Statistical Models and Mental Health:
An Analysis of Records
From a Mental Health Center

by
Edward Harris Kaplan
B.A., McGill University (1977)
S.M., Massachusetts Institute of Technology (1979)
M.C.P., Massachusetts Institute of Technology (1979)

SUBMITTED TO THE DEPARTMENT OF MATHEMATICS
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS OF THE DEGREE OF
MASTER OF SCIENCE IN MATHEMATICS
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 1982

The author hereby grants to M.I.T. permission to reproduce and to distribute copies of this thesis document in whole or in part.

Signature of Author  
Department of Mathematics  
May 7, 1982

Certified by  
Herman Chernoff  
Faculty Advisor

Accepted by  
Chairman, Departmental Graduate Committee

JUL 30 1982

LIBRARIES
STATISTICAL MODELS AND MENTAL HEALTH:  
AN ANALYSIS OF RECORDS 
FROM A MENTAL HEALTH CENTER  

by  
EDWARD HARRIS KAPLAN  

Submitted to the Department of Mathematics on May 7, 1982  
in partial fulfillment of the requirements for the Degree of Master of Science in Mathematics  

ABSTRACT  
This thesis examines lengths of stay in a New York State psychiatric hospital. Following a rationale for the study, the data are described and exploratory models for length of stay are built. The observed lengths of stay are shown to fit a beta-mixture of geometric distributions. Estimation techniques using weighted least squares and maximum likelihood are developed for this model; the statistical properties of the maximum likelihood estimates are discussed in some detail. The beta-mixture model is applied to the hospital data, and it is found that psychiatric diagnosis and legal status strongly influence observed lengths of stay. Topics for future research are discussed at the end of the thesis.
ACKNOWLEDGEMENTS

I want to thank Ron Wooldridge, Director of Forecasting and Modeling at the New York State Office of Mental Health for providing such interesting data. I also want to thank my thesis supervisor, Larry Gillick, for his enthusiastic support and insightful observations throughout the course of this project. In addition, I want to thank Rosalinda Bishop for cheerfully producing the typewritten manuscript from my handwritten scrawl. Finally, I want to thank Martha for understanding and accepting my "disappearances" over the past several months.
I. INTRODUCTION

This thesis represents a statistical study of some mental health data. In particular, we are interested in the length of time spent in the hospital by patients admitted to a New York State psychiatric center. Our basic question is, can we predict a patient's length of stay in the hospital knowing patient characteristics such as psychiatric diagnosis, legal status (voluntary or involuntary admission), age and sex?

There are basically two reasons why one would be interested in studying length of stay. The first of these relates to planning. Given a cohort of admitted patients, hospital administrators need to determine the number of patient days required for treatment. Presumably, the staff workload incurred due to any particular patient is proportional to that patient's duration of stay in the hospital. Thus, a model predicting length of stay could be used to determine staffing levels inside the hospital.

The second reason for studying length of stay concerns the relationship between treatment and psychiatric diagnosis. Most psychiatrists would claim that the decision to hospitalize a particular psychiatric patient depends primarily upon the severity of observed symptoms of mental illness. These symptoms translate into clinical diagnoses such as schizophrenia, manic depression and the like. However, studies have shown that actual hospitalization decisions are less a function of diagnosis, and more dependent upon other factors such as who performs the diagnosis and previous history of institutionalization (for an example, see Mendel and Rapport, 1973).
Until recently, the relevance of the psychiatric classifications used in practice was seriously questioned by psychiatrists, social workers and other mental health practitioners. The American Psychiatric Association refined its diagnostic classifications in 1981 with the publication of the 3rd edition of their Diagnostic and Statistics Manual (DSM-III). Unfortunately, most patients remain classified under the disputed DSM-II diagnoses; the new system has yet to be fully implemented.

For the hospital we will study, diagnoses were assigned using the DSM-II classification scheme. Thus, we have an opportunity to see whether or not treatment, as measured by length of stay, relates to these different diagnoses. If it is true that the DSM-II diagnoses are not accurate reflections of one's psychiatric condition, and if it is true that treatment does reflect upon one's mental problems, then one shouldn't expect to find a strong relationship between treatment and diagnosis.

