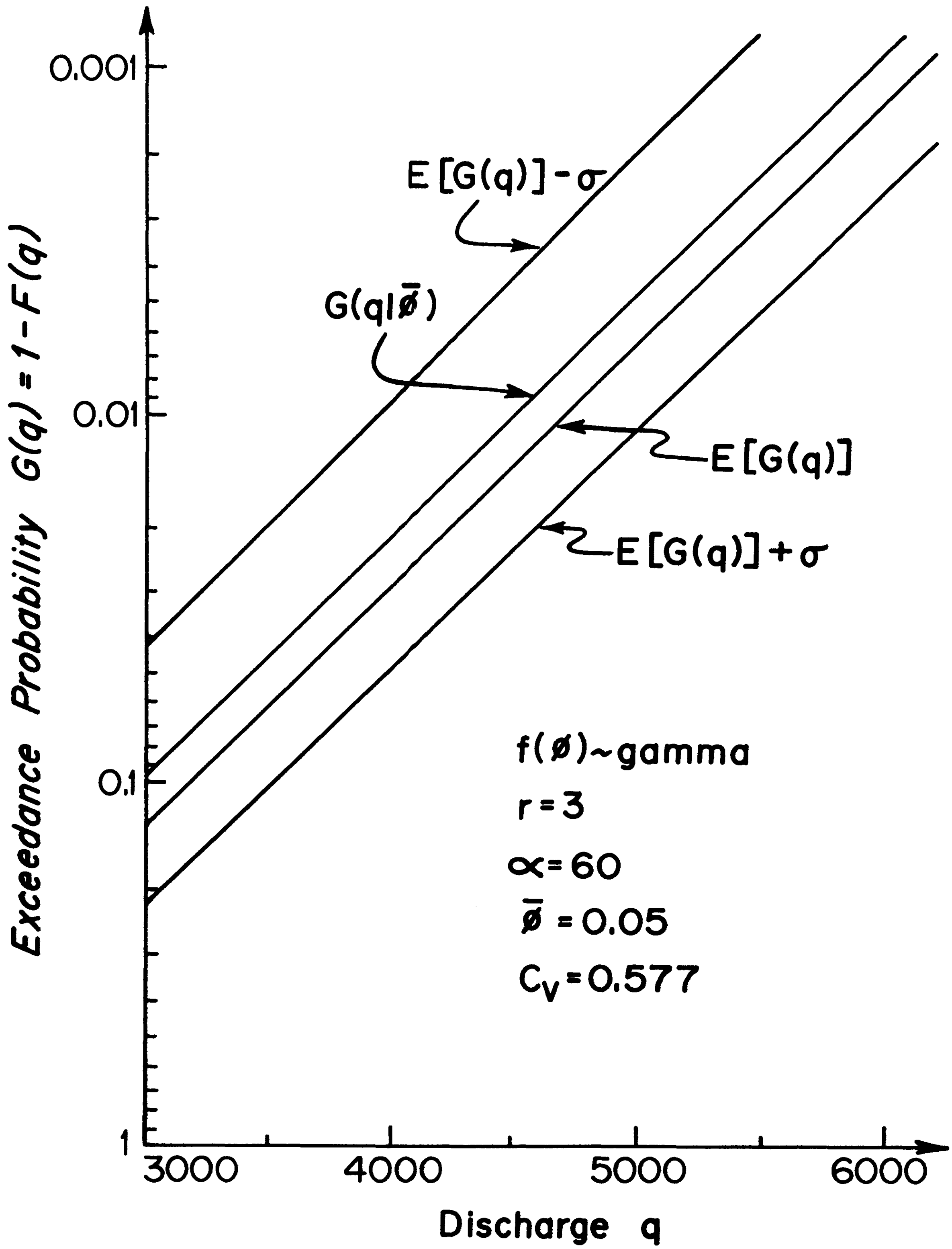


Figure 5.12: Frequency Curves for $f(\phi)$, Gamma-1 with $\bar{\phi}$ = .05 and C_v = . 577

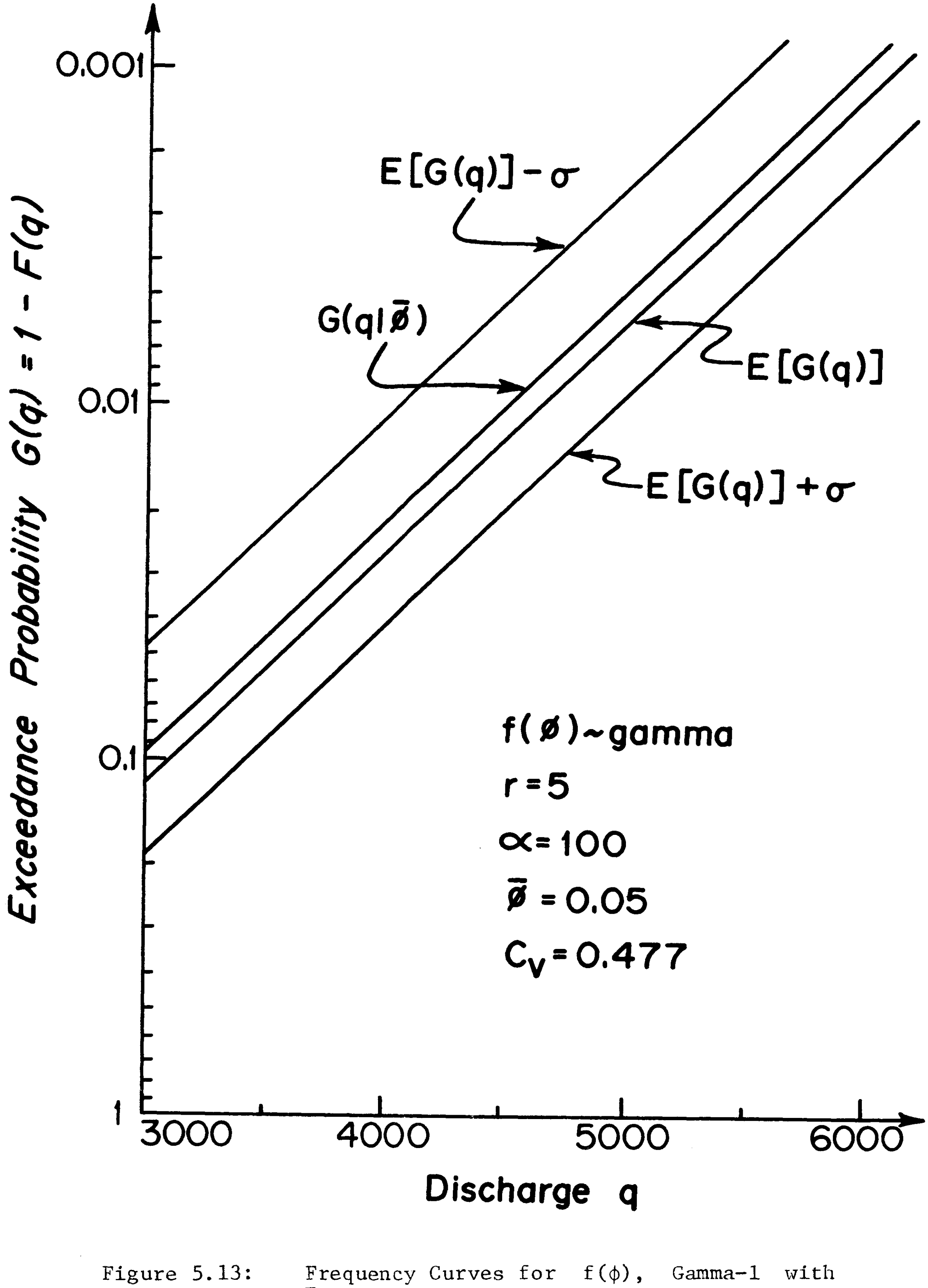


Figure 5.13: Frequency Curves for $f(\phi)$, Gamma-1 with $\bar{\phi}$ = .05 and C_v = .477

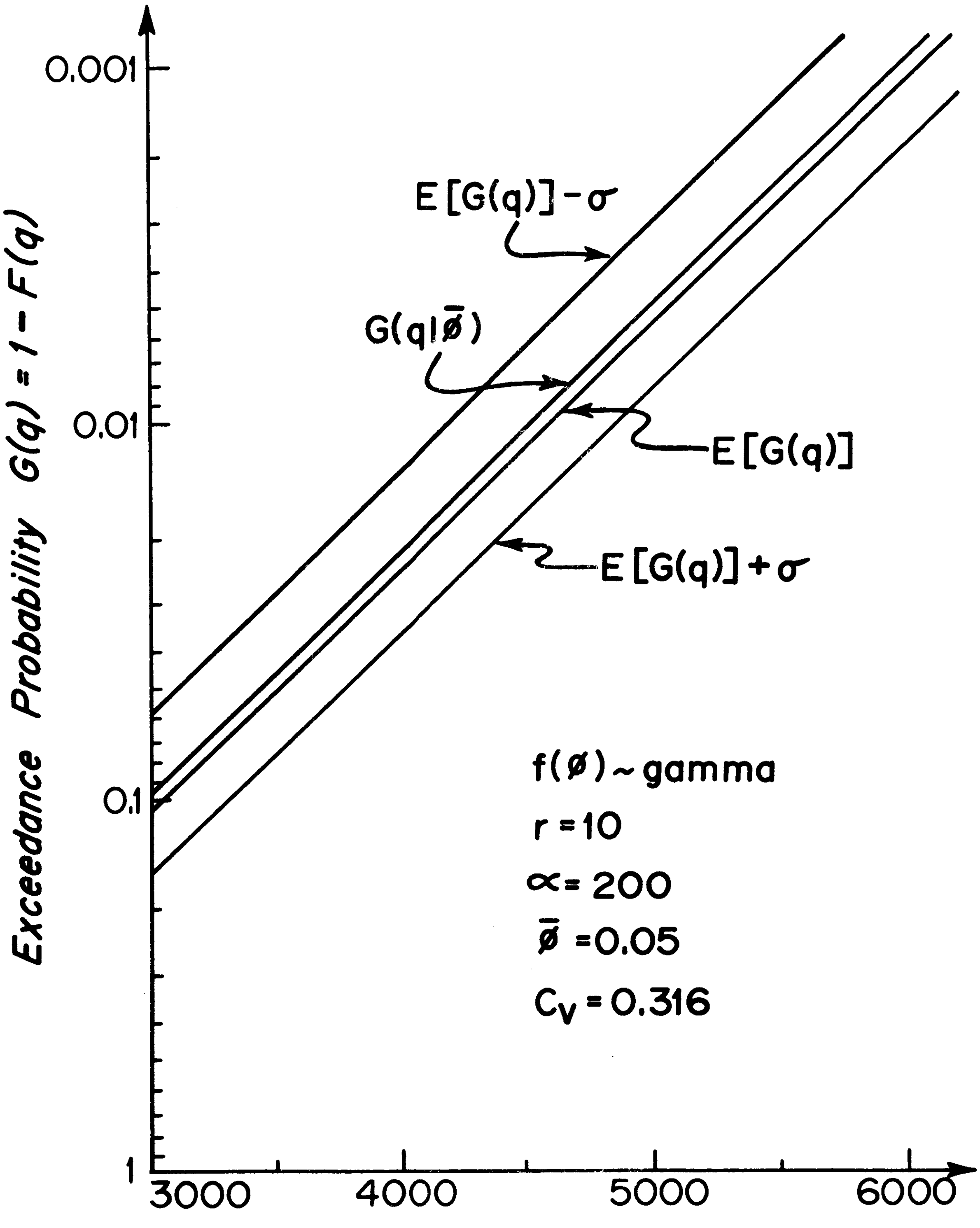


Figure 5.14: Frequency Curves for $f(\phi)$, Gamma-1 with $\bar{\phi}$ = .05 and C_v = .316

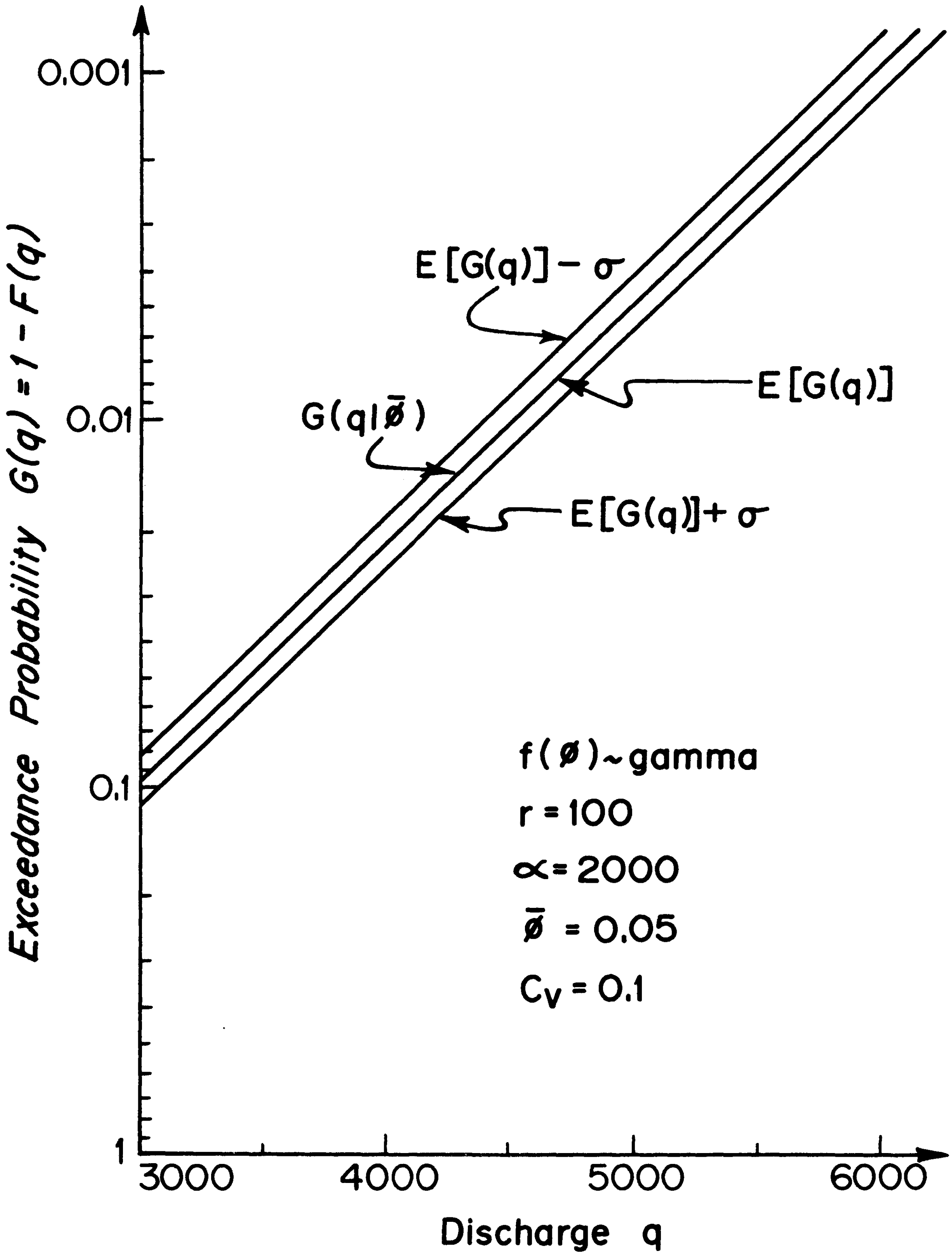


Figure 5.15: Frequency Curves for $f(\phi)$, Gamma-1 with $\bar{\phi}$ = .05 and C_v = .10

evident from the curves. In decision problems, the expected exceedance probability, $E[G(q_m)]$ whould be used. Take the case where $f(\phi)$ is exponential with a mean $\bar{\phi} = .05$. The error introduced by specifying that a peak discharge of 4500 cfs has a return period of 100 years, as predicted by the deterministic analysis, is substantial, since the stochastic analysis predicts that a peak discharge has a return period of 50 years. This error in accounting for parameter uncertainty may lead to serious design problems. When the information about ϕ is very good, which is represented by a tight distribution on ϕ (and shown in Figure 5.15), the difference between the two analyses is very small. Of course, this is expected.

This analysis only considered one uncertain parameter in the rainfall runoff modelling. The implications of considering many uncertain parameters are evident.

5.6 Conclusion

Chapter 5 analyzes the uncertainty in the output of a deterministic rainfall-runoff model due to the uncertainty in the models' parameters. Eagleson's derived flood frequency analysis is used to find the constant peak discharge boundary in the $\bar{i} - t_r$ plane, which in turn is used to define Rq_m, the region in which combinations of $\bar{i}$ and t_r yield discharges less than or equal to q_m. This boundary permitted the evaluation of the flood exceedance probability, $G(q_m)$ which is the probability that $q > q_m$. The uncertainty in the runoff model is represented by the water loss coefficient, ϕ, which results in uncertainty in the position of the constant peak discharge boundary for q_m and in the size and location of the region Rq_m. The expected flood exceedance probability, $E[G(q_m)$, is found by

$$\tilde{G}(q_m) = E[G(q_m)] = 1 - \int_{\phi} f(\phi)\, d\phi \int_{Rq_m|\phi} f(\bar{i}, t_r)\, d\bar{i}\, dt \tag{5.52}$$

which considers the uncertainty in ϕ .

Two probability density functions are obtained analytically. One is the peak discharge, conditional upon an exceedance probability level, and the other is the exceedance probability at a peak discharge level. This leads to the result that the use of a point estimate for the water loss ϕ underestimates the peak discharge for a given exceedance level, $G(q_m)$. Similarly, such a procedure underestimates the exceedance probability for a given peak discharge.

Continued research remains to be done on parameter uncertainty in rainfall runoff modelling. There are those parameters which vary from storm to storm - for example, the rainfall interior pattern, which are really stochastic processes and should be analyzed in such a framework. There are those parameters which are uncertain, due to statistical uncertainty. Their effect upon the region Rq_m has not been fully researched either. The area of parameter uncertainty in modelling the rainfall runoff process will provide many years of interesting work.

The extension of the theory presented here to other simulation models outside of hydrology - for example, water quality models, - is straightforward. If simulation models are going to be applied for prediction, where the concern is an unknown future state of nature, (an urbanized watershed, for example), then the probability distributions on the models' outputs should be estimated if the outputs are used to make meaningful decisions.

CHAPTER 6

Decision Analysis for a Flood Protection Design

6.1 Introduction

This chapter is concerned with decision-making, and the effects of uncertainty in modelling the occurrence of floods of a particular magnitude upon those decisions. The decision variable is usually an engineering variable such as the height of a dike, the capacity of a flood channel or the size of a spillway. In the example presented here, the decision problem is to determine the level of flood protection for Woonsocket, Rhode Island, which is often flooded by the Blackstone River.

The analysis of modelling the probability of flood magnitudes was presented in Chapter 4. In that chapter, four different probability models were analyzed. Each model was assumed to correctly represent the probability of the occurrence of a maximum annual flood of a particular magnitude. The model parameters were considered unknown random variables and, by combining information from historical flood records and other sources, the confidence about the parameters could be represented by a probability density function. With these density functions, it was possible to find the probability density function for the flood discharges which fully accounted for the uncertainty in the model parameters. This probability density function is the so-called Bayesian distribution of the flood discharges. Three probability models in Chapter 4 modelled the series of maximum annual floods by 1) a normal probability density function, 2) a log-normal probability density function or 3) a gamma-1 probability density function. The fourth model assumed that the extremes of the magnitudes of the partial duration series of flood peaks is best represented by an exponential probability density function and the time between the independent flood peaks is also distributed exponentially.

The four models were applied to the Blackstone River at Woonsocket, R.I. Due to their different underlying structures, each model gave a different flood frequency curve. Since inferences about flood magnitudes affect decisions, the different models may identify different decisions as being the best one. This chapter compares the decisions from the four probabilistic models. The analysis of model uncertainty is performed in Chapter 7.

The relationship between inferences about flood magnitudes and decisions on engineering variables depends upon the decision rule. The decision rule that is used in this chapter is as follows: the decision, which is chosen as being the best decision, is the one that maximizes the expected utility. This decision rule was discussed in Chapter 2, and it assumes that there exists a utility function which can be measured. The value of the utility function, $u(d_i,q)$, depends upon the decision taken, d_i, and the flood discharge q. The expected utility from any decision d_i is,

$$E[u(d_i)] = \int_q u(d_i,q) \cdot \tilde{f}(q)\,dq \tag{6.1}$$

where

$u(d_i, q)$ is the utility function

$\tilde{f}(q)$ is the (Bayesian) probability density function for flood discharges.

The best decision, d*, under the above decision rule, will satisfy the condition.

$$E[u(d^*)] = \max_{\text{all } d_i} E[u(d_i)] \tag{6.2}$$

Utility functions may be of many different forms. They can include and represent multiobjective outputs, social costs and benefits, and risk adversion toward particular outcomes. This is under the assumption that the decision maker can express such preferences and that they can be measured. In the example presented here, it will be assumed that the appropriate utility function will be net monetary benefits. The total costs will be the cost of providing flood protection, and the gross benefits will be the resulting reduction in flood damages.

Using the decision rule represented by Equation (6.2), the 'best' level of flood protection for the example problem will be determined for each of the four probability models. The effect of the different models upon the decision is discussed for two budgetary conditions. The first condition is an unlimited capital budget for flood control projects and the second condition is a budget constraint on capital expenditures.

6.2 Flood Protection for Woonsocket, Rhode Island

6.2.1 Problem Description

Woonsocket, Rhode Island is on the Blackstone River, which has an upstream drainage area of 416 square miles. The Blackstone River has a bankful capacity within Woonsocket of approximately 700 cfs (Corps of Engineers, 1958). In the last 40 years, there has been approximately one major flood every eight years. Flows exceeding bankful discharge have occurred, on the average, about once in five years.

Local flood protection can be provided by a combination of channelization, dikes and flood walls. The decision variable is the capacity of this system. Figure 6.1 presents the capital cost for different channel capacities. The cost curve is hypothesized from information contained within a flood protection study for Woonsocket by the Corps of Engineers (1958). Figure 6.2 presents two realistic flood damage curves, also based upon information from the same report. One stage-damage curve is under the condition of no flood protection. The other stage-damage curve is where the flood protection system has a design capacity of 32,000 cfs, which is equivalent to a river elevation of 129 feet. The flood stage-damage curves, with the stage-discharge curves as shown in Figure 6.3 (U.S. Geological Survey, 1958), are used to calculate a flood discharge-damage curve. For the sake of simplicity, it is assumed, in this example, that the stage-discharge relationship of Figure 6.3 will not change, due to the construction of the flood protection works.