Suppose, however, that length of stay is related to clinical diagnosis. This could indicate that both treatment and diagnosis reflect a patient's true psychiatric illness. It could also indicate that regardless of a patient's true mental state, treatment will follow a certain course given a certain diagnosis. It may be that a labeling syndrome occurs - once a patient is framed in a certain manner (via diagnosis), it may be difficult to convince people to view the patient as anything other than his or her label.

Having provided a rationale for studying length of stay, the question of which lengths of stay to study arises. One could study lengths of stay for all people admitted within a certain time frame. Or, we could look at lengths of stay for all of the patients in the hospital on a certain
date. Clearly, these two distributions will be different. The population of a mental health center is largely a residue population. At any given point in time, this population will consist of patients with unusually long lengths of stay compared to the bulk of admitted patients. A good discussion of this is found in Kramer (1957).

For our purposes, we are interested in the length of stay facing an individual upon admission to the hospital, thus we will not consider "point-in-time" length of stay distributions. Note that it is possible for the same individual to be admitted more than once over a given time interval. Unfortunately, our data on diagnoses and other variables only refer to a patient's first visit to the hospital during the time frame considered. Therefore, our study will only study lengths of stay for a patient's first visit to the hospital during a fixed time interval.

We stated at the outset that this thesis is concerned with a statistical study of length of stay. During the course of the initial data analysis, it was decided that the usual regression methods were not well suited to the problem at hand. Thus, a parallel goal of this thesis is to develop a model for length of stay from "first principles" that takes into account some of the realities of a mental health center. This model is then used to analyze our data set.

The four remaining chapters report the analysis of lengths of stay for patients admitted to a New York State psychiatric center. In the next chapter, the data are described and analyzed in an exploratory fashion using regression and proportional hazards models. Chapter III is concerned with the axiomatic development of a model for length of stay, and its attendant statistical properties. We apply this model to the
observed data in Chapter IV. In Chapter V, we discuss a number of issues that have yet to be resolved. These issues are mainly technical in nature, but they do harbor implications for future research.
II. DATA DESCRIPTION AND EXPLORATORY ANALYSIS

Our original data consisted of variables pertaining to 24866 admissions to the Central Islip Psychiatric Center. These data refer to all those in the hospital as of September 1, 1976, and all those admitted from this date through May 31, 1981. For any patient released on or before May 31, 1981, true lengths of stay are known; this includes patients already in the Center as of September 1, 1976. Patients who were released after May 31, 1981 are referred to as "censored" patients; for these individuals, length of stay is recorded as time from admission until the termination date.

The population referred to above constitutes a complete census of all patients moving into and out of the Central Islip Center. As discussed in Chapter I, we are interested in modeling length of stay for a patient's first visit during the stay period. Also, we recall the distinction between cohort and point-in-time length of stay distributions made earlier. Since the data includes all patients who were in the hospital on a fixed starting date, we are guaranteed to find a large, residue population with abnormally long lengths of stay in our data. Thus, we will only consider those visits with admission dates between September 1, 1976 and May 31, 1981 which correspond to first admission to the hospital during the period specified. This reduces the number of visits considered from 24866 to 13771.

This is still a large number of cases to work with, so it was decided to construct a random sample consisting of roughly half of the data. The resulting sample contains 6965 observations, or about 51% of the data considered. These data were further subdivided at random
into five groups of 1384, 1428, 1356, 1400 and 1397 observations. In this chapter, we will perform several analyses on each of these five samples; later we will focus on the 6965 observations as a single sample.

Several variables were obtained in addition to length of stay in the hospital. These include:

1. Multiple Visits - 1 if patient had more than one visit during the study period; 0 otherwise.
2. First Admission - 1 if first visit was also patient's first admission to a New York State institution; 0 otherwise.
3. Age at Admission - categorized as 0-18 years, 19-25 years, 26-45 years, 46-65 years, and 66+ years.
4. Sex of patient.
5. Ethnicity - white; minority; and unknown.
7. Legal Status - voluntary admission, involuntary admission.
8. Psychiatric Diagnosis - schizophrenia, alcoholic psychosis, alcoholism, major affective disorders (manic depression), psychosis associated with cerebral conditions (e.g. brain trauma), neurosis, transient situational disorders (e.g. adjustment reaction of adolescence), mental retardation, non-psychotic organic brain syndromes (e.g. senility), all other diagnoses, unknown.

In addition, for each observation it was recorded whether or not the patient was in the hospital on the last day of the study period. To
provide a feeling for these data, the marginal distributions of the
covariates for the first of our five subsamples is presented in Table 2.1.

Observed lengths of stay vary from a minimum of 1 day to a maximum
of 1731 days for the first of our five samples. With such a large range,
it is useful to quantize length of stay - this will make our analyses
more manageable, especially when considering complicated probabilistic
models for length of stay such as those pursued in Chapters III and IV.
Thus, we will measure length of stay in quantized weeks: if \( L_i \) = quantized
length of stay for ith observation, and \( \text{Los}_i \) = length of stay in days for
the ith observation, we set

\[
L_i = \left\lfloor \frac{\text{Los}_i}{7} \right\rfloor
\]

where \( \lfloor x \rfloor \) means "the smallest integer greater than or equal to \( x \)."

Now the observed distribution of \( L \) has very heavy tails; this
distribution is shown in Figure 2.1 for the first sample of 1384 patients.
If we are interested in building models relating \( L \) to our covariates, a
transformation is in order. For exploratory analysis, we would like to
keep our models simple, so a logarithmic transformation will be used.

As a first pass, stepwise regressions were run relating \( \log L \) to
the covariates. To adjust for the fact that some lengths of stay were
censored, an indicator variable stating whether or not the patient was
in the hospital at the end of the study period was included. The idea
behind these models was not to obtain good parameter estimates. Rather,
the intent was to screen out those variables which bear no relation to
TABLE 2.1
Marginal Distributions of Covariates from First Sample

1) Multiple visits - 17.6% yes; 82.4% no
2) First admission - 53.0% yes; 47.0% no
3) Age at admission - 2.8% 0-18; 14.2% 19-25; 41.8% 26-45; 24.7% 46-65; 16.6% 66+
4) Sex of patient - 61.8% male; 38.2% female
5) Ethnicity - 79.3% white; 16.1% minority; 4.6% unknown
6) Marital Status - 44.7% never married; 20.7% married; 11.7% divorced or annulled; 10.7% separated; 7.0% widowed; 5.2% other or unknown
7) Legal status - 53.5% voluntary admissions; 44.7% involuntary admissions
8) Psychiatric diagnosis - 28.5% schizophrenia; 15.3% alcoholic psychosis; 13.1% alcoholism; 6.7% major affective disorders; 3.2% psychosis associated with cerebral condition; 2.6% neurosis; 2.5% transient situational disorders; 3.1% mental retardation; 3.2% non-psychotic organic brain syndromes; 4.0% all other diagnoses; 17.8% unknown
9) Censoring indicator - 4.5% yes; 95.5% no
FIGURE 2.1