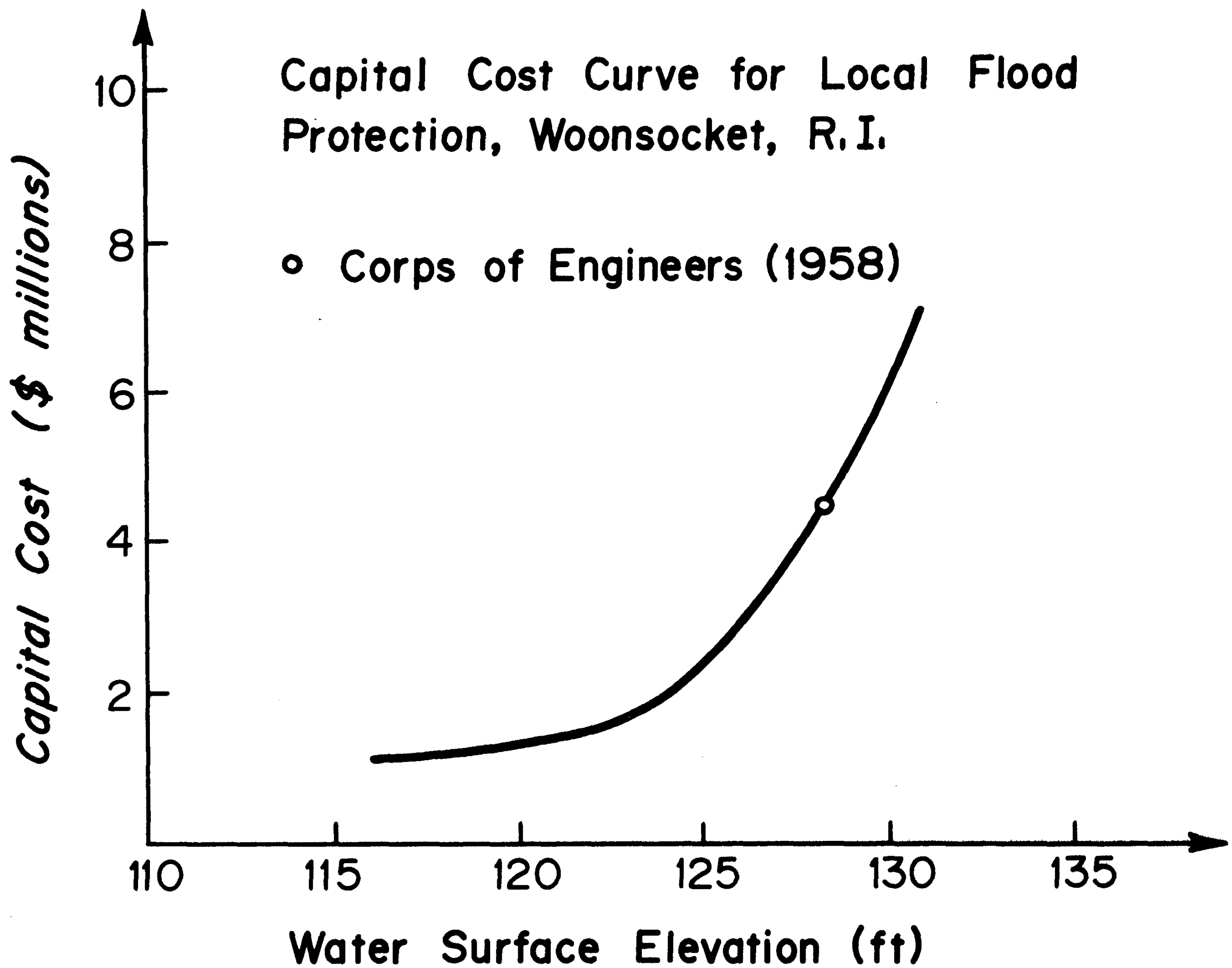


Figure 6.1: Capital Cost for Flood Protection, Woonsocket, R.I.

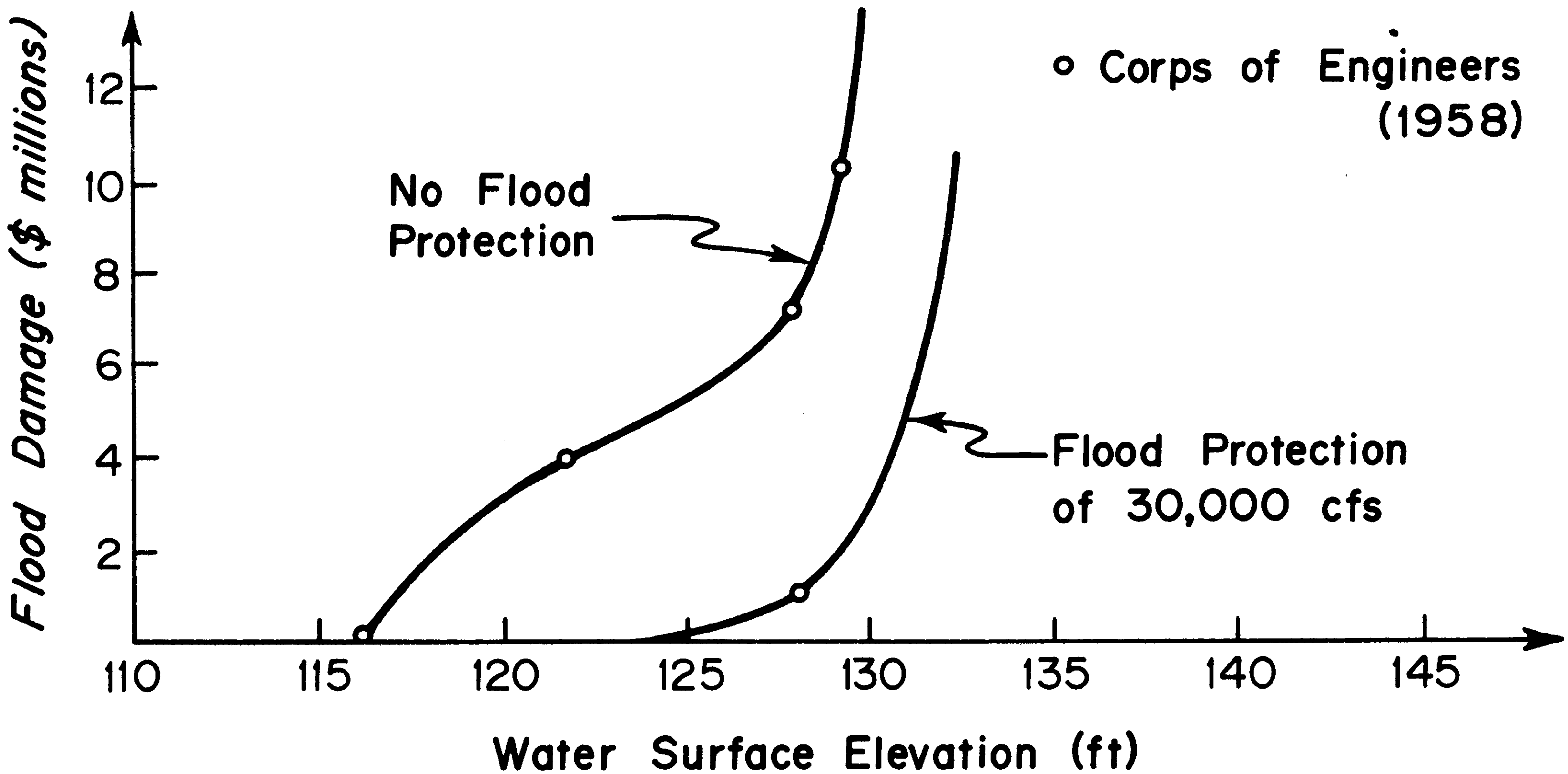


Figure 6.2: Flood Damage Curves for Woonsocket, R.I.

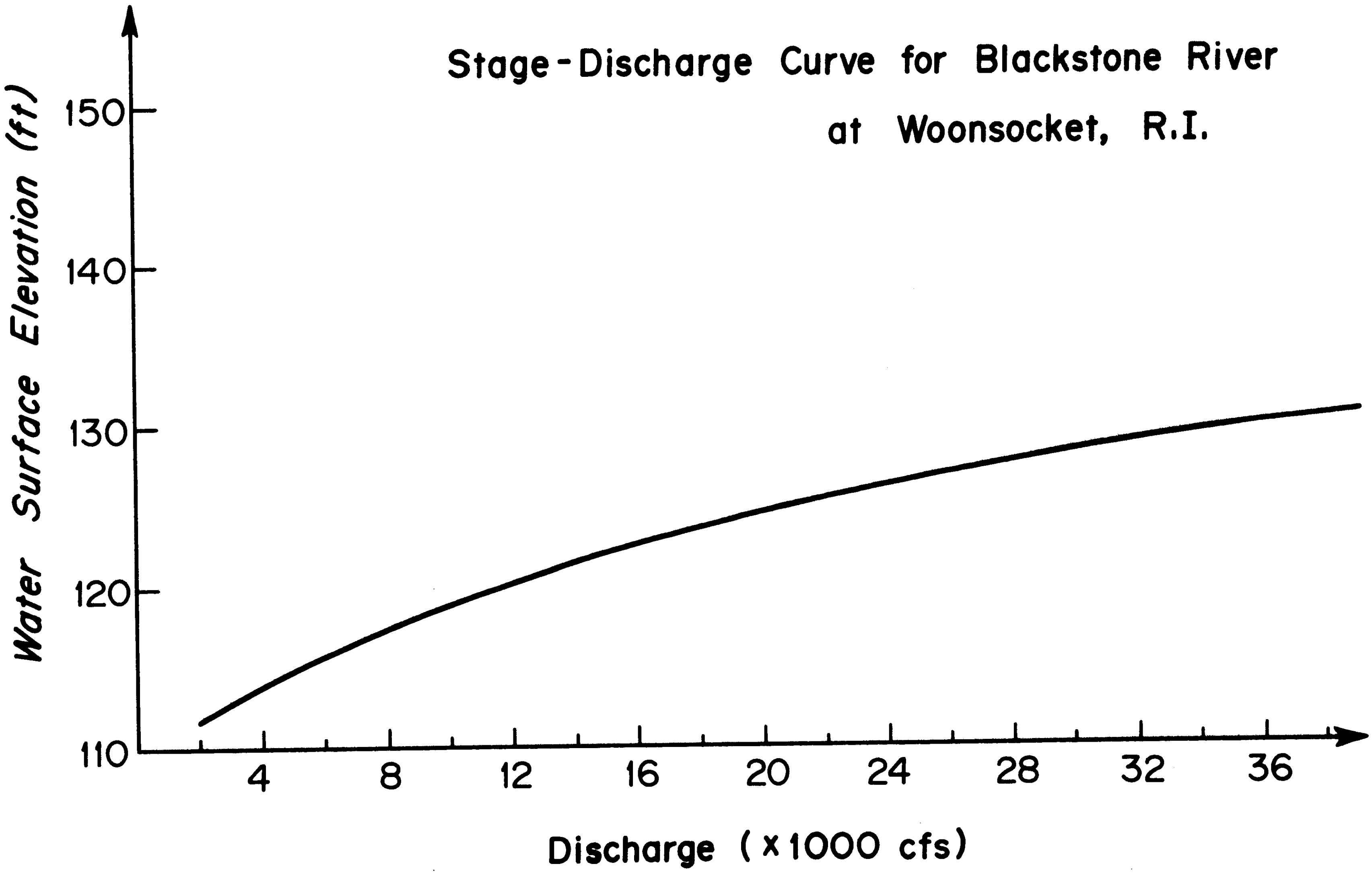


Figure 6.3: Stage-Discharge Curve for the Blackstone River, Woonsocket, R.I.

For ease of computation, the decision set is limited to five different flood protection levles. These levels are shown in Table 6.1 along with the capital cost of protection. Reasonable damage curves were obtained by the interpolation and extrapolation of the damage curves presented in Figure 6.2 for the five different levels of the decision set.

6.2.2 Discussion of Results.

The expected annual flood damage (in dollars) predicted by each model depend upon the flood damage curves and the probability model. The expected damages are calculated by Equation (6.1). The interaction between the damage curves and the probability models can be best observed by rescaling the damage curves. The damages resulting from a particular peak discharge are reached by the probability of that discharge occurring. The area under the rescaled damage curve is the expected damages. Figure 6.4 shows the expected flood damage curves for the six levels of protection (no protection plus the five levels presented in Table 6.1) rescaled by the normal probability model. Figure 6.5 is similar but uses the log-normal model, while Figure 6.6 uses the gamma-1 model and Figure 6.7 uses the exceedance model. These figures not only show graphically which floods contribute to the expected damages, but also show the expected reductions in damages (marginal flood benefits) between various designs. The marginal expected benefits of one design over another is the area between the corresponding rescaled damage curves. The graphs show from

Decision	Flood System Capacity (cfs)	Capital Cost ($ Million)
d_0	8500	0
d_1	15000	1.25
d_2	20000	2.0
d_3	25000	3.0
d_4	30000	4.25
d_5	36000	7.0

Note: Decision d_0 is for no flood protection

Table 6.1

Capital Costs for Various Flood Protection Levels, Woonsocket, Rhode Island Example

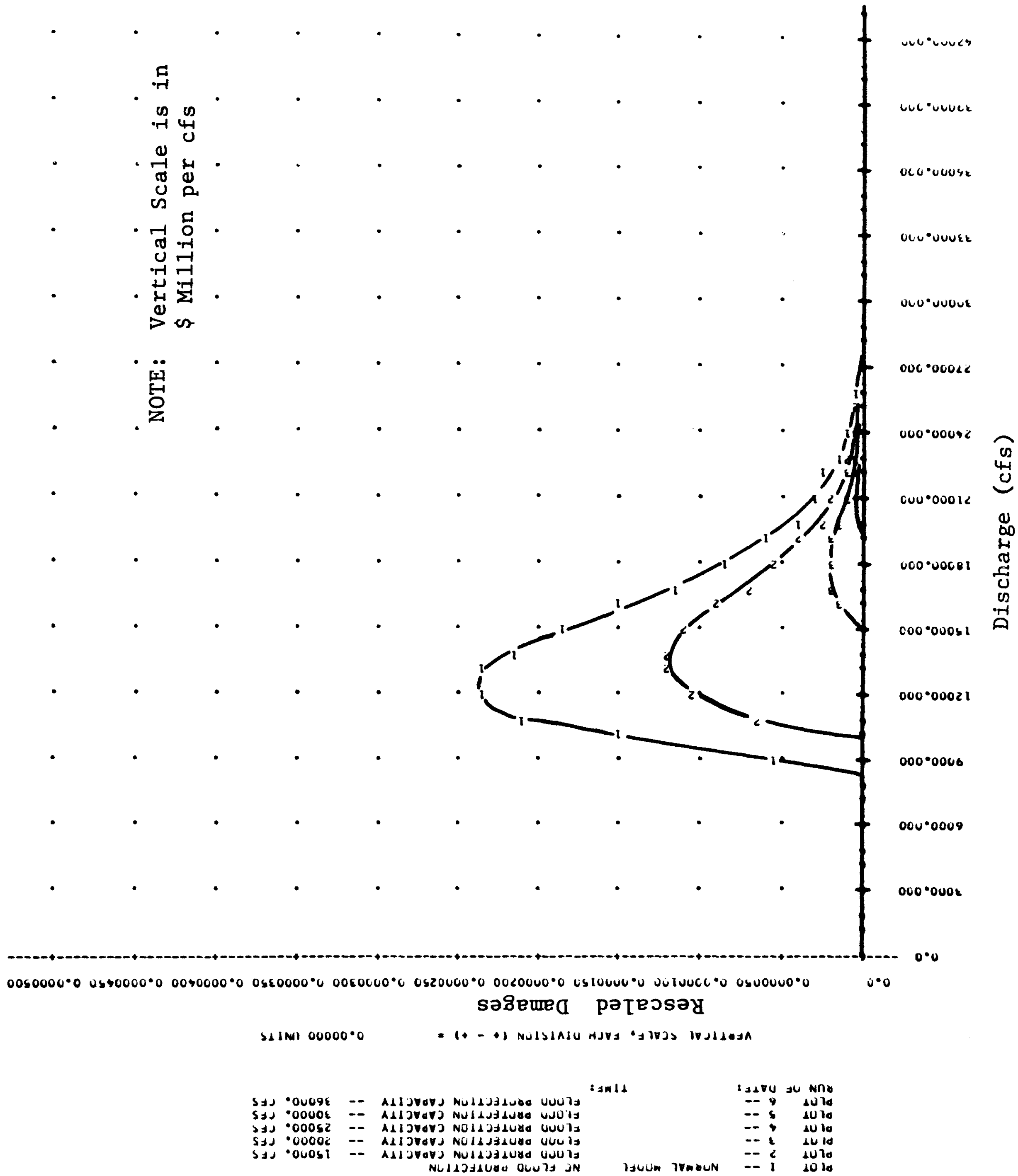


Figure 6.4: Rescaled Realistic Damage Curves for the Normal Model

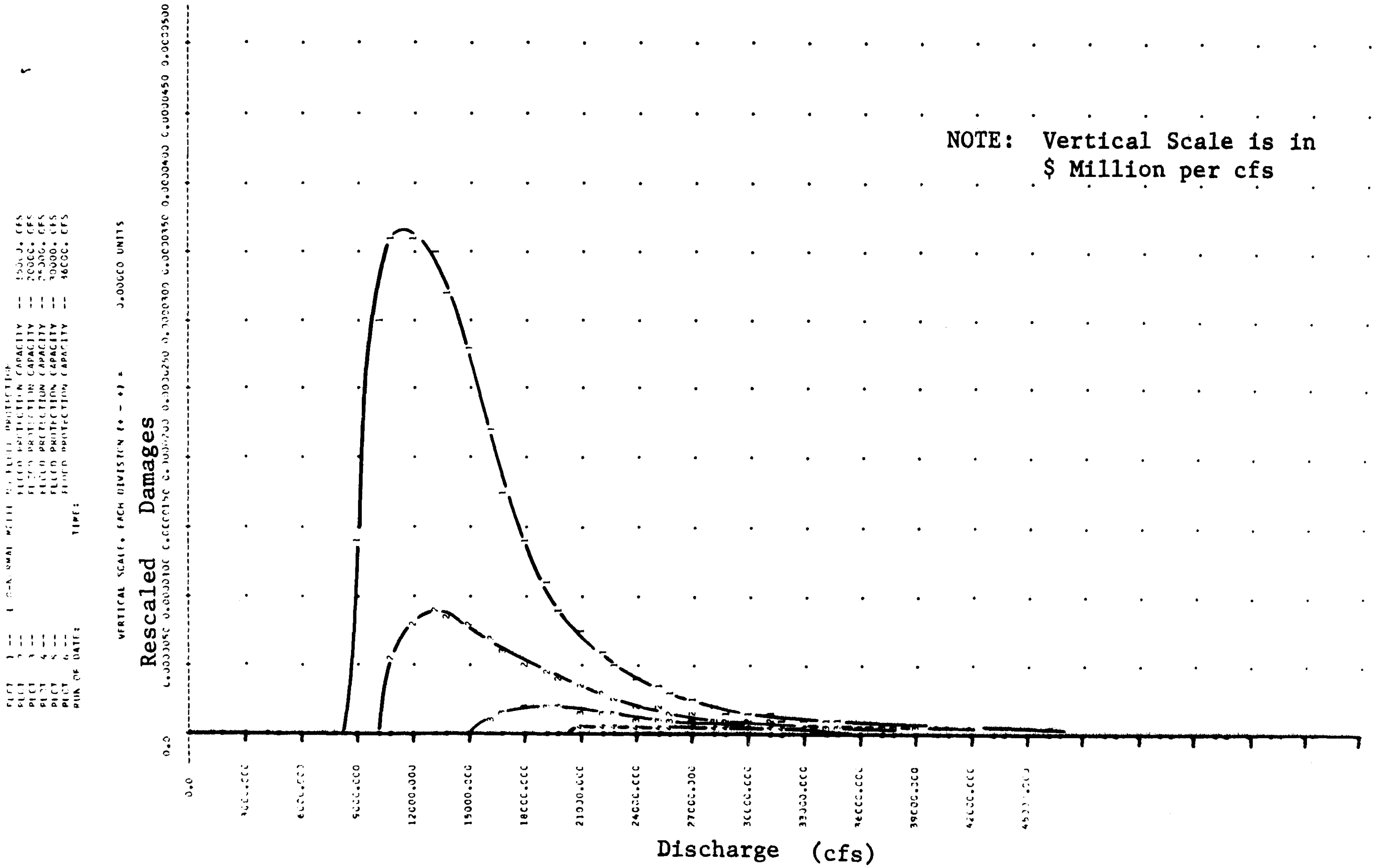


Figure 6.5: Rescaled Realistic Damage Curve for the Log-Normal Model

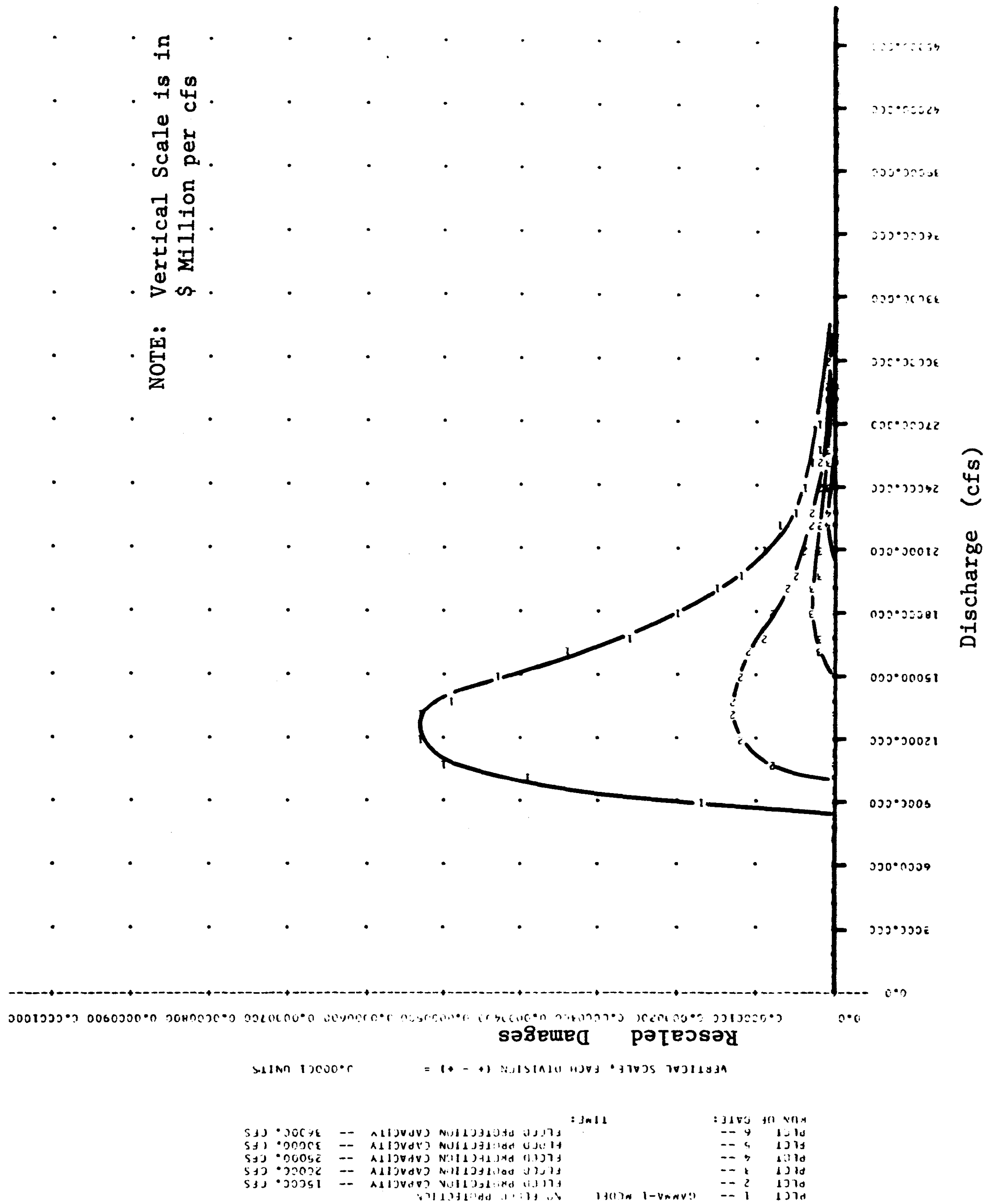


Figure 6.6: Rescaled Realistic Damage Curve for the Gamma Model

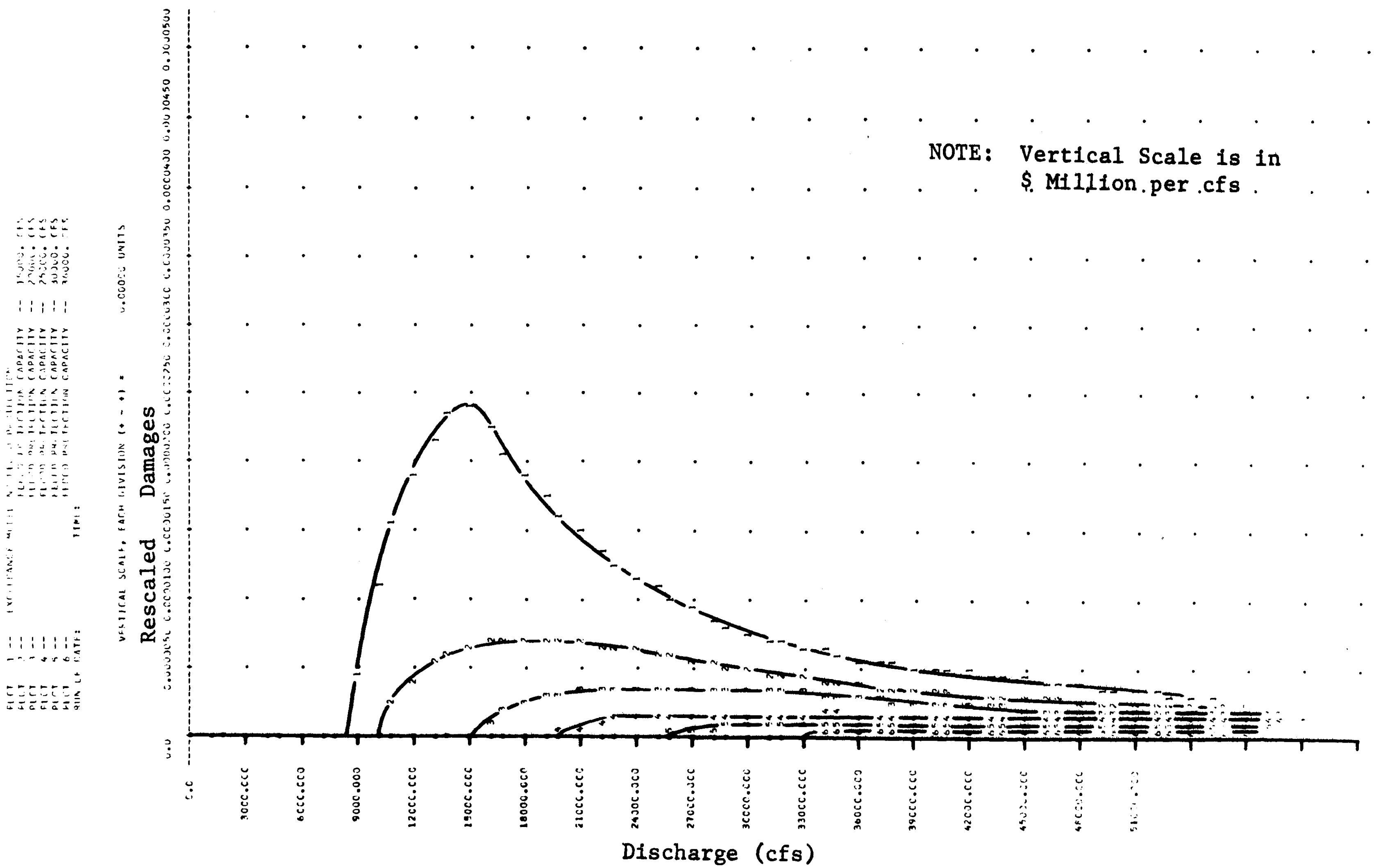


Figure 6.7: Rescaled Realistic Damage Curve for the Exceedance Model

which peak flood discharges the expected benefits are derived.