Histogram for Quantized Length of Stay from First Sample

EACH * REPRESENTS 10 OBSERVATIONS

<table>
<thead>
<tr>
<th>MIDDLE OF INTERVAL</th>
<th>NUMBER OF OBSERVATIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>500</td>
</tr>
<tr>
<td>2.</td>
<td>206</td>
</tr>
<tr>
<td>3.</td>
<td>107</td>
</tr>
<tr>
<td>4.</td>
<td>94</td>
</tr>
<tr>
<td>5.</td>
<td>61</td>
</tr>
<tr>
<td>6.</td>
<td>43</td>
</tr>
<tr>
<td>7.</td>
<td>44</td>
</tr>
<tr>
<td>8.</td>
<td>34</td>
</tr>
<tr>
<td>9.</td>
<td>31</td>
</tr>
<tr>
<td>10.</td>
<td>39</td>
</tr>
<tr>
<td>11.</td>
<td>25</td>
</tr>
<tr>
<td>12.</td>
<td>30</td>
</tr>
<tr>
<td>13.</td>
<td>17</td>
</tr>
<tr>
<td>14.</td>
<td>13</td>
</tr>
<tr>
<td>15.</td>
<td>14</td>
</tr>
<tr>
<td>16.</td>
<td>6</td>
</tr>
<tr>
<td>17.</td>
<td>8</td>
</tr>
<tr>
<td>18.</td>
<td>13</td>
</tr>
<tr>
<td>19.</td>
<td>3</td>
</tr>
<tr>
<td>20.</td>
<td>5</td>
</tr>
<tr>
<td>21.</td>
<td>8</td>
</tr>
<tr>
<td>22.</td>
<td>6</td>
</tr>
<tr>
<td>23.</td>
<td>1</td>
</tr>
<tr>
<td>24.</td>
<td>2</td>
</tr>
<tr>
<td>25.</td>
<td>6</td>
</tr>
</tbody>
</table>
length of stay, so that more formal models could be constructed with fewer covariates.

Table 2.2 lists the variables found to be significant at the 5% level for each of three subsamples. As one would expect, the censoring indicator is significant in all three models. Two other variables appear significant in all three samples: legal status and the alcoholic psychosis indicator. The alcoholism indicator appears in two of the three models. Whether or not the patient has previously been in the New York State mental health system bears little to no relation to length of stay. It also seems that age has little to do with length of stay, though one would expect age and diagnosis to interact.

We have gained some preliminary evidence that length of stay is associated with legal status and certain psychiatric diagnoses, notably those involving alcohol. However, it is dangerous to conclude on the basis of these models that all of the other variables are unimportant - we have not handled censored observations properly, and our assumption that log L is linear in the covariates is very strong. Also, the models are not predictively strong, as evidenced by the low $R^2$. To verify (or refute) these exploratory findings, we now turn to a nonparametric method for relating $L$ to the covariates that properly accounts for observed censoring.

Suppose we assume that quantized length of stay follows the general distribution

$$
Pr \{ L_i = l \} = h(l; x_i) \frac{l-1}{n} \prod_{k=1}^{l} (1 - h(k; x_i)) \tag{2.2}
$$
TABLE 2.2

Significant Covariates from Stepwise Regressions on Log-Length of Stay

First Sample (R**2 = 0.242)

Censor indicator; legal status; alcoholic psychosis; mental retardation; age=26-45; sex

Second Sample (R**2 = 0.278)

Censor indicator; legal status; schizophrenia; alcoholic psychosis; alcoholism; major affective disorders; psychosis associated with cerebral condition; neurosis; non-psychotic organic brain syndromes; all other diagnoses; multiple visits; white; minority; married

Third Sample (R**2 = 0.261)

Censor indicator; legal status; alcoholic psychosis; alcoholism; separated; widowed
where \( x_i \) represents the covariate vector for the \( i \)th patient. The function \( h(l; x_i) \) is referred to as the hazard function - \( h(l; x_i) \) is the conditional probability of release from the hospital during the \( l \)th week of stay, given a length of stay of at least \( l \) weeks. Let us further assume that

\[
h(l; x_i) = h_0(l) e^{x_i \beta}
\]

(2.3)

The model (2.3) is called the proportional hazards model - all individual hazards are multiplicative factors of some unspecified baseline hazard \( h_0(l) \) with the multiplicative factor corresponding to \( e^{x_i \beta} \).

The parameter vector can be estimated from observed data using a method similar to maximum likelihood estimation; the interested reader is referred to Kalbfleisch and Prentice (1980, p.76-78). This method does take censoring into account when computing estimates of \( \theta \).

As our analysis in this chapter remains exploratory in intent, a stepwise version of the proportional hazards model was fit to the fourth of our five subsamples using the BMDP2L program (Dixon et.al., 1981, p.576-594). Significant variables, their estimated coefficients and standard errors are shown in Table 2.3. Note that positive values of \( \theta \) increase the hazard and shorten length of stay, while negative values of \( \theta \) do the reverse. It is interesting to note that legal status (in the form of involuntary admission) and the alcoholic psychosis indicator are again included in the model; recall that these two variables were present in all three log linear regressions (see Table 2.2). Also,
**TABLE 2.3**

Significant Variables from the Stepwise Proportional Hazards Model: Sample 4 **

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Involuntary admission</td>
<td>-.2339</td>
<td>.0574</td>
</tr>
<tr>
<td>Alcoholic psychosis</td>
<td>.3107</td>
<td>.0800</td>
</tr>
<tr>
<td>Psychosis/cerebral</td>
<td>-.3613</td>
<td>.1616</td>
</tr>
<tr>
<td>Mental retardation</td>
<td>.5040</td>
<td>.1765</td>
</tr>
<tr>
<td>Married</td>
<td>.1583</td>
<td>.0713</td>
</tr>
<tr>
<td>Widowed</td>
<td>-.2910</td>
<td>.1162</td>
</tr>
</tbody>
</table>

**All variables are indicators of the 0,1 form**
these two variables are the most significant of those listed in Table 2.3.

As a last exploratory probe, a proportional hazards model was fit to the lengths of stay in our fifth sample using the variables shown in Table 2.3; the results are listed in Table 2.4. Once again, legal status and alcoholic psychosis are significant; the coefficients for these variables are similar to those in Table 2.3. All of the other variables are not significant.

To summarize, we have fit several exploratory models to quantized length of stay. It appears that legal status is significantly associated with length of stay, with involuntary admissions having longer visits than voluntary admissions. Also associated is the alcoholic psychosis indicator—patients diagnosed as alcoholic psychotic tend to have shorter lengths of stay. Certain variables can be excluded from further consideration: multiple visits, first admission to the state mental health system, ethnicity, sex, age, and marital status do not seem to be worth pursuing. Psychiatric diagnosis does affect length of stay, though we do not need to consider all of the categories initially proposed.

Thus, on empirical grounds, we should focus on psychiatric diagnosis and legal status in further analyses. From the mental health administrator's perspective, however, age and sex breakdowns are important. For this reason, we will continue to consider age and sex as possible predictors of length of stay.

Before pressing on with more analysis, we need to specify what form the analysis should take. In the next chapter, a mathematical model for length of stay is derived, and its statistical properties are
<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Involuntary admission</td>
<td>-0.2030</td>
<td>0.0581</td>
</tr>
<tr>
<td>Alcoholic psychosis</td>
<td>0.4321</td>
<td>0.0770</td>
</tr>
<tr>
<td>Psychosis/cerebral</td>
<td>-0.2096</td>
<td>0.1448</td>
</tr>
<tr>
<td>Mental retardation</td>
<td>0.1150</td>
<td>0.1575</td>
</tr>
<tr>
<td>Married</td>
<td>0.1221</td>
<td>0.0681</td>
</tr>
<tr>
<td>Widowed</td>
<td>-0.1377</td>
<td>0.1136</td>
</tr>
</tbody>
</table>

** All variables are indicators of the 0,1 form
explored in detail. We return to the analysis of our covariates with this new model in Chapter IV.
III. BETA MIXTURE MODEL

Thus far, we have avoided deriving a model for length of stay from "first principles", yet such a model is highly desirable for interpreting an observed data set. In this chapter, we will derive such a model, and explore its mathematical and statistical properties. We will then use this model in Chapter IV to analyze the Central Islip data.