Table 6.2 presents the expected annual flood damages for the four probability models and the six decision levels.

The expected annual flood damages from the normal and gamma-1 model are higher than the damages from the exceedance and log-normal models. This difference is due to the former models having a higher probability of observing a flood that causes damage and the relatively large damages that occur from fairly small, but damaging, floods. The previous figures, 6.4 through 6.7, show this interaction between probabilities and damages. Furthermore, as the flood protection capacity increases, the models with the higher probability of observing extreme annual floods have the higher expected annual damages, as best displayed by the exceedance model (Figure 6.7).

The gross benefits from a flood protection design are calculated by converting the expected annual flood benefits (reductions in flood damages) to an equivalent present value. It is assumed that 5% is the appropriate social rate of discount and that the project has a 50 year life. The expected net benefits are calculated by subtracting the cost of protection from the gross benefits. For each model, the expected net benefits for the five decision acts are given in Table 6.3. Figure 6.8 presents the benefit and cost curves obtained from each model.

Decision	Expected Damages ($ Million)			
	Normal	Log Normal	Gamma-1	Exceedance
d_0	10.63	6.09	8.17	8.18
d_1	2.58	1.68	2.12	3.84
d_2	.336	.460	.423	2.12
d_3	.030	.147	.083	1.11
d_4	.001	.055	.015	.612
d_5	0	.017	.001	.276

Note: Annual Flood Damages Discounted at 5% over a 50 Year Project Life

Table 6.2

Expected Damages for Various Decision Levels, Realistic Damage Function

Decision	Expected Benefits ($ Million)							
	Normal		Log Normal		Gamma-1		Exceedance	
	Gross	Net	Gross	Net	Gross	Net	Gross	Net
d_1	8.049	6.80	4.36	3.11	6.054	4.80	4.34	3.09
d_2	10.30	8.30	5.585	3.585	7.75	5.75	6.05	4.05
d_3	10.60	7.60	5.898	2.898	8.09	5.09	7.069	4.069
d_4	10.63	6.38	5.99	1.74	8.16	3.91	7.57	3.32
d_5	10.63	3.63	6.03	-.973	8.17	1.17	7.90	.90

Table 6.3

Expected Benefits for Various Decision Levels, Realistic Damage Function

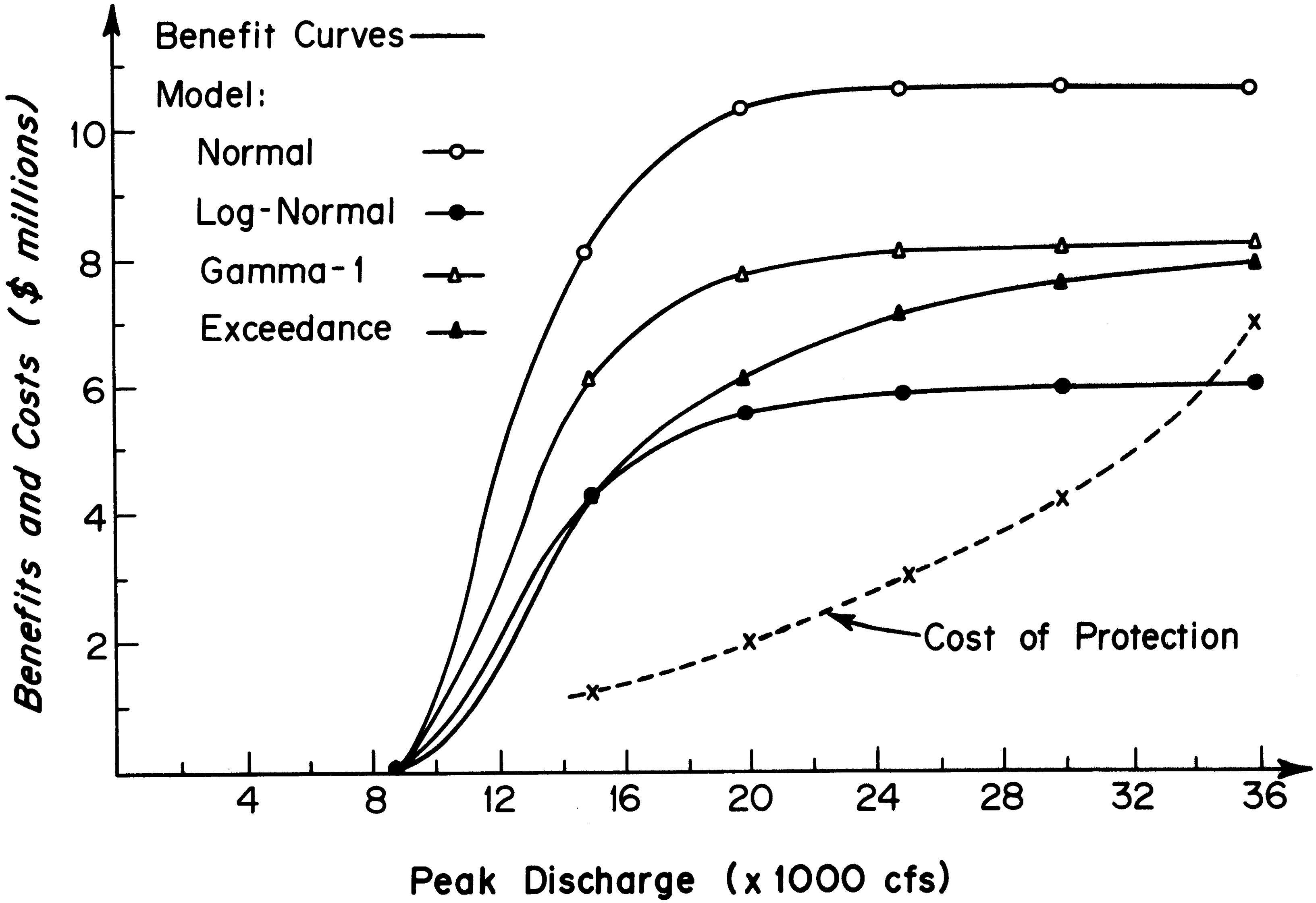


Figure 6.8: Benefit and Cost Curves for Woonsocket R.I. Example

The results indicate that the best decision is to provide a flood channel capacity of 20,000 cfs, (decision d_2) if the true model of the probability of annual flood discharges follows a normal, log-normal, or gamma-1 density function; and the best decision is a capacity of 25,000 cfs (decision d_3) if the true model is the exceedance model. Let the decision set be increased from five decision levels to a set which contains all possible flood channel capacities. Then the optimal design capacity, given a model, occurs where marginal benefits equal marginal costs, assuming there is an unconstrained budget. This capacity also maximizes net benefits, and could be estimated through the use of Figure 6.8.

To study the effect of the shape of the damage curve upon the results, the same procedures that were followed for the realistic damage curves were followed for a set of quadratic damage curves of the form

$$\text{DAMAGE} = 0.26\ (Q - Q_o)^2 \quad \text{for } Q > Q_o$$

$$= 0 \qquad\qquad \text{for } Q \leq Q_o$$

where

Q_o is the minimum discharge where damage occurs, and is a function of the decision level.

Q is peak discharge (cfs)

Damages are in dollars.

The coefficient of the quadratic damage curve is chosen in such a manner that the quadratic damage curve and the realistic damage curve will intersect at a flood discharge of 32,000 cfs. At a discharge of 32,000 cfs the Corps of Engineers (1958) estimated flood damages to be 10.25 million dollars. The capital cost curve and decision levels are the same as those used in the realistic damage curve example. Figures 6.9, 6.10, 6.11 and 6.12 present the rescaled damage curves for each probability model. Table 6.4 gives the benefits of each decision level, and Figure 6.13 presents the benefit and cost curves obtained from the models. The results indicate that if the log-normal model is the true model, then no flood control works should be built. If the normal model or the gamma-1 model is the true model, then a design capacity of 20,000 cfs (decision d_2) is best. If the exceedance model is the correct model, then the best design capacity is 25,000 cfs (decision d_3). These optimal designs are based upon the assumption that maximizing net benefits is the appropriate criterion by which decisions are made. For all the models, the benefits from the quadratic damage curve are significantly different than the benefits obtained from the realistic damage curves. The results from the various models, however, tend to be similar, but is is important to realize that this outcome can not be generalized to other decision examples.

The discussion of the best decision has, up to now, only considered the condition of an unconstrained budget. The effects of a constrained budget upon the decisions will be discussed in the next section.