Let $L_i$ be the observed length of stay for the $i$th person in our sample. If $p_i$ is the probability that the $i$th person is released from the hospital in any given week, then $L_i$ may be thought to follow the geometric distribution

$$Pr\{L_i = l\} = p_i (1-p_i)^{l-1} \quad l = 1, 2, ... \quad (3.1)$$

This model could be fit by forcing $p_i$ dependent upon a group of covariates; the logit transformation

$$\log \frac{p_i}{1-p_i} = \chi_i' \beta \quad (3.2)$$

is one parameterization that comes to mind. However, for any particular combination of covariates, this model assumes that the conditional probability of release during any given week is constant over time and over individuals, an assumption this analyst finds unpalatable. For the
Central Islip data, observed release probabilities are not constant over time (Figure 3.1). Also, even among individuals with equal covariates, one should expect to find heterogeneity in the weekly probability of release. No covariate structure short of complete individual identification could yield otherwise.

A more realistic use of the geometric distribution is as follows. For the ith individual in a particular subpopulation (group of patients with equal covariates), the release probability $p_i$ may be treated as a random draw from a mixing density $f(p)$. Then, lengths of stay would follow a continuous mixture of geometrics:

$$
Pr \{ L_i = l \} = \int_0^1 p (1-p)^{l-1} f(p) \, dp \quad l = 1, 2, \ldots \quad (3.3)
$$

The model represented by (3.1) is that special case of (3.3) corresponding to a degenerate density over the release probability.

Given the formulation of (3.3), it is not clear how to estimate the mixing density $f(p)$ from observed data, though various ideas come to mind - these are discussed in Chapter V. Rather than computing a nonparametric estimate of the "prior" distribution, we will assume that $p_i$ follows the beta density

$$
f(p) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \Gamma(\beta)} p^{\alpha-1} (1-p)^{\beta-1} \quad 0 < p < 1 \quad (3.4)
$$

$\alpha, \beta > 0$
The resultant mass function for length of stay is then given by

\[ P_r \{ L_i = l \} = \frac{\alpha \Gamma(\alpha+\beta) \Gamma(\beta+l-1)}{\Gamma(\beta) \Gamma(\alpha+\beta+l)} \quad l=1,2,\ldots \quad (3.5) \]

Also, the survivor function is given by

\[ P_r \{ L_i \geq l \} = \frac{\Gamma(\alpha+\beta) \Gamma(\beta+l-1)}{\Gamma(\beta) \Gamma(\alpha+\beta+l-1)} \quad l=1,2,\ldots \quad (3.6) \]

Finally, the hazard function (or conditional probability of release) may be written as

\[ h(l) = \frac{P_r \{ L_i = l \}}{P_r \{ L_i \geq l \}} = \frac{\alpha}{\alpha+\beta+l-1} \quad l=1,2,\ldots \quad (3.7) \]
This model alleviates our previous concerns: individual release probabilities are now drawn from a mixing density, so observations from the same subpopulation are no longer assumed to behave identically. Also, the conditional release probability for a randomly chosen patient decreases with time. This can be intuitively explained as follows: during the first week, the probability of release equals \( \alpha/(\alpha + \beta) \), the mean of the beta mixture. However, as time goes on, those individuals with high release probabilities are actually released from the hospital. This forces the average release probability for those left in the hospital to decline, as the remaining patients had lower release probabilities to begin with.

The hazard function (3.7) provides us with a convenient graphical technique for checking the appropriateness of this beta-geometric mixture model for a given data set, for the reciprocal of \( h(\lambda) \) is a linear function of length of stay in the hospital. Thus, a plot of \( h(\lambda)^{-1} \) against \( \lambda \) should yield something close to a straight line if length of stay does follow the mass function derived in (3.5).

An empirical estimate of \( h(\lambda) \) can be obtained as follows.

Let

\[
\begin{align*}
  n_1(\lambda) &= \text{number of patients released during the } \lambda\text{th week after admission} \\
  n_2(\lambda) &= \text{number of patients remaining in the hospital during the } \lambda\text{th week after admission}
\end{align*}
\]
Then a simple estimate for \( h(t) \) is given by

\[
\tilde{h}(t) = \frac{n_1(t)}{n_1(t) + n_2(t)}.
\]  

(3.8)

Note that (3.8) takes care of censoring, as censored patients contribute to \( n_2(t) \) only for the number of weeks they were observed to remain in the hospital.