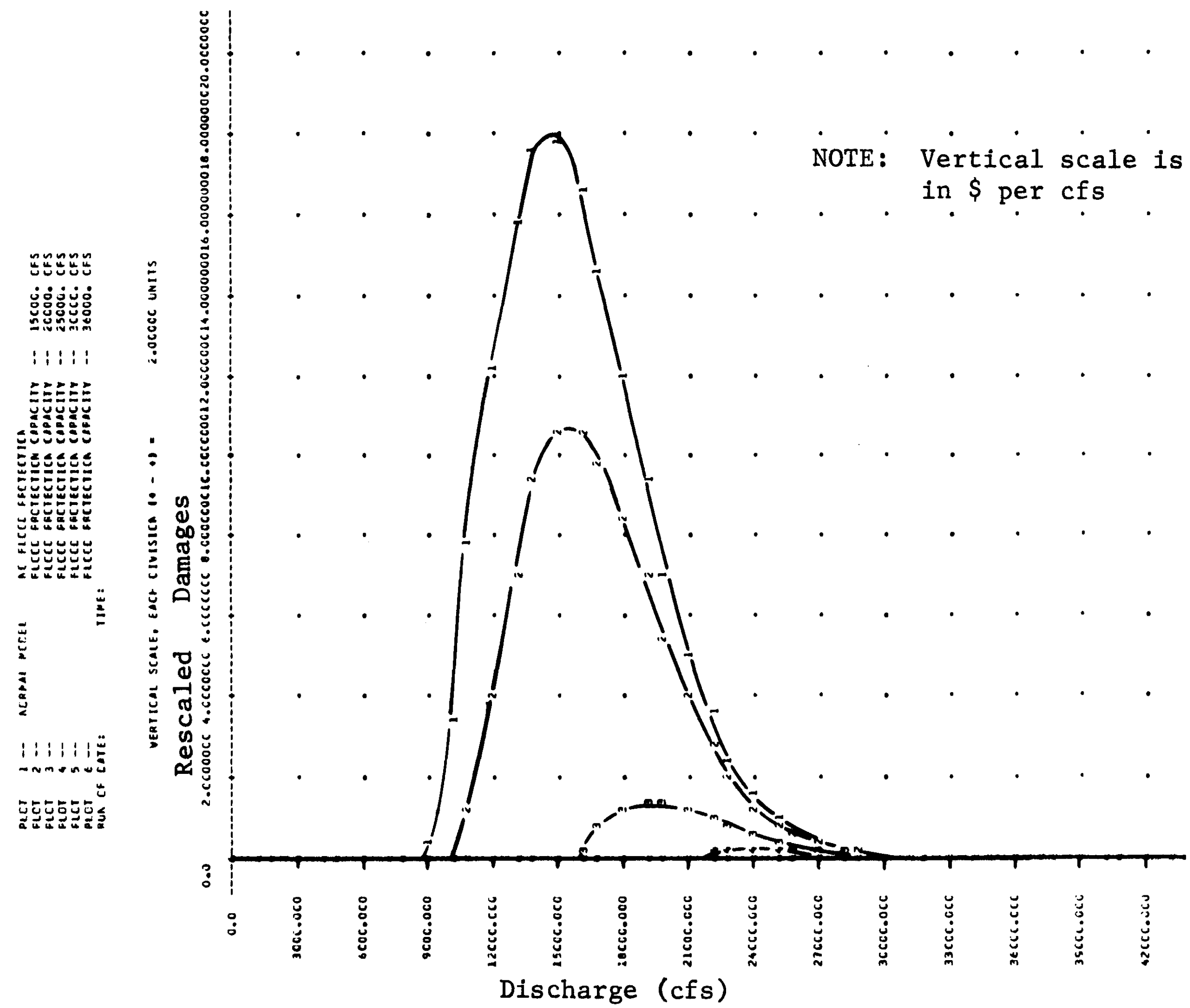


Figure 6.9: Rescaled Quadratic Damage Curve for the Normal Model

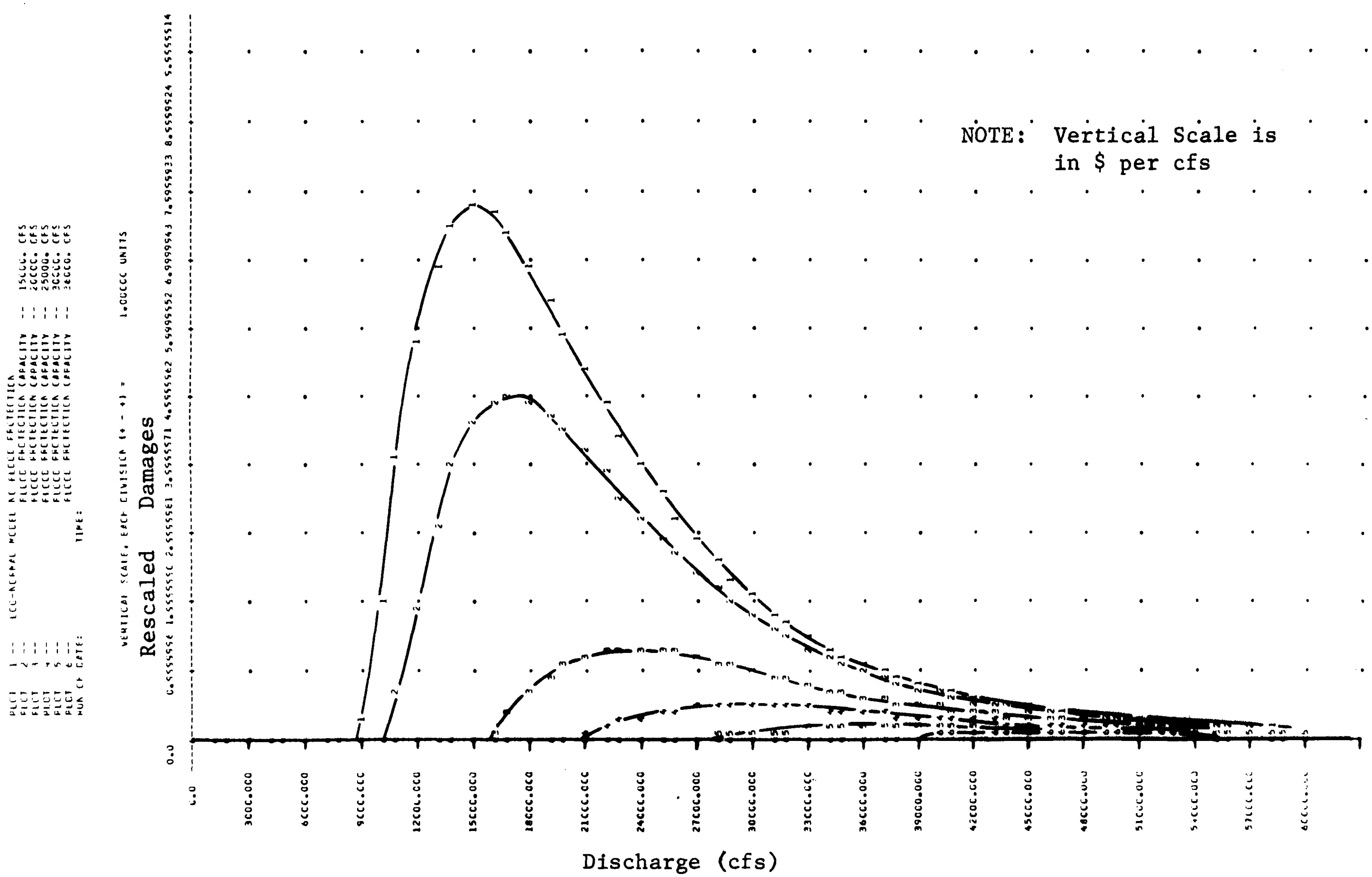


Figure 6.10: Rescaled Quadratic Damage Curve for the Log-Normal Model

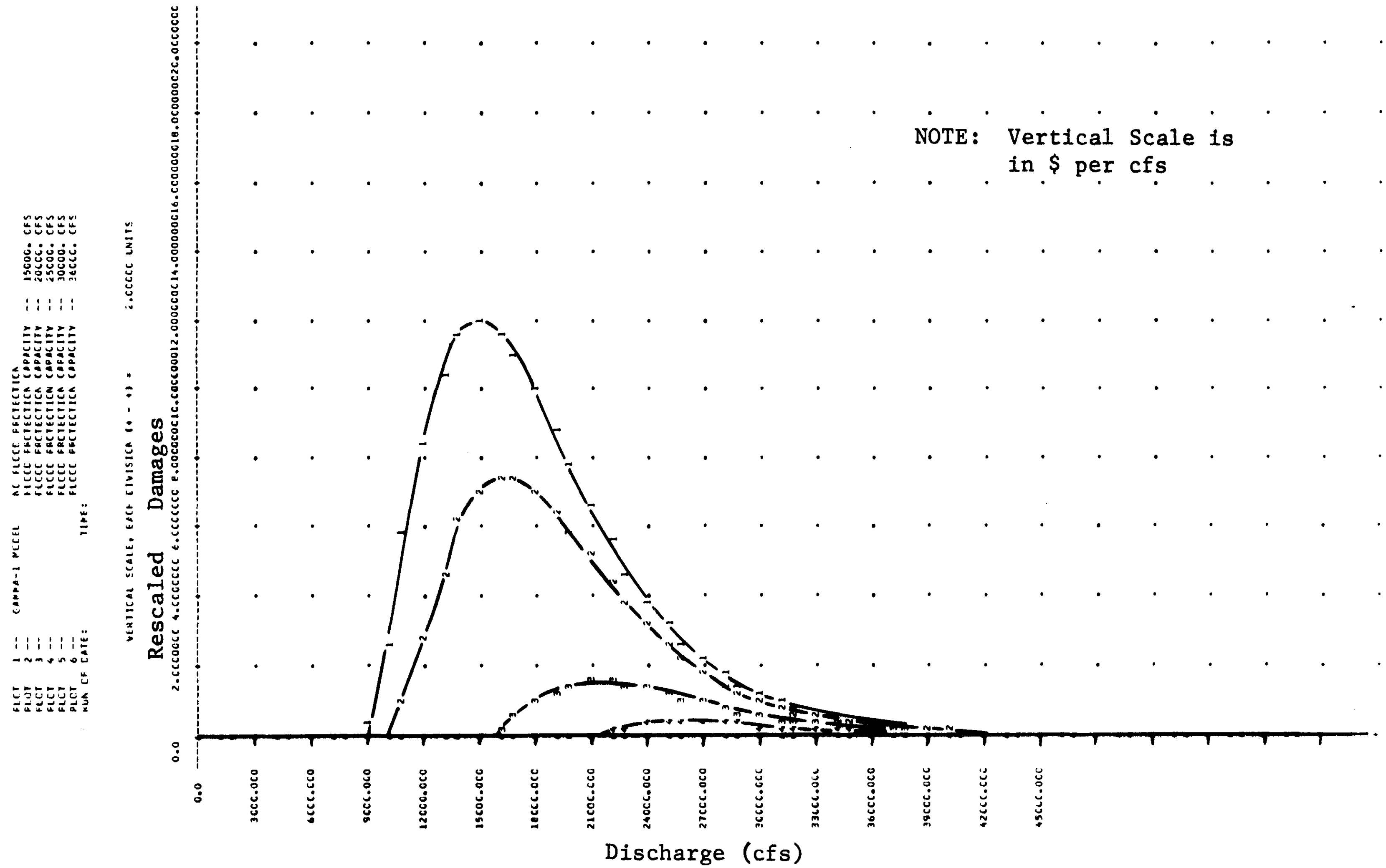


Figure 6.11: Rescaled Quadratic Damage Curve for the Gamma Model

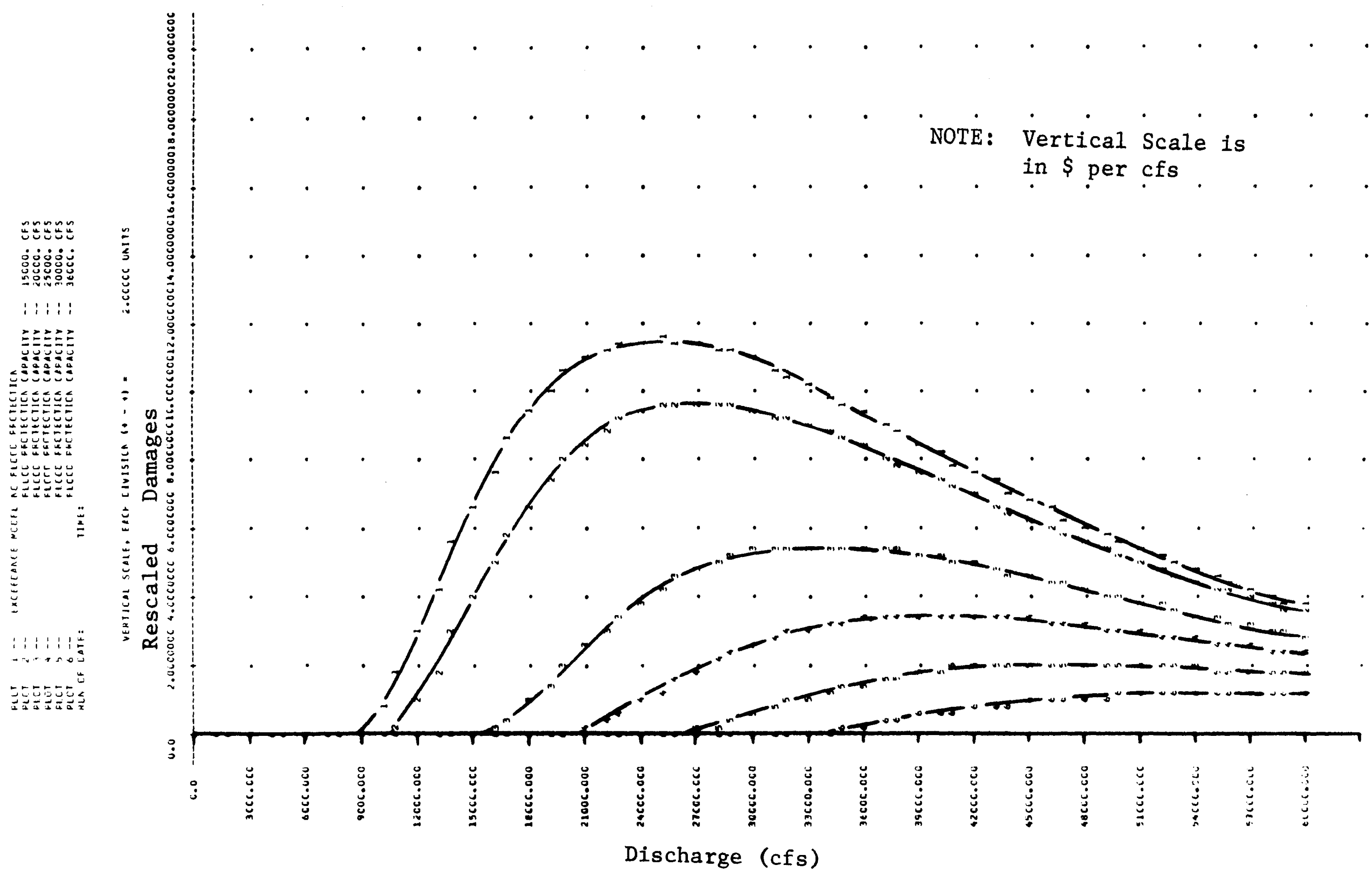


Figure 6.12: Rescaled Quadratic Damage Curve for the Exceedance Model

Decision	Expected Benefits ($ Million)							
	Normal		Log-Normal		Gamma-1		Exceedance	
	Gross	Net	Gross	Net	Gross	Net	Gross	Net
d_1	1.17	-.08	.704	-.545	.930	-.32	1.156	-.09
d_2	2.55	.55	1.737	-.263	2.17	.17	3.77	1.77
d_3	2.71	-.29	2.02	-.977	2.42	-.58	5.11	2.11
d_4	2.72	-1.53	2.15	-2.10	2.48	-1.52	6.036	1.79
d_5	2.72	-4.28	2.20	-4.80	2.49	-4.51	6.59	-.41

Table 6.4

Expected Benefits for Various Decision Levels, Quadratic Damage Curve

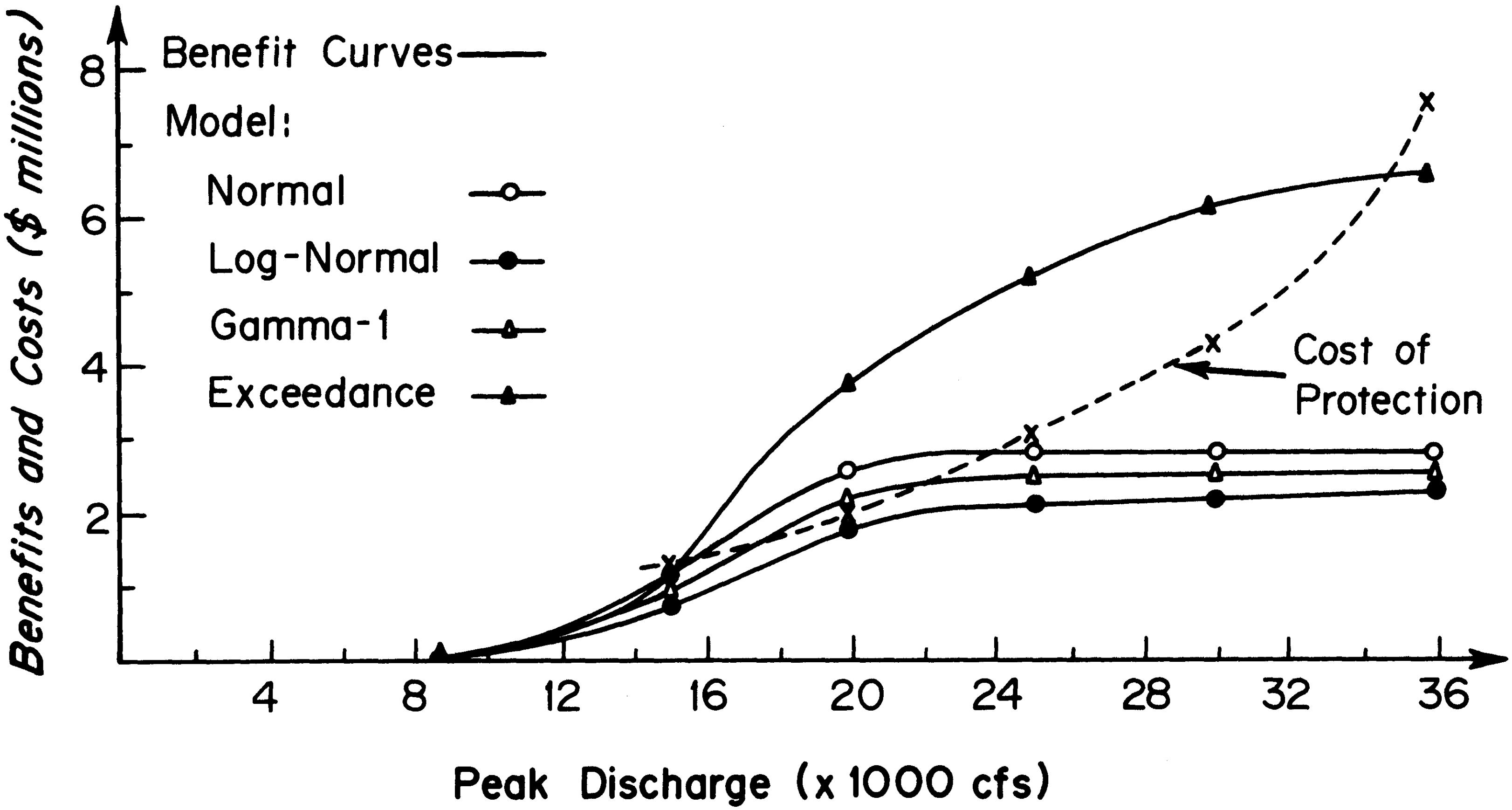


Figure 6.13: Benefit and Cost Curve, Using a Quadratic Damage Function

6.2.3 The Effects of Budget Constraints on Optimal Decisions

When a budget is unconstrained, then any project which has positive net benefits should be constructed. The optimal project size occurs where the marginal benefit to cost ratio is 1.0. At this point the net benefits are also maximized. Marginal benefits are the benefits contributed by the last increment built and marginal cost is the cost of building the last increment.

In this discussion, a budget constraint is a restriction upon the funds available for capital expenditures. The budget constraint will be binding if the total set of possible projects (of which this flood control project is but one) has expenditures that exceed available funds. Under a budget constraint, the theoretically correct procedure is to maximize expected net benefits (utility) within the budget. This procedure is achieved by redesigning the set of feasible projects so that the marginal benefits to costs are the same for all projects, and total expenditures just exhausts the budget (Major, 1973). For the last increment of flood control capacity, the marginal benefits to marginal costs should equal $1 + \lambda$, where λ is defined as the shadow premium.

From Figure 6.8, it is evident that the different models of the distribution of annual maximum flood discharges will, under a tight budget constraint, lead to quite different decisions. Table 6.5 shows the net benefits for each decision using the realistic damage curves and a shadow premium of 1.0. There is a definite shift in

Decision	Expected Net Benefits ($ Million)			
	Normal	Log-Normal	Gamma-1	Exceedance
d_1	5.55	1.86	3.55	1.84
d_2	6.30	1.59	3.75	2.05
d_3	4.6	-.10	2.09	1.07
d_4	2.13	-2.51	-.34	-.93
d_5	-3.37	-7.97	-5.83	-6.10

Table 6.5

Net Benefits under a Shadow Premium, $\lambda = 1$

decisions. Figure 6.14 presents the cost curves for the unconstrained budget, $\lambda = 0$, and for the constrained budget, $\lambda = 1$. Also presented are the benefit curves from each model. The change in the optimal design levels can be clearly seen. The shift in optimal design levels is larger for the log-normal model and the exceedance model than for the gamma-1 or normal model. The difference in the shift is due to the shape of the benefit curve, and depends upon the model.

For both the unconstrained and constrained budget condition, the optimal decision level depends upon the interaction of the damage curves, cost curves, and the probability models. Results can not be generalized, and it is incorrect to make general statements about the 'best' probability curve to use in real engineering flood design problems.

6.3 Summary

This chapter looks at a simple decision problem of determining the best size for a local flood protection works. Four probabilistic models of the magnitude of the maximum annual flood are considered. The decision problem applied realistic cost and damage curves from the Corps of Engineers (1958). Using the decision rule of maximizing net benefits, the four models gave similar optimal decision or design levels. The design levels depend upon the damage curve as well as upon the probability model. At the optimal decision, the risk varies significantly among the four models. For each model, the return

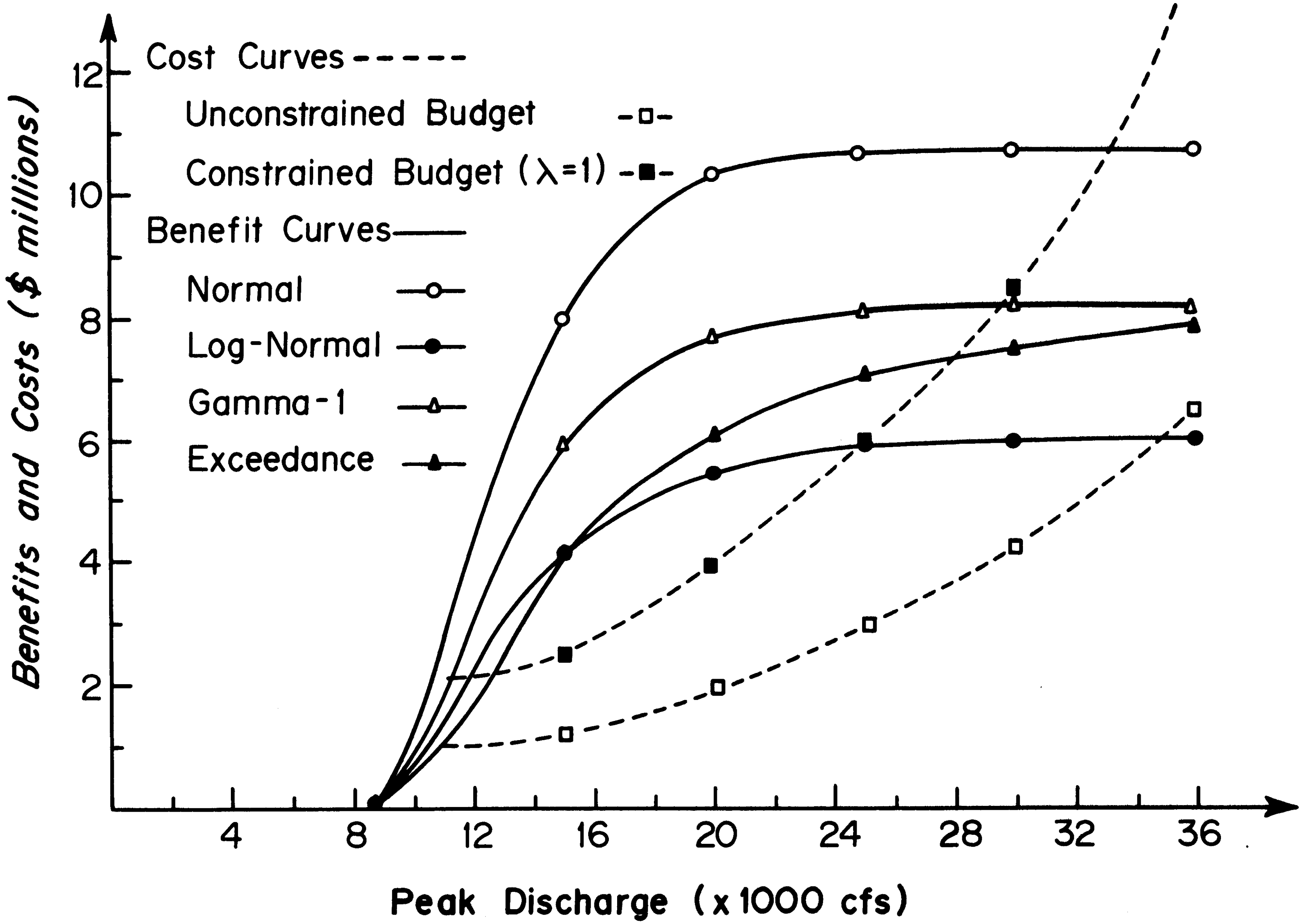


Figure 6.14: Benefit and Cost Curves with Shadow Premium, $\lambda = 0.0$ and $\lambda = 1.0$

period for a flood to exceed the optimal decision is:

normal	T = 250 years
log-normal	T = 84 years
gamma-1	T = 55 years
exceedance	T = 60 years.

The optimal design level for the exceedance model, which is the most expensive, represents a capital outlay of about 50% higher than the decision from the log-normal model, which has the lowest capital outlay.

For the example presented here, the different models did not lead to vastly different results. This outcome is not necessarily true for all decision problems. Furthermore, the risk of a flood exceeding the optimal decision varies among models and may also influence decision makers. This chapter considers the affect upon decisions of various probabilistic models. Chapter 7 will consider the problem of model uncertainty and selection.

Chapter 7

Inference and Decision Making Under Model Uncertainty

7.1 Introduction

Hydrologists are often confronted with the problem of choosing one statistical model from many contending models. This selection problem is most frequently encountered in flood frequency analysis. Here, many models seem to fit the available data very well, but often the models lead to different decisions. In recent years, considerable progress has been made on the development of statistical procedures for comparing alternative models. The most significant work has been the application of Bayesian statistics for the comparison of econometric models. (Gaver and Geisel, 1972 .) Summarize extensively the recent work in model selection for both Bayesian and new-Bayesian procedures. The work by Leamer (1973) also considers the econometric, model selection problem.

The model selection problem most often addressed in the econometrics literature is the decision of which variables should be included in a regression equation. This model selection problem is conceptually easier, since the likelihood functions are all of the same form. The procedures found in the Bayesian econometrics literature can be extended to cover the more general problem of selecting among models that have different functional forms. Smallwood (1968) uses a Bayesian framework for model selection where the functional forms among the models vary.

The approaches to model selection by non-Bayesian procedures will not be discussed in detail here. An excellent summary is given by Gaver and Geisel (1972). The difficulties of model selection by non-Bayesian procedures are similar to the problems that classical statistical approaches encounter, when dealing with uncertain parameters. These problems are discussed in Chapter 2. Such problems in model selection can be illustrated by the model selection approach taken by Dumonceaux et al (1973).

Let $\underline{x}$ be a random sample from a distribution with a probability density function $f_o(x|a,b)$ or $f_1(x|a,b)$ where

$$f_i(x|a,b) = \frac{1}{b} g_i\left(\frac{x-a}{b} \right), \qquad -\infty < a < \infty, \; b > 0 \tag{7.1}$$

The problem considered is that of selecting f_o or f_1 as the model for the observations. The problem is formulated as a test of hypothesis H_o against H_1, where

$$\begin{aligned} H_o &: X \sim f_o(x|a,b) = \frac{1}{b} g_o\left(\frac{x-a}{b} \right) \\ H_1 &: X \sim f_1(x|a.b) = \frac{1}{b} g_1\left(\frac{x-a}{b} \right) \end{aligned} \tag{7.2}$$

Dumonceaux et al define a statistic, R, where

$$R = \frac{\underset{a,b}{\text{Max}} \prod^{n} f_1(x|a,b)}{\underset{a,b}{\text{Max}} \prod^{n} f_o(x|a,b)} \tag{7.3}$$

which is independent of the unknown location and scale parameters, a and b. Essentially R is a ratio of the maximum likelihood estimators. The distribution for R will depend upon n, f_o, and f_1. Tables are set up of critical values of R, R_c, such that if $R^{1/n} > R_c^{1/n}$, the hypothesis H_o is rejected and the hypothesis H_1 accepted. When Dumonceaux et al tested to discriminate between a normal and a two-sided exponential, the results are extremely ambiguous. With a sample size of 40 (a typical length for hydrology) and an $\alpha = .05$, the power of the test, $1 - \beta$, is only .64 (for 30 years, $1 - \beta = .54$). When the two-sided exponential is the null hypothesis and when n is 40 and α is .05, the power of the test is .58. The low power of the test emphasizes the ambiguity of this classical testing procedure. It is obvious that samples can lead a null hypothesis being accepted, and then, when the null hypothesis and the alternative hypothesis are interchanged, the new hypothesis also being accepted. Furthermore, the testing of more models (in pairs) could lead to more models being 'accepted'. The competing models that are considered in hydrology are statistically closer than the normal and couble exponential, and they would, therefore, lead to more ambiguity.

The choice of one model over another competing model should consider the larger decision problem, where the probability model is only one component. This choice implies the consideration of loss functions. The inability of the classical approach to incorporate the larger decision problem within its testing procedures is discussed in Chapter 2. Bayesian procedures can be included, explicitly, within the total decision problem, which makes the application of Bayesian procedures very attractive. Bayesian procedures have been used to choose one model from a set of models. This selection is analogous to choosing a point estimate for an uncertain parameter within a probability model. In Chapter 2, it is shown that point estimates for uncertain parameters underestimate the variance of the process. Similarly, selecting one model from a set of competing models does not fully account for the uncertainty in the process. When parameter uncertainty is fully accounted for, the resulting model is called the Bayesian distribution. In the case of model uncertainty, the resulting model will be called the composite Bayesian distribution. The development of this distribution is given in the next section.

The Bayesian distribution is used in making decisions when only parameter uncertainty is considered, as in Chapter 6. When both parameter uncertainty and model uncertainty are considered, then the composite Bayesian distribution should be used. Section 7.4 takes the decision problem of Chapter 6 and finds the decision using the composite Bayesian distribution. This decision is compared to making decisions when one model is selected from the model set.

7.2 Composite Bayesian Distribution

In Chapter 4, a number of probability models which describe the frequency of floods are considered. In the analysis of Chapter 4, parameter uncertainty is considered, but the problem of model uncertainty is not considered. Model uncertainty can be considered by defining a composite model of the form

$$\hat{f}(q|\underline{A},\theta) = \theta_1 f_1(q|\underline{A}) + \ldots + \theta_n f_n(q|\underline{A}) \qquad (7.4)$$

The composite model, $\hat{f}(q|\underline{A},\theta)$, is conditional upon a set of unknown model parameters $\underline{A}$ and an unknown composite model parameter set $\underline{\theta}$. $f_1(q|\underline{A})$, ..., and $f_n(q|\underline{A})$ is the set of models that make up the composite model. These models are conditional upon a general unknown parameter set $\underline{A}$. $\theta_1, \ldots,$ and θ_n are parameters that take on a value of either 0 or 1; their value is uncertain. If $\theta_i = 1$, then model $f_i(q|\underline{A})$ is the true model. The constraint

$$\sum_{i=1}^{n} \theta_i = 1 \qquad (7.5)$$

is imposed, which implies that one and only one model is the true model. The definition for $\underline{\theta}$ and the constraint will be used throughout the development of the composite model.

For notational simplicity, consider the case where $n = 2$. The likelihood function for a set of observations $\underline{Q}$ is just:

$$L(\underline{A}|\underline{\theta}) = \theta_1 \prod_{\text{all } i} f_1(q_i|\underline{A}) + \theta_2 \prod_{\text{all } i} f_2(q_i|A)$$

$$= \theta_1 \cdot L_1(\underline{A}|\underline{Q}) + \theta_2 \cdot L_2(\underline{A}|\underline{Q}) \qquad (7.6)$$

There are no cross products of the models, due to the limitation imposed on the values that θ_i can take on; and the constraint on $\underline{\theta}$. $L_i(\underline{A}|\underline{Q})$ is just the likelihood function of model i, conditional upon the observations, $\underline{Q}$.

Define a composite prior distribution on the parameters $\underline{A}$ and $\underline{\theta}$. The prior will be of the form

$$f'(\underline{A}, \underline{\theta}) = \theta_1 f'_1 (\underline{A}|\theta_1 = 1) \cdot p'(\theta_1=1)$$

$$+ \theta_2 f'_2 (\underline{A}|\theta_2 = 1) \cdot p'(\theta_2=1) \qquad (7.7)$$

$f'_i(\underline{A}|\theta_i = 1)$ is the prior distribution on the parameter set $\underline{A}$, conditional upon θ_i equals 1. $P'(\theta_i = 1)$ is the prior probability that model i is the 'true' model.

Bayes' rule can be written as

$$f''(b|\text{data}) = \frac{1}{K} L\ (b|\text{data}) \cdot f'(b) \qquad (7.8)$$

$f''\ (b|\text{data})$ is the posterior distribution of the b, conditional upon the data; $L(b|\text{data})$ is the likelihood function for b; $f'(b)$ is the prior distribution of b; K is a normalizing constant.

The normalizing constant K is often called, in the econometrics literature, the marginal density of the observations or the marginal likelihood (Leamer, 1973; Gaver and Geisel, 1972, and Zellner, 1971) and can be found by

$$K = \int_b L(b|\text{data, model}) \cdot f'(b|\text{model}) \qquad (7.9)$$

Let K, for model i, be K_i. K_i, the marginal likelihood function for model i, can be thought of as the probability of observing the data, given model i, $f(\text{data}|\text{model } i)$.

The posterior density function for $\underline{A},\underline{\theta}$ is calculated from Bayes' rule; it is

$$f''(\underline{A},\underline{\theta}) \propto [\theta_1 \cdot L_1(\underline{A}|\underline{Q}) + \theta_2 \cdot L_2(\underline{A}|\underline{Q})] \cdot$$

$$\cdot [\theta_1 \cdot f'_1 \underline{A}|\theta_1{=}1) \cdot p'(\theta_1{=}1)$$

$$+ \theta_2 \cdot f'_2(\underline{A}|\theta_2{=}1) \cdot p'(\theta_2{=}1)]$$

$$\propto \theta_1 K_1 f''(\underline{A}|\theta_1{=}1) \cdot p'(\theta_1{=}1)$$

$$+ \theta_2 K_2 f''(\underline{A}|\theta_2{=}1) \cdot p'(\theta_2{=}1)$$

$$= \theta_1 \frac{K_1}{K^*} p'(\theta_1{=}1) \cdot f''(\underline{A}|\theta_1{=}1)$$

$$+ \theta_2 \frac{K_2}{K^*} p'(\theta_2{=}1) \cdot f''(\underline{A}|\theta_2{=}1) \qquad (7.10)$$

where K^* is a normalizing constant and equal to

$$K^* = K_1 \cdot p'(\theta_1{=}1) + K_2 \cdot p'(\theta_2{=}1) \qquad (7.11)$$

The posterior model probabilities, $p''(\theta_i)$ are

$$p''(\theta_1{=}1) = \frac{K_1}{K^*} p'(\theta_1{=}1) \qquad (7.12)$$

$$p''(\theta_2{=}1) = \frac{K_1}{K^*} p'(\theta_2{=}1) \qquad (7.13)$$

These posterior probabilities for θ_i are the same as those found by Leamer (1973), Gaver and Geisel (1972), and Smallwood (1968), even though their approaches to the problem were slightly different.

The composite Bayesian distribution can also be found by applying first principles.

$$
\begin{aligned}
\tilde{f}(q) &= \int_{\underline{A},\underline{\theta}} \hat{f}(q|\underline{A},\underline{\theta}) \cdot f''(\underline{A},\underline{\theta})\ d\underline{A}d\underline{\theta} \\
&= \int_{\underline{A},\underline{\theta}} [\theta_1 \cdot f_1(q|\underline{A}) + \theta_2 \cdot f(q|\underline{A})] \cdot \\
&\cdot [p''(\theta_1=1) \cdot f''(\underline{A}|\theta_1) + \\
&\quad p''(\theta_2=1) \cdot f''(\underline{A}|\theta_2=1)] \cdot d\underline{A}d\underline{\theta} \\
&= p''(\theta_1=1) \cdot \tilde{f}_1(q) + p''(\theta_2=1) \cdot \tilde{f}_2(q) \qquad (7.14)
\end{aligned}
$$

The composite Bayesian distribution is simply the Bayesian distributions of the models weighted by the posterior probability that particular model is the true model. This result is extremely convenient and intuitively appealing. The marginal likelihoods K_i, can be found either by analytical or by numerical procedures.

7.3 Marginal Density Function of the Observations

The marginal density function of the observations (marginal likelihood) is calculated from Equation (7.9), and is the probability of the set of observations. The marginal density function depends upon the probability model of the flood discharges, the prior pdf of the parameters of the model and the set of observed flood discharges.

To allow for a more thoughtful analysis of the marginal density function, only those probability models which have a marginal density function that can be calculated analytically, will be included in the model selection analysis of the four probability models. In Chapter 4, the normal, log-normal and exceedance models have marginal density functions that can be evaluated analytically. Appendix F presents the analytical derivation of their marginal density functions for a set of observations.

For the normal model, the marginal density (likelihood) is

$$K_N = \left(\frac{n'}{n''} \right)^{1/2} (2\pi)^{-\nu/2} \frac{\Gamma(1/2\nu'')}{\Gamma(1/2\nu')} \cdot \frac{(1/2\nu' v')^{\nu'/2}}{(1/2\nu'' v'')^{\nu''/2}} \tag{7.15}$$

The parameters are defined in Section 4.2.

For the log-normal model, the marginal density (likelihood) is

$$K_{LN} = \frac{1}{\prod_{i=1}^{n} q_i} \left(\frac{n'}{n''} \right)^{1/2} (2\pi)^{-\nu/2} \frac{\Gamma(1/2\nu'')}{\Gamma(1/2\nu')} \frac{(1/2\nu' v')^{\nu'/2}}{(1/2\nu'' v'')^{\nu''/2}} \tag{7.16}$$

The parameters are defined in Section 4.3.

In Chapter 4, the derivation of the exceedance model considered only those flood discharges greater than a specified base level, such discharges were called exceedance discharges. The probability density function of exceedance discharges was assumed to follow an exponential distribution. The probability density function of discharges less than the base level did not enter into the analysis. One result is that the number of exceedances (discharges larger than the base discharge) is much smaller than the number of flood discharges contained in the flood series considered by the normal or log-normal model. This difference in the number of flood discharges in the series makes the comparison of marginal likelihoods impossible. The difficulties in model comparison, when different 'sample' sizes are observed by each model, has not appeared in the literature and extensive research on this problem is required. To circumvent this problem in model selection, it is assumed that the exceedance model is made up of two parts. The first part is for flood discharges greater than the base level. Here, the assumptions and analyses of Chapter 4 will hold, and the exceedance probability for large discharges will be the same as the probabilities found in Chapter 4. The second part of the model concerns flood discharges less than the base level. Here, flood discharges will assume to follow a uniform distribution. This assumption will lead to marginal likelihoods that are less than those found using the true distribution. This result arises from the uniform pdf being lower than the true pdf, around the mean of the complete flood discharge series, where most of the sample discharges will occur. The lower uniform pdf leads to a smaller sample likelihood function which leads to a smaller marginal likelihood, as can be seen from Equation (7.9). The use of the uniform pdf implies that the posterior probability for the exceedance model is underestimated or conservative.

The probability density function for the (modified) exceedance model is:

$$
\begin{aligned}
f(q|\nu,\alpha) &= \nu\alpha \exp\,[-\alpha(q-q_b)] \quad \text{for} \quad q \geq q_b \\
&= \frac{1-\nu}{q_b} \qquad \text{for} \quad 0 \leq q \leq q_b
\end{aligned}
\qquad (7.17)
$$

For this form of the exceedance model, the marginal likelihood (density) is calculated in Appendix F.3, and is

$$K_E = \left(\frac{1}{q_b}\right)^m \cdot \frac{s'^{u'}}{s''^{u''}} \cdot \frac{\Gamma(u'')}{\Gamma(u')} \cdot \frac{\ell'^{v'}}{\ell''^{v''}} \cdot \frac{\Gamma(v'')}{\Gamma(v')} \qquad (7.18)$$

All parameters are defined in Section 4.5.

7.3.1 Posterior Model Probabilities for Samples Drawn from Known Processes.

Consider a sample of flood discharges that are drawn from some process. The sample has the property that, when the discharges are plotted on normal probability paper, they fall along a perfectly straight line. Table 7.1 presents the flood discharges from such a sample, along with the natural logarithms of the discharges. The probability that the underlying process is normal should be very high. Assuming diffuse prior distributions on the probability model parameters, the numerical values for the parameters of the marginal likelihoods can be found from the data, and are presented in Table 7.2. The marginal likelihoods for the normal model is

$$K_N = 6.477 \times 10^{-69} ;$$

for the log-normal model is

$$K_{LN} = 7.519 \times 10^{-73}$$

and for exceedance model

$$K_E = 8.092 \times 10^{-77}$$

From Equation (7.12), it can be easily seen that the posterior pro-

Table 7.1

Rank	Discharge	L_n Discharge
1	9000	9.105
2	8500	9.048
3	8100	8.9996
4	7800	8.962
5	7550	8.929
6	7300	8.896
7	7100	8.868
8	6900	8.839
9	6700	8.810
10	6500	8.779
11	6300	8.748
12	6100	8.716
13	5900	8.683
14	5700	8.648
15	5500	8.612
16	5200	8.556
17	4900	8.497
18	4600	8.434
19	4000	8.294

Table 7.1: Sample of Flood Discharges from a Normal Process

Table 7.2

Normal Model

m	=	6507 cfs
v	=	1839516 cfs^2
n	=	19 years
ν	=	18 years

Log - Normal Model

m	=	8.759 log cfs
v	=	.04690 log cfs^2
n	=	19 years
ν	=	18 years

Exceedance Model

m	=	16 events
u	=	3 events
v	=	3 events
ℓ	=	1600 cfs
S	=	35 events
q_b	=	8000 cfs

Table 7.2: Marginal Likelihood Parameters for Normal, Log-Normal, and Exceedance Model, for a Sample from a Normal Process.

babilities for the normal model is, virtually, 1.0.

Now consider another sample of flood discharges, which has the property that the discharges plot along a straight line on log-normal probability paper. Such a sample is presented in Table 7.3. Again, assuming diffuse prior distributions on the probability model parameters, the marginal likelihoods are calculated. The marginal likelihood is, for the normal model,

$$K_N = 7.63 \times 10^{-77}$$

for the log-normal model,

$$K_{LN} = 8.06 \times 10^{-77}$$

and for the exceedance model,

$$K_E = 4.23 \times 10^{-78}$$

Table 7.4 presents the numerical values for the parameters of the marginal likelihoods. These values are calculated from the data. Assuming uniform priors on the model probability, that is

$$P'(\theta_N = 1) = P'(\theta_{LN} = 1) = P'(\theta_E = 1) = \frac{1}{3} ,$$

then the posterior probabilities for the three models are calculated from Equation (7.12) and (7.13). The posterior model probabilities are

$$P''(\theta_N) = .4735$$

$$P''(\theta_{LN}) = .5002$$

$$P''(\theta_E) = .0263$$

Table 7.3

Rank	Discharge	L_n Discharge
1	15000	9.616
2	10000	9.210
3	7700	8.949
4	6300	8.748
5	5300	8.575
6	4500	8.412
7	3800	8.243
8	3300	8.102
9	2900	7.972
10	2500	7.824
11	2200	7.696
12	1900	7.550
13	1700	7.438
14	1400	7.244
15	1200	7.090
16	1000	6.908
17	840	6.733
18	640	6.461
19	450	6.109

Table 7.3: Sample of Flood Discharges from a Log-Normal Process.

Table 7.4

Normal Model

m	=	3823 cfs
v	=	13987600 cfs^2
n	=	19 years
ν	=	18 years

Log-Normal Model

m	=	7.84 log cfs
v	=	.91 log cfs^2
n	=	19 years
ν	=	18 years

Exceedance Model

m	=	17 events
u	=	2 events
v	=	2 events
ℓ	=	9000 cfs
S	=	36 events
q_b	=	8000 cfs

Table 7.4: Marginal Likelihood Parameters for Normal, Log-Normal and Exceedance Model, for a Sample from a Log-Normal Process

The reason that the normal model has a high posterior model probability is directly related to the sample and the density functions for the normal model and the log-normal model. Only at low discharges (500 - 3000 cfs) and at high discharges (> 13000 cfs) is the log-normal density function higher than the normal density function. Since the marginal likelihood is a multiplicative process, the ratios of the density functions are of prime importance. The sample did not contain any 'extreme' events, where the ratios are very large; thus the marginal likelihoods will be similar.

The posterior log-normal probability should go to 1 when the number of samples becomes very large and when sampling is from a log-normal process. The test for the log-normal probability to approach 1.0 requires that the ratios K_{LN}/K_N and K_{LN}/K_E be much greater than 1, for large n. This test can be done numerically by sampling from a known log-normal process and calculating the appropriate marginal likelihood ratios.

7.3.2 Posterior Model Probabilities for the Blackstone River, at Woonsocket, R.I.

The Blackstone River, at Woonsocket, R.I., has been analyzed in Chapter 3 for prior information, in Chapter 4 for the Bayesian pdf of flood discharges (for four different probability models), and in Chapter 6 for a decision problem concerning local flood protection. Model uncertainty was not considered in the previous chapter even though competing models were considered. This section calculates the posterior model probabilities. The marginal likelihood function is evaluated using parameter values from Chapter 4. These parameter values for the marginal likelihood functions are summarized in Table 7.5. The marginal likelihood for the normal model is

$$K_N = 7.46 \times 10^{-191} ,$$

for the log-normal model is

$$K_{LN} = 4.76 \times 10^{-160}$$

and for the exceedance model is

$$K_E = 1.14 \times 10^{-156}$$

Assuming uniform prior probabilities on the three models, the posterior probabilities for the models are

$$P''(\theta_N = 1) = 0.0$$

$$P''(\theta_{LN} = 1) = .00418$$

$$P''(\theta_E = 1) = .99582$$

The composite Bayesian distribution of flood discharges is, from Equation (7.14)

$$\tilde{f}(q) = .99582\ \tilde{f}_E(q) + .00418\ \tilde{f}_{LN}(q) \qquad (7.19)$$

where $\tilde{f}_E(q)$ is the Bayesian pdf for the exceedance model, and

$\tilde{f}_{LN}(q)$ is the Bayesian pdf for the log-normal model.

The composite Bayesian distribution of Equation (7.19) is the pro-

Table 7.5

Normal Model

n'	=	7 years,	n''	=	44 years
ν	=	36 years	ν''	=	43 years
v'	=	9.22×10^6 cfs^2	v''	=	24.7×10^6 cfs^2

Log-Normal Model

n'	=	4 years	n''	=	41 years
ν'	=	36 years	ν''	=	40 years
v'	=	.22 log cfs^2	v''	=	.689 log cfs^2

Exceedance Model

u'	=	6 events	u''	=	11 events
v'	=	3 events	v''	=	8 events
S'	=	50 years	S''	=	87 (S''+m=119) years
ℓ'	=	10850 cfs	ℓ''	=	49468 cfs
m	=	32 events	n	=	5 events
q_b	=	8500 cfs			

Table 7.5: Marginal Likelihood Parameters for Normal, Log-Normal and Exceedance Model, for the Blackstone River, Woonsocket, R.I.

bability model which should be used in making inferences about future flood discharges. The composite Bayesian model rationally accounts for both parameter and model uncertainty. It is interesting to note that the form of composite Bayesian model is not fixed but is dynamic and changes, as more data becomes available.

7.4 Decision Making with Model Uncertainty

The theoretically correct Bayesian procedure for making decisions, when model uncertainty exists, is to calculate the composite Bayesian distribution of flood discharges, which is then used, in conjunction with an appropriate utility function, to maximize expected utility. Chapter 6 discusses the decision aspects of maximizing expected utility. In Chapter 6, four competing inference models were considered. These inference models can now be replaced by the composite Bayesian distribution of flood discharges, which accounts for model uncertainty.

In many cases, a decision maker may want to chose a 'best' model, from the set of competing models. This chosen model will lead to a 'best' decision. It is recognized that basing decisions on one model, from a set of models, is a sub-optimal Bayesian procedure to the procedure of basing decisions upon the composite Bayesian distribution (Gaver and Geisel, 1972). A decision maker may consider decisions based upon one model for a variety of reasons. The calculation or application of the composite Bayesian distribution may be computationally difficult; or the sensitivity of the decisions to the probabilities of the models may be of interest. In either case, the problem of choosing one model, from a set of models, is equivalent to choosing the 'best' overall decision, from a set of 'best' decisions (based on each model being the true model). The application of decision analysis is most effective in choosing the 'best' overall decision, taking into account model uncertainty. Of the procedures within decision analysis, the normal-form-analysis (Raiffa, 1968), not only identifies the best overall decision but also allows for sensitivity analysis of the decision to the probability that a particular model is the true model.

The most common form of decision analysis is the extensive-form, which consists of four basic steps (Raiffa, 1968):

1. Chart the decision-flow diagram
2. Assign payoffs or utilities for outcomes.
3. Assign or determine probabilities at all chance forks.
4. Average out and fold back (find expected utilities).

The normal form of analysis does not require, initially, that all probabilities be evaluated. Instead, expected utility of each possible decision is determined, conditional to the unknown probabilities, $E[u|d_j, p(\theta_i)]$. The best decision will depend upon the utility of that decision and the probability of the model, $p(\theta_i)$. The following example will show the interaction between choosing a 'best' decision and evaluating the model probabilities.

Consider the decision example for the Blackstone River, presented in Chapter 6. Assume that either the log-normal model or the exceedance model is the true probability model of flood discharges. The decision problem of Chapter 6 can be represented by the simple decision tree presented in Figure 7.1. The evaluation of the outcomes was done in Chapter 6 and was presented in Table 6.3.

The extensive-form of analysis would evaluate the probabilities $P(\theta_{LN})$ and $P(\theta_E)$. Then, the expected utility of the decisions is calculated, and the decision that maximizes expected utility is picked as being the 'best' decision. The normal-form of analysis graphs the expected utility of each decision, conditional upon the model probabilities. The joint conditional evaluation diagram is presented in Figure 7.2. Lines of constant expected utility can also be drawn in Figure 7.2. These lines are represented by an equation of the form

$$E[u] = P(\theta_{LN}) \cdot u|\theta_{LN} + P(\theta_E) \cdot u|\theta_E$$

As lines of expected utility move away from the origin to the northwest, their value increases. Thus, the efficient set of decision, those that will maximize expected utility, will lie along the northwest boundary of the enclosed region in the joint conditional evaluation space. In Figure 7.2, the efficient set of decisions is made up of d_2 and d_3. Which decision maximizes the expected utility depends upon the model probabilities $P(\theta_E)$ and $P(\theta_{LN})$. Exact evaluation of the model probabilities is not necessarily needed to determine to best decisions. If $P(\theta_E)$ falls into a specific interval, then d_3 will maximize the expected utility; if $P(\theta_E)$ falls outside this interval, then d_2 will maximize expected utility. For the sample

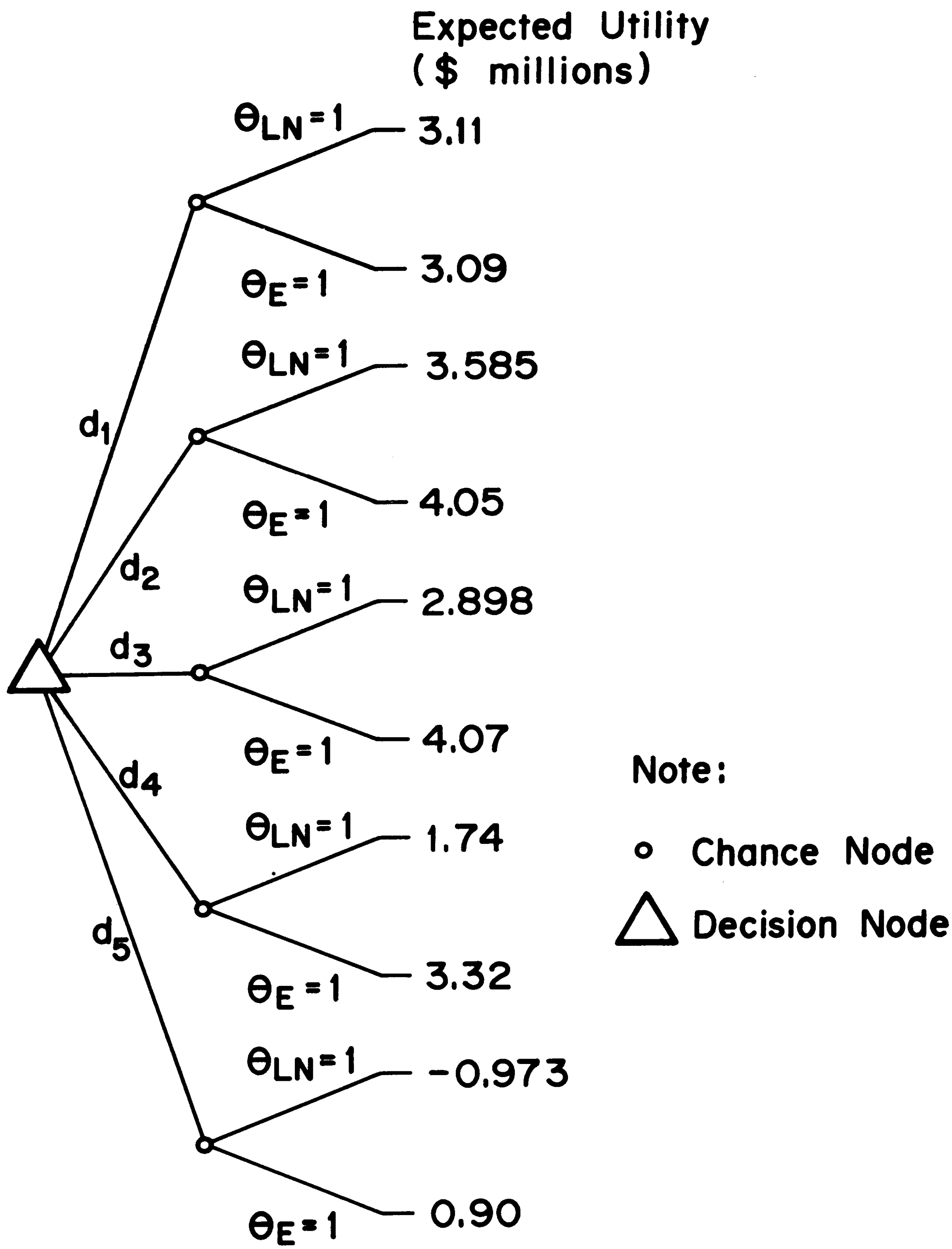


Figure 7.1 Decision Tree for Model Uncertainty

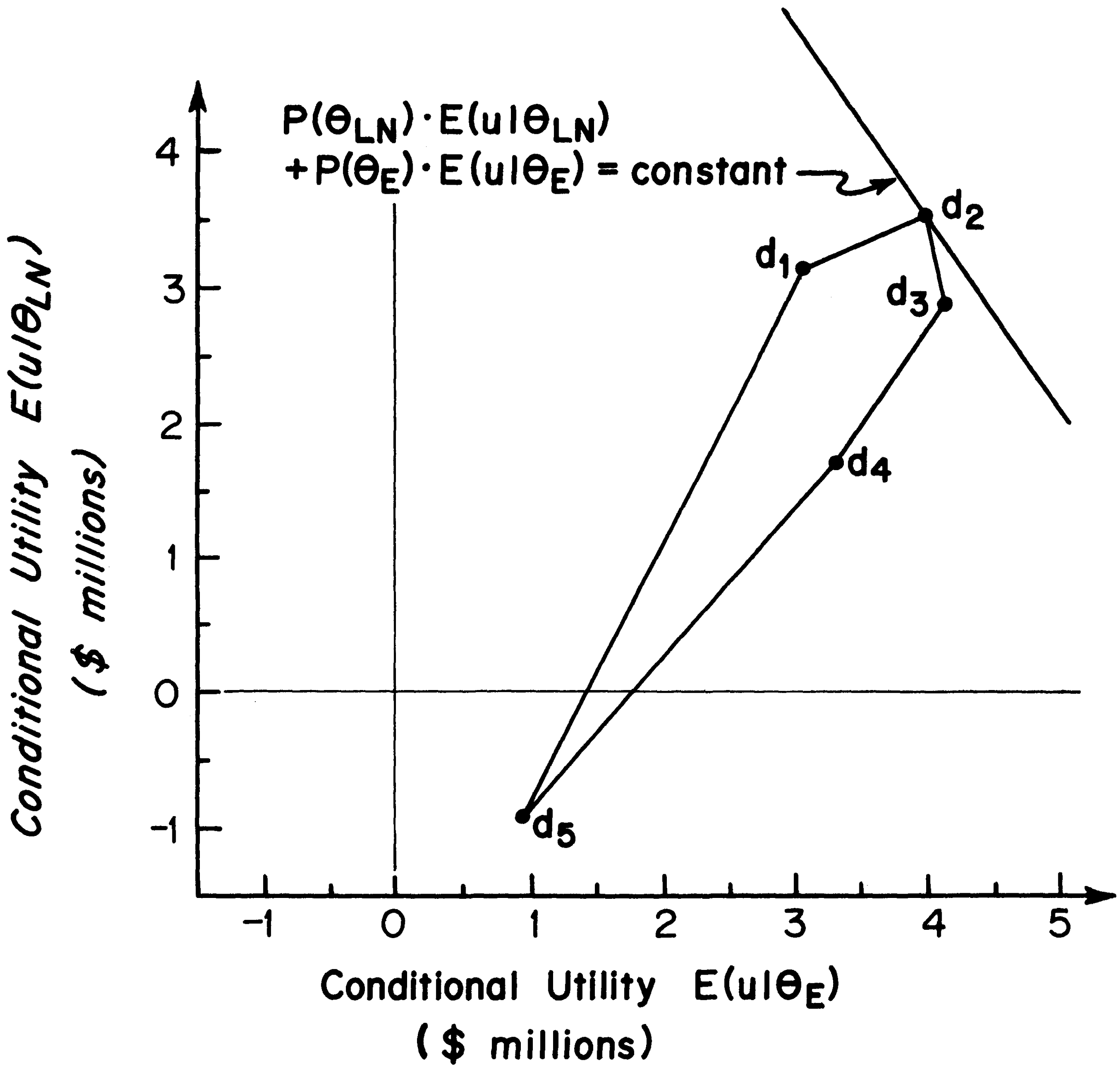


Figure 7.2 Conditional Evaluation Diagram

presented here, the probability intervals, and the corresponding 'best' decision, are as follows:

If $.861 \quad P(\theta_E) \leq 1 \quad (0 \leq P(\theta_{LN}) \leq .138)$

then choose d_3

If $0 \leq P(\theta_E) \leq .861 \quad (.138 \leq P(\theta_{LN}) \leq 1)$

then choose d_2.

The advantages of the normal-form of decision analysis is that exact evaluation of probabilities is not needed. The sensitivity of the expected utility and of the 'best' decision to the model probabilities is clearly seen.

If the decision space is discrete (consists of only these 5 decisions), then the 'best' decision from the normal form of analysis, using the posterior model probabilities of the previous section, will be the same decision as would be found using the composite Bayesian distribution. If the decision space is continuous, then the decisions found by the two procedures will be different. The decision from the composite Bayesian distribution will be optimal and from the normal analysis, merely good.

7.5 Conclusions

This chapter considered the problem of model uncertainty. When there is a set of competing probability models for flood discharges, Bayesian analysis leads to a composite Bayesian model. The composite Bayesian model is a linear model consisting of the Bayesian distribution of the individual models, weighted by the posterior model probability that the individual model is the true model. The posterior model probabilities are calculated from the marginal likelihood function of the observed data and the prior model probability.

Decision making under model uncertainty follows the procedures of Chapter 6, but the inference model of the flood discharges is now the composite Bayesian model. A sub-optimal Bayesian procedure is to apply a normal form of decision analysis to a model selection problem. First the 'best' decision is found for each model, and then from this set an overall, 'best' decision is calculated, by considering model probabilities and expected utilities. In general, the model selection procedure is sub-optimal to decision making with the composite Bayesian distribution of flood discharges.

CHAPTER 8

Summary and Conclusions

Inference -making probability statements about future state of nature- and decisions -determining engineering variables such as the height of a dike- together form an indivisible pair which results in an engineering design. This procedure of inferences and decisions is called decision making, and it contains many sources of uncertainty. This study is concerned with the analysis of the uncertainty and with the development of procedures to rationally design for extreme hydrologic events in the light of such uncertainty.

In Chapter 2, arguments are presented which advocate the use of Bayesian statistical decision theory for decision making. Decision theory allows the decision maker to consider, together, the uncertainty of the modelled process, the quantifying of decision outcomes and the preferences for these outcomes. The decision theory approach to making decisions seems to be a more rational approach than other procedures which separate inferences, decisions, and preferences.

The results of this study confirm the arguments that Bayesian statistical procedures can be used to make engineering decisions. Most of the emphasis of this study is focused upon the uncertainty of the modelled process of extreme flood discharges and not upon the preferences toward decision outcomes. The criterion applied for evaluating alternative decisions is the maximization of expected monetary benefits. A more appropriate utility function, for real world flood control decisions, would probably include monetary benefits, social benefits (such as decrease in loss of life, or reduction in the disruption of community services) and risk adversion towards large floods. Considerable work has been done in multiple utility theory, but more work is required in assessing utility functions before they can be applied to general engineering projects.

When discussing inferences about flood discharges, it is important to keep two sources of uncertainty separate. The first source of uncertainty is the 'modelled' uncertainty of the underlying stochastic process. This stochastic process is the future flood discharges. The second source of uncertainty is statistical uncertainty. Statistical uncertainty is concerned with quantities which are 'fixed but unknown' due to a lack of information. Such uncertain quantities include the uncertainty in the form of the true underlying stochastic process (model uncertainty) and the uncertainty in the values of the parameters of the probability models that are used to represent the underlying process (parameter uncertainty). The uncertainty in these quantities can be reduced with additional information.

Chapter 3 considers the use of information, other than historical records, to reduce statistical uncertainty. A regional flood regression and an analytical flood frequency analysis (Eagleson, 1972) are two sources of information that are studied in detail. It is shown that they provide information equivalent to between 4 and 7 years of historical record for the river basin studied. This information could be significant in the analysis of river basins with short historical records. Research in evaluating sources of prior information would be a significant contribution to Bayesian analysis. There is room here for a large amount of fruitful research.

One source of prior information, not considered directly in this study, is rainfall-runoff simulation. Chapter 5 did, though, analyze the uncertainty in the flood discharge from a rainfall-runoff analysis due to the uncertainty in the rainfall and infiltration processes. The theory for considering all uncertain parameters is presented. The work in Chapter 5 is especially useful for flood analysis in regions with no historical records or in regions that have undergone urbanization. Recently rainfall-runoff simulation models have been applied to find flood frequency curves (Leclerc and Schaake, 1973; Ott and Linsley, 1972; among others) for urbanized areas or areas with limited historical records. These studies have not analyzed the uncertainty in the flood frequency curve, due to parameter uncertainty. It is evident from the results of Chapter 5 that ignoring such uncertainty can lead to grave errors in decision.

Chapter 4 shows how the prior information of Chapter 3 and the historical record of flood discharges can be combined, with the model of the true stochastic process of flood discharges, to yield a 'Bayesian probability distribution' of flood discharges, which fully accounts for parameter uncertainty. The Bayesian probability distribution is obtained by taking the probability model of flood discharges, which has uncertain parameters, and integrating that model, weighted by the probability of observing particular values of the uncertain parameters, over all values of the uncertain parameters.

The Bayesian probability distribution of flood discharges is obtained for four different models of the underlying stochastic process. The four models are: a normal model, a log-normal model, a gamma-1 model and an exceedance model. The first three models consider the complete annual series of flood discharges. The fourth model only considers flood discharges that exceed a particular level, and such exceedance discharges are exponentially distributed.

The resulting models are applied to the Blackstone River, and comparisons are made among the different flood frequency curves from the models. Comparisons between the Bayesian probability model and the classical approach to frequency analysis is also shown. The general result is that the Bayesian procedure will lead to higher flood discharges for the same exceedance probability, which leads to more conservative designs. This conservative design reflects the parameter uncertainty, and current practice of using point estimates for uncertain parameters should be reconsidered since it may lead to incorrect decisions.

In Chapter 7, the statistical uncertainty of which model represents the true stochastic process is analyzed. This analysis of model uncertainty leads to a composite Bayesian distribution. The composite Bayesian distribution is a linear model of the individual Bayesian probability models of Chapter 4,weighted by the posterior probability that a particular model is the true model. The composite Bayesian probability model accounts for all sources of statistical uncertainty - both parameter uncertainty and model uncertainty. This model is the model that should be used in the decision analysis, for it best represents the knowledge of the decision maker about future flood discharges.

Chapter 6 applies the decision analysis technique to a case study. The decision problem is to determine the optimal capacity of local flood protection works for Woonsocket, Rhode Island. Initially, the analysis is performed for each Bayesian model of flood discharges. The sensitivity of the optimal designs to budgetary considerations is also analyzed.

In Chapter 7, the optimal designs with model uncertainty are considered. The theoretically correct procedure is to apply the composite Bayesian distribution for flood discharges. Chapter 7 also presents a procedure for finding decisions when the composite Bayesian distribution is computationally difficult to obtain. It is also possible to perform sensitivity analysis on the optimal decisions, due to model uncertainty.

This study presents procedures which should help decision makers consider uncertainty in a more complete manner, allowing them to design more rationally under uncertainty.

BIBLIOGRAPHY

Andersen, Hiskey, Lachawathana (1971), "Application of Statistical Decision Theory to Water Use Analysis in Sevier County, Utan", Water Resources Research, Vol. 7, No. 3, June, 1971

Benson, M.A. (1962), "Factors Influencing the Occurrence of Floods in a Humid Region of Diverse Terrain," U.S. Geological Survey Water Supply Paper No. 1580-B, U.S. Government Printing Office, Washington, D.C.

Benjamin, Jack R. and C. Allin Cornell (1970), Probability, Statistics and Decision for Civil Engineers, McGraw-Hill, New York, 1970

Birnbaum, Allan (1962), "Foundations of Statistical Inference" Journal of American Statistical Association, Vol. 57, pp. 269-305

Chambers, M.L. (1970), "A Simple Property with Strikingly Different Frequentist and Bayesian Solutions", Journal of the Royal Statistical Society, Series B, Vol. 32, No. 2, pp. 278-282

Chow (1964), "Statistical and Probability Analysis of Hydrologic Data", in Handbook of Applied Hydrology, edited by V.T. Chow, McGraw-Hill, New York, 1964

Cornfield, Jerome (1909), "The Bayesian Outlook and its Application", Biometrics, Vol. 25, pp. 617-657

Corps of Engineers (1958), "Blackstone River Basin, Lower Woonsocket, Rhode Island", Report from Secretary of the Army to the U.S. Senate, 85th Congress, 2nd Session, TC 423, A3, S.D. 6-87, 1958

Davis, Kisiel, and Duckstein (1972), "Bayesian Decision Theory Applied to Design in Hydrology", Water Resources Research Vol. 8, No. 1, February, 1972

Dumonceaux, Anthe, Haas (1973), "Lieklihood Ratio Test for Discrimination Between Two Models with Unknown Location and Scale Parameters", Technometrics, Vol. 15, No. 1, February, 1973

Draper, N.R. and H. Smith (1966), Applied Regression Analysis, John Wiley & Sons, Inc., New York

Eagleson, P. (1972), "Dynamics of Flood Frequency", Water Resources Research, Vol. 8, No. 4, August, 1972

Fishburn, P.C. (1964), Decision and Value Theory, John Wiley & Sons, New York, 1964

Fishburn, P.C. (1969), "Preferences, Summation and Social Welfare Functions", Management Science, Vol. 16, pp. 179-186 (1969)

Gaver and Geisel (1972), Discriminating Among Alternative Models: Bayesian and Non-Bayesian Methods", to appear in Frontiers of Econometrics, Paul Zaremkka, editor, Academic Press, forthcoming

Gray, D.M. (1970), ed., Handbook on the Principles of Hydrology, Canadian National Committee for the International Hydrologic Decade, Ottawa, Canada

Grayman and Eagleson (1971), "Evaluation of Radar and Raingage Systems for Forcasting", Ralph M. Parsons Laboratory for Water Resources and Hydrodynamics, Technical Report No. 138, Department of Civil Engineering, M.I.T., Cambridge, Mass.

Ibbitt, R.P. (1972), "Uncertainties in Deterministic Models", General Report Presented at the International Symposium on Uncertainties in Hydrologic and Water Resource Systems, Tucson, Arizona, December, 1972

Jeffreys, H. (1961), Theory of Probability, 3rd Edition, Oxford University Press, 1961

Kaufman (1972), "Two Distributions Involving a Bessel Function", Journal of the American Statistical Association, Vol. 67, No. 340, Theory and Methods Section, December, 1972

Keeney, R.L. (1969), "Multidimensional Utility Functions: Theory, Assessments and Application", T,R, No. 43, Operations Research Center, Massachusetts Institute of Technology, Cambridge, Mass. October, 1969

Keeney, Ralph and Howard Raiffa (1971), Untitled Manuscript, to be published

Keynes, John M. (1921), A Treatise on Probability, Macmillan, 1921

Kirkwood, C.W. (1972), "Decision Analysis Incorporating Preferences of Groups", T.R. 74, Operations Research Center, Massachusetts Institute of Technology, Cambridge, Mass., June, 1972

Leamer (1973), "Probabilities of Linear Hypothesis", unpublished working paper, Harvard University

Leclerc, G. and J.C. Schaake (1973), "Methodology for Assessing the Potential Impact of Urban Development on Urban Runoff and the Relative Effeciency of Runoff Control Alternatives", Ralph M. Parsons Laboratory Technical Report No. 167, Department of Civil Engineering, Massachusetts Institute of Technology, March, 1973

Lindley, D.V. (1972), Bayesian Statistics, A Review, Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, 1972

Matalas, N.C. and M.A. Benson (1961), "Effect of Interstation Correlation on Regression Analysis", Journal of Geophysical Research, Vol. 66, No. 10, pp. 3285 - 3293

Matalas, N.C. and E.J. Gilroy (1968), "Some Comments on Regionalization in Hydrologic Data", Water Resources Research, Vol. 4, No. 6, pp. 1361 - 1369

McGilchrist and Woodyer (1970), "Recurrence Intervals between Exceedances of Selected River Levels", 3. Estimation and Use of a Prior Distribution", Water Resources Research, Vol. 6, No. 2, April, 1970

Pratt, Raiffa, and Schlaifer (1965), Introduction to Statistical Decision Theory, McGraw-Hill, New York, (preliminary ed.)

Pratt, John W. (1965), "Bayesian Interpretation of Standard Inference Statements", Journal of the Royal Statistical Society, Series B, Vol. 27, pp. 169 - 203

Raiffa, Howard (1968), Decision Analysis, Introductory Lectures on Choices under Uncertainty, Addeson-Wesley, Reading, Mass., 1968

Raiffa, H. and R. Schlaifer (1961), Applied Statistical Decision Theory, M.I.T. Press, Massachusetts Institute of Technology, Cambridge, Mass.

Ramsey, F.P. (1926), Reprinted in Studies in Subjective Probabilities, ed. Kyberg and Smokler, Wiley, New York, 1964

Savage, L.J. (1954), The Foundation of Statistics, John Wiley and Sons, New York, 1954

Shane and Gaver (1970), "Statistical Decision Theory Techniques for the Revision of Mean Flood Flow Regression Estimates", Water Resources Research, Vol. 6, No. 6, December, 1970

Shane and Lynn (1964), "Mathematical Model for Flood Risk Evaluation", Journal of Hydraulics Division, American Society of Civil Engineers, Vol. 90, No. HY-6, 1964

Shaw, Elizabeth and P.P. Lynn (1972), "Areal Rainfall Evaluation Using Two Surface Fitting Techniques", Hydrological Sciences Bulletin, Vol. 17, No. 4, December, 1972

Smallwood (1968), "A Decision Analysis of Model Selection", IEEE Transactions on Systems Science and Cybernetics, Vol. SSC-4, No. 3, September, 1968

Stein, C.M. (1951), "A Property of Some Tests of Composite Hypothesis", Annals of Mathematical Statistics, Vol. 22, pp. 475 - 476

Thomas, D.M. and M.A. Benson (1970), "Generalization of Streamflow Characteristics from Drainage-Basin Characteristics", U.S. Geological Survey Water-Supply Paper No. 1975, U.S. Government Printing Office, Washington, D.C.

Todorovic and Zelenhasic (1970), "A Stochastic Model for Flood Analysis", Water Resources Research, Vol. 6, No. 6, December, 1970

Tschannerl (1971), "Designing Reservoirs with Short Streamflow Records", Water Resources Research, Vol. 7, No. 4, August, 1971

von Neumann, John and O. Morgenstern (1947), Theory of Games and Economic Behavior, 2nd Ed., Princeton University Press, Princeton, New Jersey, 1947

Wald, Abraham (1950), Statistical Decision Functions, John Wiley and Sons, New York, 1950

Zellner (1971), "Bayesian and Non-Bayesian Analysis of the Log-Normal Distribution and Log-Normal Regression", Journal of the American Statistical Association, Vol. 66, No. 334, Theory and Methods Section, June, 1971

Zellner (1971), An Introduction to Bayesian Inference in Econometrics, Wiley, New York

Addum:

Major, David, (1973) "Multiobjective Water Resource Planning," American Geophysical Union Water Resources Monograph 4, Draft Copy, Cambridge, Mass.

Ott and Linsley (1972) "Streamflow Frequency Using Stochastically Generated Hourly Rainfall," Proceedings of the International Symposium or Uncertainties in Hydrologic and Water Resources Systems Dec. 11-14, 1972, Tucson, Ariz.

Appendix A

Proof that Bayesian pdf of Flood Discharges from a Normal Process is Student.

This appendix presents the proof that the Bayesian pdf for q is student, when the underlying process is a normal process with uncertain parameters mean μ and precision h. From Equation (4.1)

$$\tilde{f}(q) = \int_\mu \int_h f_N(q|\mu,h) \cdot f''(\mu|h) \cdot f''(h)\, d\mu\, dh \qquad \text{(A-1)}$$

solving for $f(q|h) = \int_\mu f_N(q|\mu,h) \cdot f(\mu|h)\, d\mu$ (A-2)

which is

$$\int_\mu \frac{1}{\sqrt{2\pi}} h^{1/2} \exp[-\frac{1}{2} h(q-\mu)^2] \cdot \frac{1}{\sqrt{2\pi}} n^{1/2} h^{1/2} \exp[-\frac{1}{2} nh(\mu-m)^2] d\mu$$

(A-3)

substituting for $f_N (q \mid \mu,h) \sim N(\mu,h)$

and $f(\mu|h) \sim N(m,nh)$

Solving (A-3)

$$\tilde{f}(q \mid h) = \frac{1}{\sqrt{2\pi}} \int_\mu \frac{1}{\sqrt{2\pi}} n^{1/2} h \exp \left\{- \frac{1}{2} h \left[\mu - \frac{(q+nm)}{n+1}\right]^2 + \frac{n}{n+1}(q-m)^2 \right]\right\} d\mu$$

$$= (\frac{n}{n+1})^{1/2} \frac{h^{1/2}}{\sqrt{2\pi}} \exp[-\frac{1}{2} h(\frac{n}{n+1}) \cdot (q-m)^2] \int_\mu f_N(\frac{q+nm}{n+1}, (n+1)\, h)\, d\mu$$

$$\tilde{f}(q|h) = \frac{(r\, h)^{1/2}}{\sqrt{2\pi}} \exp\, [-\frac{1}{2} rh(q-m)^2] \qquad \text{(A-4)}$$

where $r = \frac{n}{n+1}$

Thus $\tilde{f}(q|h)$ is distributed $N(m,rh)$

$$\tilde{f}(q) = \int_h f_N(q|m,rh) \cdot f_{\gamma 2}''(h|v,\nu)\, dh \qquad \text{(A-5)}$$

where h is distributed gamma-2, as

$$f_{\gamma 2}''(h|v,\nu) = \frac{(\frac{1}{2}\nu v)^{\nu/2}}{\Gamma(\nu/2)} \cdot h^{\nu/2-1} \cdot \exp(-\frac{1}{2}\nu v h) \qquad \text{(A-6)}$$

$$\tilde{f}(q) = \int \frac{(rh)^{\frac{1}{2}}}{\sqrt{2\pi}} \exp[-\frac{hr}{2}(q-m)^2] \cdot \frac{(\frac{1}{2}\nu v)^{\nu/2}}{\Gamma(\nu/2)} \cdot h^{\nu/2-1} \exp[-\frac{h}{2}\nu v]\, dh$$

(A-7)

now let $\nu' = \nu + 1$

and combine $v' = \frac{r(q-m)^2 + \nu v}{\nu'}$

so that (A-7) can now be written as:

$$\tilde{f}(q) = \frac{r^{\frac{1}{2}}}{\sqrt{2\pi}} \frac{(\frac{1}{2}\nu v)^{\nu/2}}{\Gamma(\nu/2)} \cdot h^{\nu'/2-1} \cdot \exp[-\frac{h}{2}\nu' v']\, dh \qquad \text{(A-8)}$$

$$= \frac{r^{\frac{1}{2}}}{(2\pi)^{\frac{1}{2}}} \frac{(\frac{1}{2}\nu v)^{\nu/2}}{\Gamma(\nu/2)} \cdot \frac{\Gamma(\nu'/2)}{(\frac{1}{2}\nu' v')^{\nu'/2}} \cdot \int \frac{(\frac{1}{2}\nu' v')^{\nu'/2}}{\Gamma\, \nu'/2} \cdot h^{\nu'/2-1} \cdot$$

$$\cdot \exp[-\frac{h}{2}\nu' v']\, dh$$

after simplifying

$$\tilde{f}(q) = \frac{\nu^{\nu/2}}{B(\frac{1}{2},\frac{1}{2}\nu)} \cdot [\nu + r\frac{(q-m)^2}{v}]^{-\frac{(\nu+1)}{2}} \cdot (\frac{r}{v})^{\frac{1}{2}} \qquad \text{(A-9)}$$

$$= \frac{1}{B(\frac{1}{2}, \frac{1}{2}\nu)} \cdot [1 + r \frac{(q-m)^2}{\nu v}]^{-(\frac{\nu+1}{2})} \cdot (\frac{r}{\nu v})^{\frac{1}{2}}$$

where $B(\frac{1}{2}, \frac{1}{2}\nu) = \frac{\pi^{\frac{1}{2}}\Gamma(\nu/2)}{\Gamma(\nu/2 + 1/2)}$

which is Student. The moments of $\tilde{f}(q)$ are

$$\mu_q = m \quad \text{(A-10)}$$

$$\sigma_q^2 = \frac{v}{r} (\frac{\nu}{\nu-2}) \quad \text{(A-11)}$$

$$= \left(\frac{n+1}{n}\right) v \left(\frac{\nu}{\nu-2}\right)$$

Appendix B

The Bayesian Distribution from Log-Normal Process

In Section 4.3 the analysis of the log-normal (LN) process was approached by recognizing that if q is LN, then $x = \ell n\, q$ is normally distributed. It was shown in Section 4.3 that the Bayesian distribution of x is student, thus the Bayesian distribution of q will be log-student. This result was found by applying derived distribution theory. A disadvantage of this procedure is that one can not make inferences on or from the distributions of the parameters of the log-normal process.

This appendix derives the Bayesian pdf of the log-normal process by first deriving the posterior pdf of the parameters, then integrating over the parameters to derive the Bayesian distribution on the annual flood.

If $x_i = \ell n\, q_i$ is normally distributed with mean μ and variance σ^2 then q_i, by definition, is log-normally distributed and has mean $\eta = \exp\,[\mu + \sigma^2/2\,]$.

B.1 Posterior Distribution of Mean η

It can be shown (Raiffa and Schlaifer, 1961; Zellner, 1971) that given n observations on x_i, $\underline{x}$, the likelihood function for μ and σ is normal-gamma and of the form

$$\ell(\mu, \sigma \mid \underline{x}) \propto \sigma^{-1} \exp\left[\frac{-n}{2\sigma^2}(\mu-\overline{x})^2\right] \cdot \sigma^{-(n-1)} \exp\left[\frac{-\nu s^2}{2\sigma^2}\right] \qquad \text{(B-1)}$$

where n = number of observations

$$\nu = n - 1$$

$$\bar{x} = \frac{1}{n} \sum^{n} x_i$$

$$\nu s^2 = \sum^{n} (x_i - \bar{x})^2$$

Let $\Theta = \ln \eta = \mu + \sigma^2/2$. The posterior pdf for Θ, $f''(\Theta|\sigma,\underline{x})$, follows from

$$f''(\Theta|\sigma, \underline{x}) \propto \ell(\Theta|\sigma,\underline{x}) \cdot f'(\Theta \mid \sigma,\underline{x}) \qquad \text{(B.2)}$$

From Equation (B.1) it follows that

$$\ell(\Theta|\sigma_1 x) \propto \sigma^{-1} \exp\left[-\frac{n}{2\sigma^2}(\Theta - \bar{x} - \sigma^2/2)^2\right] \qquad \text{(B.3)}$$

and from inspection of (B.3) the natural conjugate prior will be of the form

$$f'(\Theta|m', n'/\sigma^2) \propto \sigma^{-1} \exp\left[-\frac{n'}{2\sigma^2}(\Theta - m')^2\right] \qquad \text{(B.4)}$$

Thus

$$f''(\Theta/\sigma,\underline{x}) \propto \sigma^{-1} \exp\left[-\frac{n''}{2\sigma^2}(\Theta - m'' - n\cdot\sigma^2/2\cdot n'')^2\right] \qquad \text{(B.5)}$$

where

$$n'' = n' + n$$

$$m'' = \frac{1}{n''}(n' \cdot m' + n \cdot \bar{x})$$

Similarly, the posterior pdf $f''(\sigma|\underline{x})$ can be found from the product of its likelihood and prior pdf. It can be shown that

$$\ell(\sigma|\underline{x}) \propto \sigma^{-\nu} \exp\left[-\frac{\nu s^2}{2\sigma^2}\right] \tag{B.6}$$

The natural conjugate prior to the likelihood function of σ is Inverted Gamma-2 with parameters (ν', s'^2)

$$f'(\sigma|\nu', s'^2) \propto \exp\left[-\frac{\nu' s'^2}{2\sigma^2}\right] \cdot \sigma^{-(\nu'+1)} \tag{B.7}$$

And the posterior pdf for σ is:

$$f''(\sigma|\nu'', s''^2) \propto \exp\left[-\frac{\nu'' s''^2}{2\sigma^2}\right] \cdot \sigma^{-(\nu''+1)} \tag{B.8}$$

where $\nu'' = \nu' + \nu$

$$s''^2 = \frac{1}{\nu''}\left[\nu' \cdot s'^2 + \nu \cdot s^2\right]$$

To obtain $f''(\Theta|\underline{x})$ we apply

$$f''(\Theta|\underline{x}) \propto \int_\sigma f''(\Theta|\sigma,\underline{x}) \cdot f''(\sigma) \; d\sigma \tag{B.9}$$

$$\propto \int_\sigma \sigma^{-(\nu''+2)} \cdot \exp\left[-\frac{n''}{2\sigma^2}(\Theta - m'' - n\sigma^2/2n'')^2 + \frac{\nu'' s''^2}{2\sigma^2}\right] d\sigma \tag{B.10}$$

$$\propto \exp(n''\cdot\Theta/2) \cdot \int_\sigma \sigma^{-(\nu''+2)} \cdot \exp\left\{-\frac{\nu'' s^2 + n''(\Theta - m'')^2}{2\sigma^2} + \right.$$

$$\left. + \frac{n^2}{n''}\frac{\sigma^2}{8}\right]\right\} d\sigma \tag{B.11}$$

The Jacobian transform from $f(\Theta)$ to $f(\eta)$ is $\frac{1}{\eta}$, thus

(B.11) becomes

$$f''(\eta|\underline{x}) \propto \eta^{(n-2)/2} \cdot \int (\cdot) \; d\sigma$$

This integral can be evaluated by recognizing that it may be rewritten as

$$\int_0^\infty t^{\nu^*-1} \cdot \exp\left[-\left(1/4\,t + at\right)\right] dt \qquad \text{(B.12)}$$

where $\nu^* = (\nu''+1)/2$

$t = 2n''/n^2\,\sigma^2$

$a = \frac{n^2}{4n''} \cdot \left[\nu''\,s''^2 + n''\,(\ln\eta - m'')^2\right]$

The value of the integral (Zellner, 1971) is $2(1/4a)^{\nu'/2} K_{\nu'}(\sqrt{a})$ where $K_{\nu'}(\sqrt{a})$ denotes a modified K Bessel function. Substituting into (B.11) gives

$$f''(\eta|\underline{x}) \propto \eta^{(n-2)/2} \cdot 2(1/4a)^{\nu'/2} K_{\nu'}(\sqrt{a}) \qquad \text{(B.13)}$$

Figure B-1 shows the probability density function of the mean of the discharges, η, for the case where

$n = n'' = 37$

$s^2 = .27$

$m'' = 8.6$

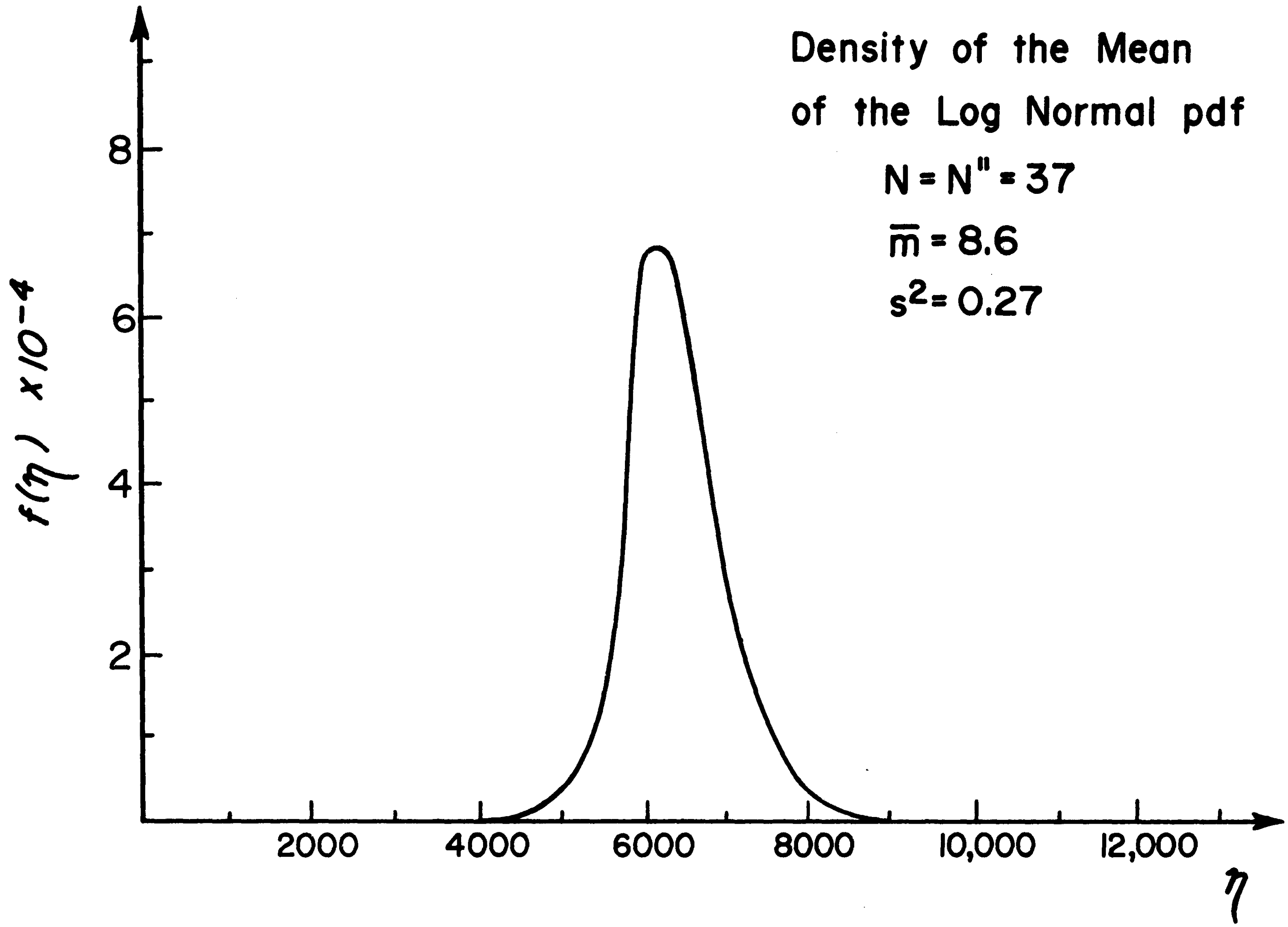


Figure B.1: Density of the Mean of the Log-Normal pdf

It is the case that the posterior mean of η does not exist (Zellner, 1971) but from Figure B-1 it is noted that $f(\eta)$ is unimodel and positively skewed.

B.2 Bayesian Distribution of Discharges, $\tilde{f}(q)$.

From the posterior distributions represented by Equations (B.9) and (B.10) the Bayesian distribution for discharges can be found by

$$\tilde{f}(q) = \int_{\Theta}\int_{\sigma} f_{LN}(q|\Theta,\sigma) \cdot f''(\Theta|\sigma,\underline{x}) \cdot f''(\sigma|\underline{x})\, d\Theta d\sigma \tag{B.14}$$

where

$$f_{LN}(q|\Theta,\sigma) \propto q^{-1}\sigma^{-1}\exp[-\frac{1}{2\sigma^2}(\ell n\, q - \Theta + \sigma^2/2)^2]$$

$$\Theta = \ell\eta$$

Substitute for $f''(\Theta|\sigma,\underline{x})$ from Equation (B.5) and for $f''(\sigma|\underline{x})$ from (B.8) into Equation (B.14). $\tilde{f}(q)$ can then be represented by

$$\tilde{f}(q) \propto q^{-1}\exp[\Theta/2 + n'' \cdot \Theta/2] \cdot \int_{\Theta}\int_{\sigma} \sigma^{-(\nu''+3)} \cdot$$

$$\cdot \exp\left\{ -\left| \frac{(\Theta-\ell n\, q)^2 + n''(\Theta-m'')^2 + \nu'' s^{2''}}{2\sigma^2} + \frac{n^2+n''}{n''}\ \frac{\sigma^2}{8} \right| \right\} d\Theta d\sigma \tag{B.15}$$

The integral over σ can be expressed by a modified Bessel function of type K. Thus (B.15) can be expressed, after substitution for Θ as:

$$\tilde{f}(q) \propto q^{-1} \int_{\eta=0}^{\infty} \eta^{(n''-1)/2} \cdot 2(1/4a)^{\nu^*/2} K_{\nu^*}(\sqrt{a})\, d\eta \qquad (B.16)$$

where $\nu^* = (\nu + 2) / 2$

$$a = \frac{n^2 + n''}{4n''} [(\ell n\, \eta - \ell n\, y)^2 + n''(\ell n\, \eta - m'')^2 + \nu'' s^2]$$

$K_{\nu^*}(\sqrt{a})$ is a modified K-Bessel function. The density function $\tilde{f}(q)$ must be evaluated by numerical procedures. Appropriate numerical procedures are available but their investigation is outside the scope of this thesis. Equation (B.16) was integrated by the trapezoidal rule which will introduce errors that are progressive - especially where the density functions are steep.

Figure (B-2) shows Bayesian PDF for q. The parameters were the same as those for the posterior pdf for the mean η. Also given in Figure (B-2) is the density function from the Log-Student pdf with the same parameters for comparison. The two curves are almost identical and the difference is attributed to numerical error in the evaluation of Equation (B.16).

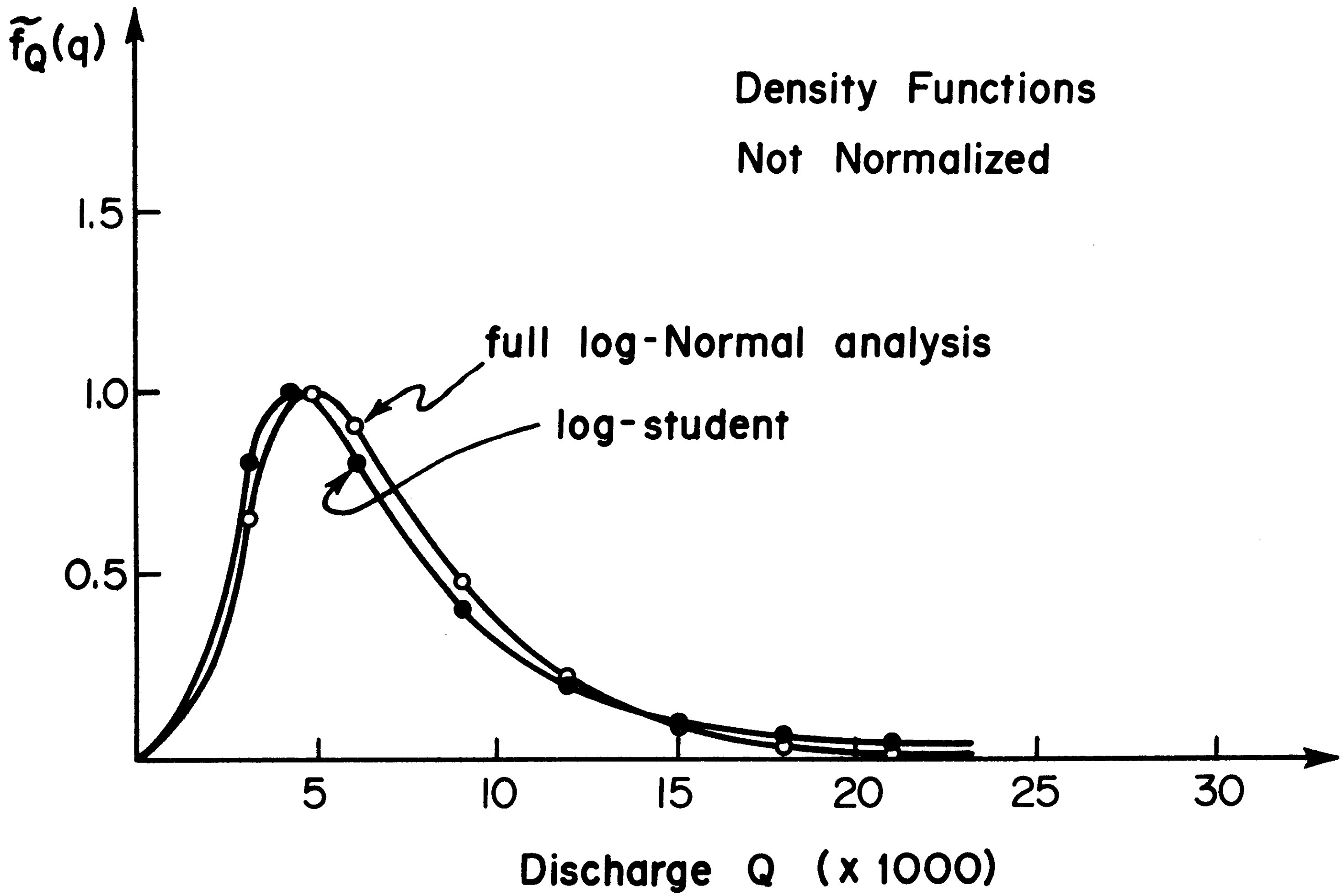


Figure B.2: Comparison of Bayesian Distributions from the Log-Normal Analysis

Appendix C

Bayesian Distribution of Floods from the Exceedance Model.

The modelled CDF of exceedance floods is given by Equation (4.29) and is

$$\tilde{F}_Z(z|\alpha,\nu) \cong 1 - \nu t \exp[-\alpha z] \qquad \text{(C-1)}$$

under the assumption that z is far, out on the tail. Thus the pdf is

$$f_Z(z|\alpha,\nu) \cong \nu t \alpha \exp[-\alpha z] \qquad \text{(C-2)}$$

The posterior pdf for ν and α are both gamma-1 and are given in Equations (4.32) and (4.34) respectively.

The Bayesian pdf can be found by Equation (4.2) which is:

$$F_Z(z) = \int_\nu \int_\alpha f_Z(z|\alpha,\nu)\, f_{\gamma 1}''(\nu|u'',s'') \cdot f_{\gamma 1}''(\alpha|v'',\ell'') \cdot d\alpha \cdot d\nu \qquad \text{(C-3)}$$

substituting into (B-3), (B-2) and rearranging

$$f_Z(z) = t \int_\nu \int_\alpha \nu f_{\gamma 1}''(\nu|u'',s'') \cdot \alpha \exp[-\alpha z] \cdot f_{\gamma 1}''(\alpha|v'',\ell'')\, d\alpha\, d\nu \qquad \text{(C-4)}$$

the integration over ν gives $\bar{\nu}$, the expected value of $f_{\gamma 1}''(\nu|u'',s'')$, by definition.

The integration over α can be accomplished by changing parameters, that is

$$\int_{\alpha} \alpha \exp[-\alpha z] \cdot \frac{\ell^{v''+1}}{\Gamma(v''+1)} \alpha^{v''} \exp[-\ell'' \cdot \alpha] \, d\alpha$$

$$\frac{\ell^{v''+1}}{\Gamma(v''+1)} \cdot \int_{\alpha} \alpha^{v''+1} \cdot \exp[-\alpha(\ell+z)] \cdot d\alpha = \frac{\Gamma(v''+2)}{\Gamma(v''+1)} (\ell+z)^{-(v''+2)} \cdot \ell^{v''+1} \tag{C-5}$$

which can be simplified to

$$\bar{\alpha} \left(1 + \frac{\bar{\alpha} z}{v''+1}\right)^{-(v''+2)} \tag{C-6}$$

where $\bar{\alpha} = E[\alpha] = \frac{v''+1}{\ell}$

Thus the integration of (B-4) gives

$$\tilde{f}_Z(z) = \bar{\nu} \, \bar{\alpha} t \left[1 + \frac{\bar{\alpha} z}{v''+1}\right]^{-(v''+2)} \tag{C-7}$$

The Bayesian CDF $\tilde{F}_Z(z)$ can also be found by

$$\tilde{F}_Z(z) = \int_{\nu}\int_{\alpha} F_Z(z|\nu,\alpha) \, f_{\gamma 1}''(\nu|u'',s'') \cdot f_{\gamma 1}'' (\alpha|v'',\ell'') \cdot d\alpha \cdot d\nu$$

$$= 1 - \bar{\nu} t \left[1 + \frac{\bar{\alpha} z}{v''+1}\right]^{-(v''+1)} \tag{C-8}$$

This result can be easily found by using similar procedures as were used for $\tilde{f}_Z(z)$.

Appendix D

Prior Estimation of ν, the Arrival Rate for the Poisson Exceedance Model, using a Distribution from Plotting Position Theory.

This Appendix sets forth one possible statistical procedure of estimating a distribution of the arrival rate of floods greater than some level Q_b. The problem of estimating ν is related to the problem of estimating the return period of the base flood Q_b. Q_b may correspond to some level, say bank level in the river, so that floods greater Q_b are those of interest. Q_b must be large enough so that the flood events are independent and so that the probability distribution of events greater than Q_b can be approximated by an exponential.

The procedure herein is to choose a discharge Q_b that fulfills the requirements of the exceedance model. Often the level can be related to a T year flood. Subjective assessment of the distribution of the T year event, that Q_b relates to, can be done using procedures set out in Chapter 3. Instead, here the engineer will answer the question, "if we had a record n years long, a flood equal to or greater than Q_b would be observed on the average m times." Thus he would estimate the rank of Q_b in a record n years in length. Assuming that such an assessment can be made what is the distribution on ν, the arrival rate.

Let (1-P) be the probability that an event (flood) has a value less than $Q_b - \frac{dQ_b}{2}$. There are (n-m) of these events.

Let dp be the probability that an event lies in the interval $Q_b \pm \frac{dQ_b}{2}$. There is one such event.

Let p − dp be the probability that an event has a value greater than $Q_b + \frac{dQ_b}{2}$. There will be (m−1) events.

The probability of that the m^{th} value lies in the discharge band $Q_b \pm \frac{dQ_b}{2}$ will be $d\Theta$; which is

$$d\Theta \propto p^{m-1}(1-p)^{n-m}\,dp \qquad \text{(D-1)}$$

assuming that $(p-dp) \cong p$, let

$$f(p) \propto p^{m-1}(1-p)^{n-m} \qquad \text{(D-2)}$$

which will be the probability density function associated with the probability that the m^{th} rank in a record n is equal to or greater than $Q_b - \frac{dQ_b}{2}$.

The probability that the cumulative density function of the n^{th} event is less than p_o is

$$F_p(p_o) = \int_o^{p_o} f(p)\,dp = \binom{n}{m} \cdot m \cdot \int_o^{p_o} (1-p)^{n-m}\,p^{m-1}dp \qquad \text{(D-3)}$$

The integral is related to the incomplete Beta function which is defined as

$$B_I(r,s,p) \equiv \int_o^{p} t^r\,(1-t)^{s-r-1}\,dt \qquad \text{(D-4)}$$

So that (D-3) is

$$F_p(p_o) = \binom{n}{m} \cdot m \cdot B_I(m-1,\, n,\, p_o) \qquad \text{(D-5)}$$

p was defined as the probability that an event of magnitude greater than $Q_b - \frac{dQ_b}{2}$ (or $\cong Q_b$) had a distribution given by (D-2). The expected value of p is easily shown to be:

$$E[p] = \frac{m}{n+1} \tag{D-6}$$

ν is the arrival rate of events greater than Q_b. This is the probability p.

Thus the pdf of ν is

$$f(\nu) = \binom{n}{m} m \; \nu^{m-1}(1-\nu)^{n-m} \tag{D-7}$$

$$\text{with } E[\nu] = \frac{m}{n+1} \tag{D-8}$$

If an engineer can subjectively choose m and n, in the manner explained earlier, then a distribution on ν can be found by applying some theory from the distributions of plotting positions. Both the pdf and the CDF are easily found. Some distributions of ν are shown in Figures D-1, D-2 for various m and n for $E[\nu] = .1$ and $.067$. The corresponding $E[T]$ year floods are 10 years and 15 years respectively.

The pdf on ν given by (D-7) can be used directly as a prior. Since its form is not a natural conjugate, the posterior pdf on ν must be found through numerical procedures.

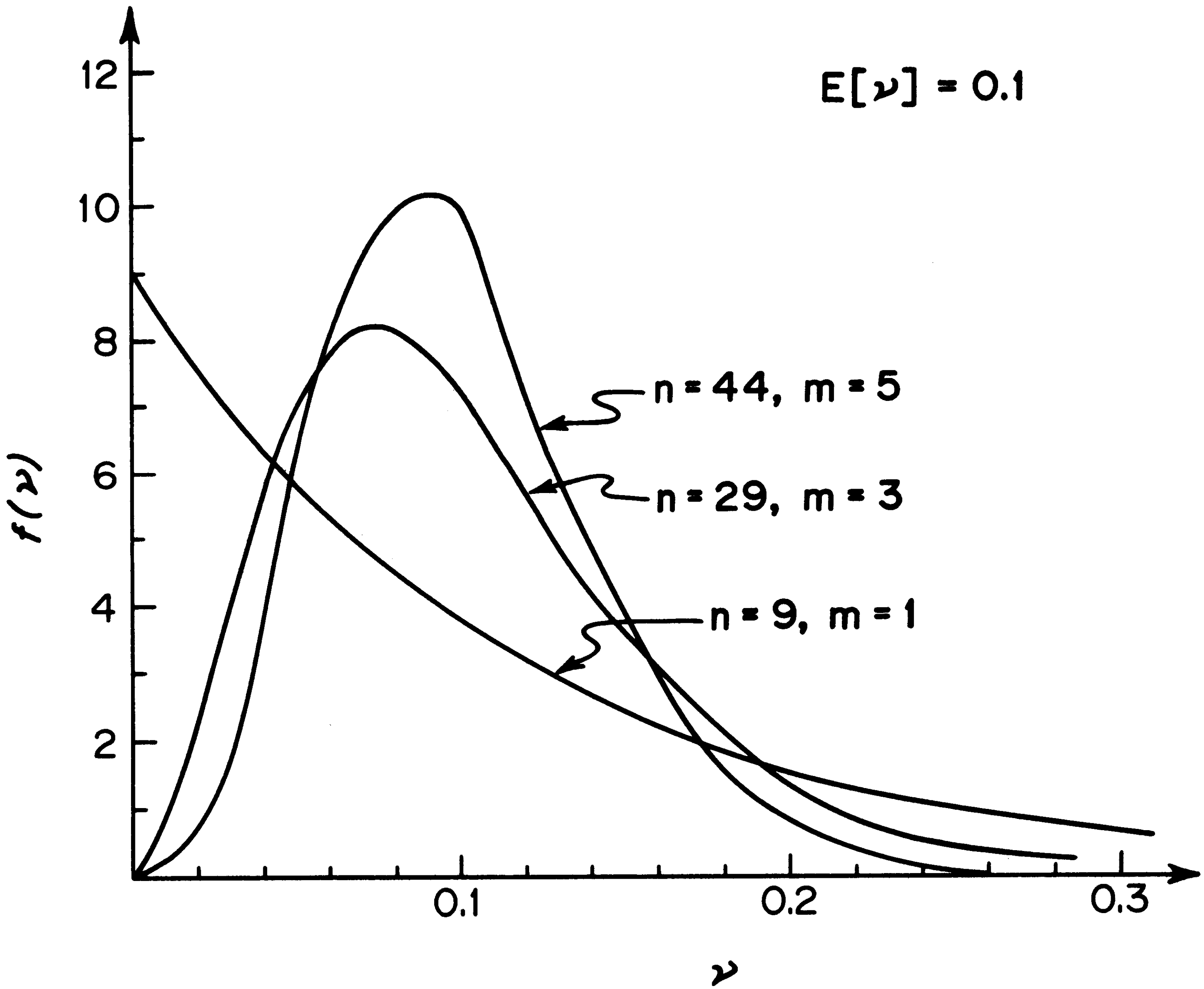


Figure D.1 Distribution for ν, for E[ν] = .1

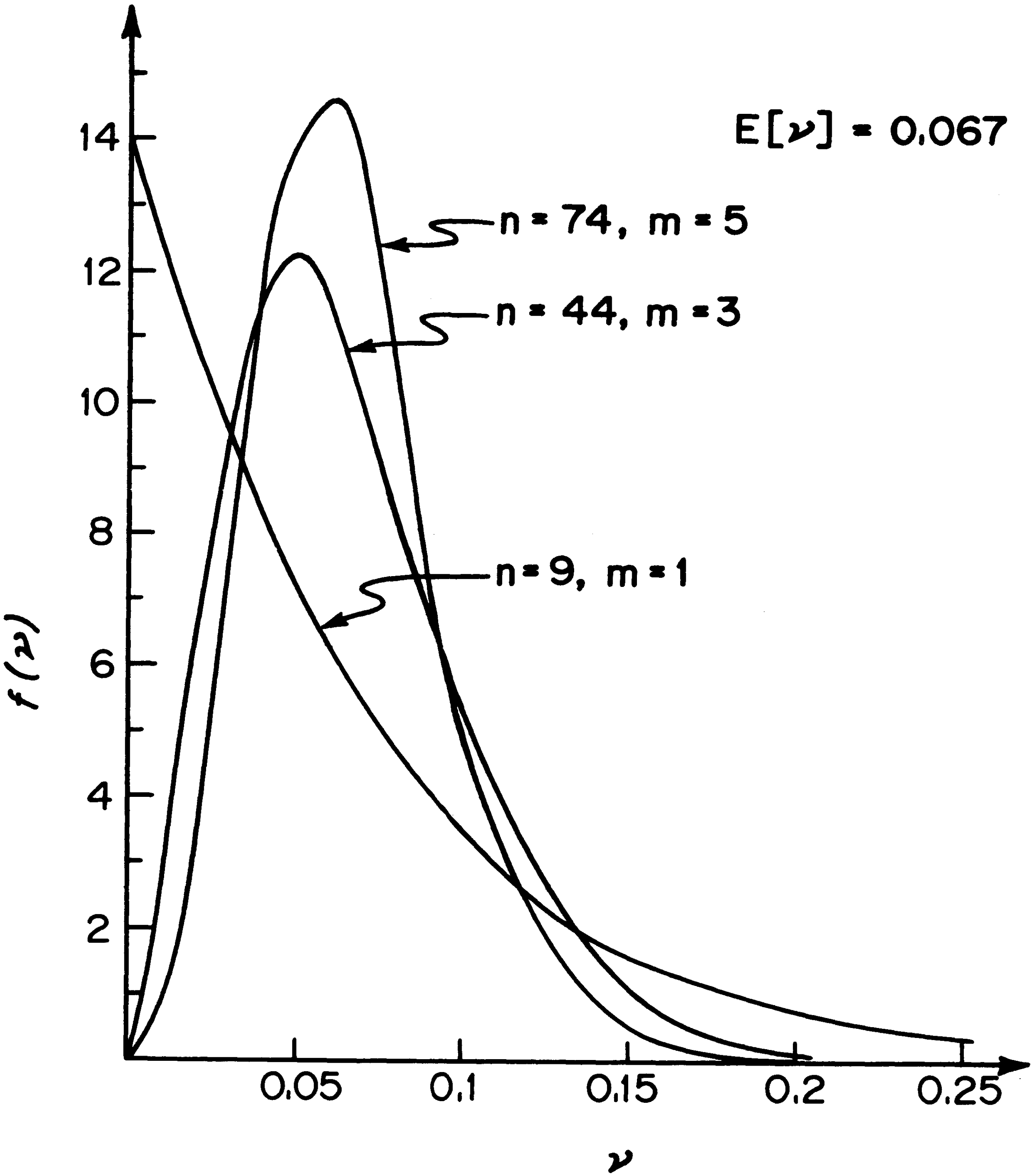


Figure D.2 Distributions for ν for E[ν] = .067

APPENDIX E

This appendix presents the derivation $G(q_p)$, the exceedance probability of the peak discharge of a natural catchment as derived by Eagleson (1972).

<u>Stochastic Model of the Rainfall</u>

Eagleson, (1972), modeled the rainfall during an individual storm event in terms of two random variables; the average rainfall intensity, $\bar{i}_o$ [in/hr], and the storm duration t_r[hr]. The first order joint density function (conditional on the occurrence of a storm event) is

$$f(\bar{i}_o, t_r) = \beta\lambda \exp[-\beta\bar{i}_o - t_r] \tag{E-1}$$

where β and λ are parameters

For Boston, Ma., β = 30 hr/in and λ = 0.13 hour^{-1} (Eagleson, 1972). The average rainfall excess intensity, $\bar{i}_e$ [in/hr] is assumed to be

$$\bar{i}_e = \bar{i}_o - \emptyset$$

where $\emptyset$ is a constant rate of water loss [in/hr].

It can be shown, (conditional on the occurrence of a rainfall excess event) that the joint pdf of $\bar{i}_e$ and t_{re} is

$$f(\bar{i}_e, t_r) = \beta\lambda \exp[-\beta\bar{i}_e - \lambda t_{re}]$$

where t_{re} is the duration [hr] of the rainfall excess intensity.

The total number of storms during a year is Θ but only $N \leq \Theta$ of these produce rainfall excess. Each storm is assumed to have a rectangular storm interior.

Mechanics of Overland Flow

Eagleson models catchments as an idealized flow plane. Kinematic wave theory is applied to determine the peak discharge with the added assumption that the time of concentration for the streams, t_s, is treater than the time of concentration for the overland flow catchment, t_c. The peak discharge, q_m, is found to be

$$q_m = \alpha_s [A_s (L_s, t'_*)]^{3/2} \qquad \text{(E-4)}$$

where

α_s is a parameter of the streamflow (assumed by Eagleson to be .1 sec^{-1})

A_s is the cross-sectional area of the streamflow (ft^2) which is derived from

$$A_s(L_s, t'_*) = 16.7\ \alpha_c\ i_e^{\,2}\ t_c^{\,3}(\frac{3tr_e}{t_c} - 2) + \frac{7720R_c^2}{\alpha_c t_c} [\frac{t'_* - tr_e}{t'_* - tr_e + (t_c/2)}] \qquad \text{(E-5)}$$

for $t_c < t_{re} < t_s + t_c$

and where

α_c = catchment parameter (assumed to be equal to 10 sec^{-1}

by Eagleson)

R_c = dimension perpendicular to the stream of the area producing direct runoff (miles)

t'_* = a time found from

$$\frac{(t'_* - t_c)}{[t'_* - (t_c/2)]} \simeq 1 \qquad \text{(E-6)}$$

for the case where $t_{re} > t_s + t_c$, q_m is found to be

$$q_m = 645\ A_r\ i_e \qquad \text{(E-7)}$$

where

A_r is the area producing direct runoff (sq. mi.) assumed to be 1/2 the catchment area.

i_e is the excess rainfall intensity (in/hr).

(E-7) is a particular result of the kinematic wave analysis of rectangular storm events that converges to well recognized rational formula.

$F(q_m)$ for Individual Storms

The flood cumulative probability density for individual events can be found from

$$F(q_m) = \int_{R_{q_m}} f(\bar{i}_e, t_r)\ d\bar{i}_e\ dt_r \qquad \text{(E-8)}$$

where the region of integration is defined by R_{q_m} and has boundaries $\bar{i}_e = 0$, $t_r = 0$, q_m = constant. This region is shown in Figure E-1.

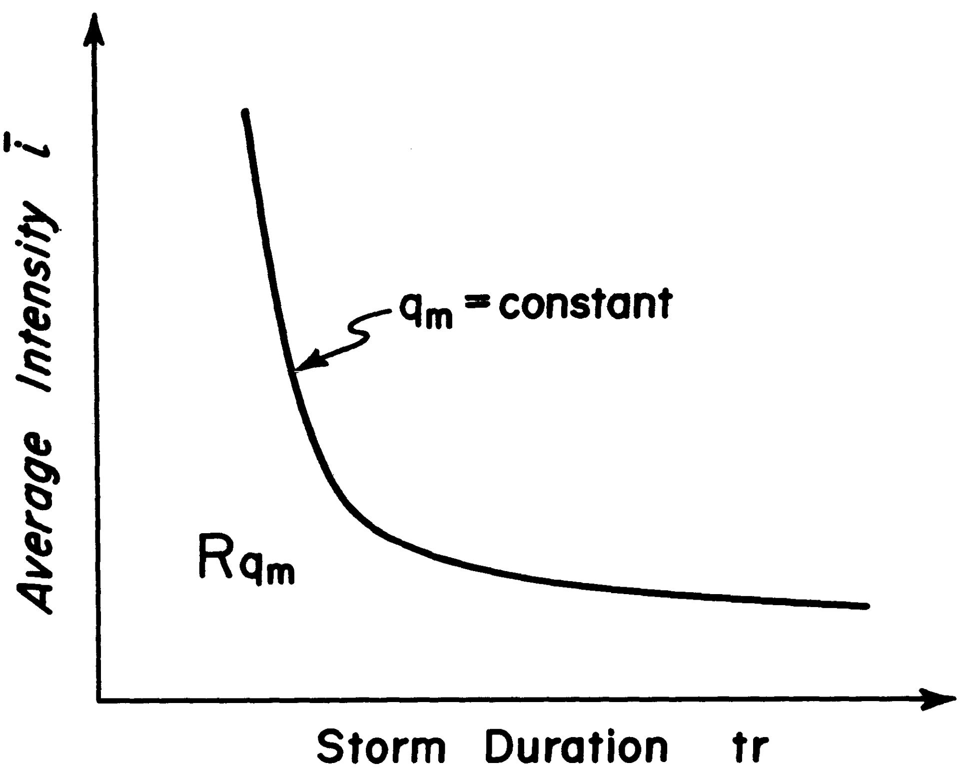


Figure E.1: $\bar{i}$, t_r Plane Showing Peak Discharge

The boundary q_m = constant is defined by the mecahnics of overland flow and represented by Equations (E-4) and (E-7). All combinations of $\bar{i}_e$ and t_r to the southwest of the boundary q_m = constant produce discharges less than or equal to q_m.

The analytical structure of this region precludes an exact analytical integration of (E-8). Eagleson approximates this region by substituting for the boundary q_m = constant a boundary defined as

$$t_r = B/\bar{i}_e^m \tag{E-9}$$

where

$$B = 2.97\left(\frac{A_r}{\alpha_c L_s}\right)^{1/2}\left[1 - \frac{655\ \alpha_s^{4/3}\ A_r^2}{\alpha_c\ L_s^3\ q_m^{1/3}}\right]$$

and for m = 1/2

Under this approximation (E-8) can be integrated to yield

$$F(q_m) = 1 - e^{-2\sigma}\sigma^{-\sigma+1}\Gamma(\sigma)\ \exp\left[\frac{-\beta q_m}{645KA_r}\right] \tag{E-10}$$

where

$$\sigma = \left[2.21\ \frac{\beta\lambda^2 A_r}{K\alpha_c L_s}\ \left(1 - \frac{655\alpha_s^{4/3}A_r^2}{\alpha_c L_s^3 q_m^{1/3}}\right)\right]^{1/3}$$

K = factor reducing point rainfall depth to average areal depth over area A_r

The total flood discharge peak, q_p, will include the peak due to direct runoff, q_m, and the average contribution from groundwater, $\bar{q}_b$, which Eagleson estimates to be

$$\bar{q}_b = .074\ (1 - \Phi_2)\ \Phi_1\ PA_c \qquad \text{(E-11)}$$

where

Φ_1 = ratio of average annual runoff to average annual point rainfall. (Estimated to be about .50 for New England)

Φ_2 = ratio of average annual direct runoff to average annual runoff

P = average annual point rainfall (inches)

Substituting $q_p - \bar{q}_b$ for q_m in Equation (E-10) and taking the compliment of $F(q_m)$ which is the flood exceedance probability, $G(q_m)$, results in Eagleson's derived flood frequency formula which is

$$G(q_p) = e^{-2\sigma}\sigma^{-\sigma+1}\Gamma(\sigma)\ \exp\left[\frac{-\beta(q_p - \bar{q}_b)}{K645A_r}\right] \qquad \text{(E-12)}$$

APPENDIX F

Analytic Derivation of the Marginal Likelihood Function

This appendix presents the analytical derivation for the marginal likelihood or marginal density function of a probability model.

The marginal likelihood, K, is found from

$$K = \int_{\underline{\Theta}} L(\underline{\Theta}|\underline{Q}) \cdot f'(\underline{\Theta}) d\underline{\Theta} \tag{F.1}$$

where

$\underline{\Theta}$ is the set of model parameters

$L(\underline{\Theta}\ \underline{Q})$ is the likelihood function for a set of observations $\underline{Q}$

$f'(\underline{\Theta})$ is the prior pdf for the parameters $\underline{\Theta}$.

This appendix derives the marginal likelihood for the normal, log-normal and exceedance models.

F.1 Marginal Likelihood for the Normal Model

Let q be distributed from a normal pdf. Then, given n independent observations of q, $\underline{Q}$, the likelihood function for μ and h is:

$$L(\mu,h|\underline{Q}) = \prod_{i=1}^{n} f_N(q_i|\mu,h)$$

$$= (2\pi)^{-n/2} n^{n/2} \exp[-\frac{h}{2} \sum (q_i-\mu)^2] \tag{F.2}$$

Define the following

$$m = \frac{1}{n} \sum q_i \tag{F.3}$$

$$v = \frac{1}{n-1} \sum (q_i - m)^2 \tag{F.4}$$

$$\nu = n-1$$

then

$$L(\mu,h|\underline{Q}) = (2\pi)^{-n/2} \exp[-\frac{1}{2}\, h\, \nu v - \frac{1}{2}\, hn(m-\mu)^2] \cdot h^{n/2} \tag{F.5}$$

Assume the prior on (μ,h) is a natural conjugate prior of the form

$$f'(\mu,h) = (2\pi)^{-1/2} \exp[-\frac{1}{2}\, hn'(\mu-m')^2 - \frac{1}{2}\, h\nu'v'] \cdot$$

$$\cdot\; n'^{1/2} h^{1/2} \;.\; h^{1/2\nu'-1} \frac{(1/2\ \nu'v')^{1/2}}{\Gamma(1/2\ \nu')}' \tag{F.6}$$

define

$$m'' = \frac{n'm' + nm}{n'+n} \tag{F.7}$$

$$n'' = n' + n \tag{F.8}$$

$$v'' = \frac{1}{\nu''}(\nu'v' + nm'^2 + \nu v + nm^2 - n''m''^2) \tag{F.9}$$

$$\nu'' = \nu' + \nu + 1 = n'' - 1 \tag{F.10}$$

The product of F.5 and F.6 yields as an exponent on exp as

$$-\frac{1}{2}h[n'(\mu-m')^2 + \nu'v' + \nu v + n\ (m-\mu)^2] \tag{F.11}$$

which can be expanded to

$$- \frac{1}{2} h[n'\mu^2 - 2n'm'\mu + n'm'^2 + \nu'v' + \nu v + nm^2 - 2nm\mu + n\mu^2$$

$$+ n''m''^2 - n''m''^2] \qquad (F.12)$$

$\nu''v''$ can be isolated easily

$$-2\mu n''m'' = - 2\mu(n'm' + nm), \text{ and}$$

$$n''\mu^2 = (n' + n')\mu^2$$

Therefore (F.12) can be rewritten as

$$- \frac{1}{2} h[n''(\mu-m'')^2 + \nu''v''] \qquad (F.12)$$

and (F.1) is, from F.12 and the constants in (F.5) and (F.6),

$$K_N = n'^{1/2}(2\pi)^{-\nu/2} \frac{(1/2\ \nu'v')^{1/2\ \nu'}}{\Gamma(1/2\ \nu')} \int_{\mu}\int_{h} (2\pi)^{-1/2} h^{1/2} \exp[- \tfrac{1}{2} hn'' \cdot$$

$$\cdot(\mu-m'')^2] \cdot h^{1/2\nu''-1} \exp[- \tfrac{1}{2} h\nu''v''] d\mu dh$$

(F.13)

The integral is equal to

$$\frac{1}{n''^{1/2}} \cdot \frac{\Gamma(1/2\ \nu'')}{(1/2\ \nu''v'')^{1/2\nu''}} \qquad (F.14)$$

Thus

$$K_N = (\frac{n'}{n''})^{1/2} \cdot (2\pi)^{-\nu/2} \cdot \frac{\Gamma(1/2\ \nu'')}{\Gamma(1/2\ \nu')} \cdot \frac{(1/2\ \nu'v')^{1/2\nu'}}{(1/2\ \nu'v')^{1/2\nu''}} \qquad (f.15)$$

F.2 Marginal Likelihood for the Log-Normal Model

Let $x_i = \ell n\ q_i$ be normally distributed with mean μ and precision h. Then q_i is log normally distributed by definition. The probability density function for q is

$$f(q|\mu,h) = \frac{1}{q}(2\pi)^{-1/2}h^{1/2}\exp[-\frac{1}{2}h(x-\mu)^2] \qquad \text{(F.16)}$$

The likelihood function for μ and h, given n independent observations of q is

$$L(\mu,h|\underset{\sim}{Q}) = \frac{1}{\prod_{i=1}^{n} q_i}(2\pi)^{-n/2}\cdot h^{n/2}\cdot\exp[-\frac{1}{2}hn\sum(x_i-\mu)^2] \qquad \text{(F.17)}$$

Assume a normal gamma prior for μ and h of the same form as Equation (F.6). The marginal likelihood, K_{LN}, is just the integration of μ and h over the product of the likelihood and the prior pdf.

$$K_{LN} = \frac{1}{\prod_{i=1}^{n} q_i}\int_{\mu,h}(2\pi)^{-n/2}h^{n/2}\exp[-\frac{1}{2}hn\sum(x_i-\mu)^2]\cdot f'(\mu,h)d\mu dh \qquad \text{(F.18)}$$

The integral is of the same form as the marginal likelihood for the normal model. Then, from Equation (F.15), K_{LN} is:

$$K_{LN} = \frac{1}{\prod_{i=1}^{n} q_i}\cdot\left(\frac{n'}{n''}\right)^{1/2}(2\pi)^{-\nu/2}\frac{\Gamma(1/2\ \nu'')}{\Gamma(1/2\ \nu')}\cdot\frac{(1/2\ \nu'v')^{1/2\ \nu'}}{(1/2\ \nu''v'')^{1/2\ \nu''}} \qquad \text{(F.19)}$$

F.3 Marginal Likelihood for the Exceedance Model

Assume that the probability of flood discharges, q, are distributed such that discharges greater than or equal to some base flow, q_b, are exponentially distributed and that the arrivals of such events are Poisson occurrences. As explained in Chapter 7, Section 7.3, the pdf for discharges less than q_b will follow a uniform pdf. Then the pdf for flood discharges, q, can be written as

$$f(q|\nu,\alpha) = \nu\alpha t \exp[-\alpha(q-q_b)] \quad \text{if} \quad q \geq q_b$$

$$= \frac{1-\nu}{q_b} \quad \text{if} \quad 0 \leq q < q_b \qquad (F.20)$$

Given a sample of n independent discharges, $\underline{Q}$, of which m are discharges less than q_b and n-m are discharges greater than or equal to q_b, then the likelihood function for ν and α is, from Appendix C,

$$L(\nu,\alpha|\underline{Q}) = \frac{(1-\nu)^m}{{q_b}^m} \cdot \nu^{n-m} \alpha^{n-m} \exp[-\alpha \sum_{i=1}^{n-m} (q_i-q_b)]$$

$$\cdot \exp[-\nu\sum t_i] \qquad (F.21)$$

The marginal likelihood function, K_E, is defined as:

$$K_E \int_{\nu,\alpha} L(\nu,\alpha|\underline{Q}) \cdot f'(\nu) \cdot f'(\alpha) d\nu d\alpha \qquad (F.22)$$

The prior pdf for ν and α are of the form

$$f'(\nu) = \exp(-s'\cdot\nu)\cdot\nu^{u'-1} \frac{s'^{u'}}{\Gamma(u')} \tag{F.23}$$

$$f'(\alpha) = \exp(-\ell'-\alpha)\cdot\alpha^{v'-1} \frac{\ell'^{v'}}{\Gamma(v')} \tag{F.24}$$

$$K_E = q_b^{-m} \cdot \frac{s'^{u'}}{\Gamma(u')} \int_\nu \exp[-(s'' + m)\nu]\cdot\nu^{(u'+n-m)-1} \, d\nu \cdot$$

$$\cdot \frac{\ell'^{v'}}{\Gamma(v')} \int_\alpha \exp[-(\ell' + \sum(q_i - q_b)]\cdot\alpha^{(v'+n-m)-1} \, d\alpha \tag{F.25}$$

The integral over ν equals

$$\frac{\Gamma(u'')}{(s''+m)^{u''}} \tag{F.26}$$

where

$$u'' = u' + n - m$$

$$s'' = s' + T \quad (\text{or } s'' = s' + \sum t_i)$$

and the integral over α equals

$$\frac{\Gamma(v'')}{(\ell'')^{v''}}$$

where

$$v'' = v' + n - m$$

$$\ell'' = \ell' + \sum(q_i - q_b)$$

Thus, K_E equals

$$K_E = q_b^{-m} \cdot \frac{s'^{u'}}{(s''+m(^{u''}} \cdot \frac{\Gamma(u'')}{\Gamma(u')} \cdot \frac{\ell'^{v'}}{\ell''^{v''}} \qquad \frac{\Gamma(v'')}{\Gamma(v')} \qquad \text{(F.28)}$$

BIOGRAPHY

Eric Franklin Wood was born on 22 October 1947 in Vancouver, Canada, and received all his primary and secondary education in that city.

He attended the University of British Columbia from September 1965 to May 1970, receiving a Bachelor of Applied Sciences (Honours) in Civil Engineering. He continued his education at M.I.T. starting September 1970. He received the S.M. Degree in Civil Engineering in June 1972 and the C.E. Degree in June 1973.

He is a member of the American Geophysical Union, Chi Epsilon and Sigma Xi honorary fraternaties. His publications include:

Wood, E.F., B. Harley, and F. Perkins, "Operational Characteristics of a Numerical Solution for the Simulation of Open Channel Flow", Ralph M. Parsons Laboratory, Technical Report #150, Department of Civil Engineering, M.I.T., Cambridge, Mass, June 1972

Harley, B.M., E. Wood, and J.C. Schaake, "The Application of Hydrologic Models to Urban Planning", Presented at the 54th Annual Meeting of the American Geophysical Union, April 16-20, 1973, Washington, D.C.

Wood, E.F., "Flood Control Design with Limited Data - A Comparison of the Classical and Bayesian Approaches", Presented at the Symposium on the Design of Water Resources Projects with Inadequate Data, Madrid, June 1973

Schaake, J.C., G. Vicens, E. Wood, R. Lenton and A. Rhenals, "Discussion of Drought Severity and Water Supply Dependability by L. Beard and H. Kubik", J. of the Irrigation and Drainage Division, American Society of Civil Engineers, Vol. 99, No. 1R3, September 1973

Wood, E.F., B. Harley and F. Perkins, "Transient Flood Routing in Channel Networks", forthcoming