A one year (52 week) plot of \( \tilde{h}(t) \) for our sample of 6965 Central Islip patients is shown in Figure 3.1. That the empirical hazard is decreasing is undeniable, but whether the observed hazard conforms to (3.7) is unclear. Figure 3.2 is a plot of the reciprocal hazard, and it is clear that \( \tilde{h}^{-1}(t) \) is very unstable for large values of \( t \). This follows from the fact that \( n_1(t) \) declines quite rapidly, so the estimates \( \tilde{h}^{-1}(t) \) have large standard errors for lengths of stay greater than 10 weeks or so.

Table 3.1 shows the first ten weeks of data. Given that 6617 patients were actually observed to leave the hospital, these ten weeks account for 88% of observed releases, so we will now focus our attention on this subset of the sample.

For our reduced data set, Figure 3.3 shows \( \tilde{h}(t) \) while \( \tilde{h}(t)^{-1} \) is shown in Figure 3.4. Figure 3.4 is not perfectly linear, but it must be remembered that the variability of \( \tilde{h}(t)^{-1} \) is increasing with \( t \). To see this, we will obtain a rough estimate of the variance of \( \tilde{h}(t)^{-1} \). Treating \((n_1(t) + n_2(t)) \) as fixed, and regarding \( n_1(t) \) as a binomial random...
FIGURE 3.1

Empirical Hazard Function
FIGURE 3.2

Reciprocal Hazard Function
### TABLE 3.1

Length of Stay Data

<table>
<thead>
<tr>
<th>( \kappa )</th>
<th>( n_1(\kappa) )</th>
<th>( n_2(\kappa) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2488</td>
<td>4477</td>
</tr>
<tr>
<td>2</td>
<td>1082</td>
<td>3373</td>
</tr>
<tr>
<td>3</td>
<td>572</td>
<td>2781</td>
</tr>
<tr>
<td>4</td>
<td>435</td>
<td>2328</td>
</tr>
<tr>
<td>5</td>
<td>333</td>
<td>1984</td>
</tr>
<tr>
<td>6</td>
<td>224</td>
<td>1751</td>
</tr>
<tr>
<td>7</td>
<td>201</td>
<td>1544</td>
</tr>
<tr>
<td>8</td>
<td>175</td>
<td>1363</td>
</tr>
<tr>
<td>9</td>
<td>155</td>
<td>1201</td>
</tr>
<tr>
<td>10</td>
<td>128</td>
<td>1068</td>
</tr>
</tbody>
</table>
FIGURE 3.3

Empirical Hazard Function

HAZA

.420+

.350+

.280+

.210+

.140+

.070+

+---------+--------------

0 2.0 4.0 6.0 8.0 10.0

LOS
FIGURE 3.4
Reciprocal Hazard Function

INVH
10.0+
8.5+
7.0+
5.5+
4.0+
2.5+

+---------+---------------+---------------
    0.0     2.0     4.0     6.0     8.0     10.0

LOS
variable with success probability $h(\ell)$, we estimate the variance of $\hat{h}(\ell)$ as

$$\text{Var} \left( \hat{h}(\ell) \right) = \frac{\hat{h}(\ell) \left( 1 - \hat{h}(\ell) \right)}{n_1(\ell) + n_2(\ell)} .$$

(3.9)

Application of the delta method yields

$$\text{Var} \left( \frac{1}{\hat{h}(\ell)} \right) \approx \frac{\text{Var} \left( \hat{h}(\ell) \right)}{\hat{h}(\ell)^4} .$$

(3.10)

Now we see from Figure 3.3 how $\hat{h}(\ell)$ is decreasing, and (3.10) shows how this fact inflates the estimated variance of $\hat{h}(\ell)^{-1}$. Thus, we should give more weight to the linear part of Figure 3.4, and this suggests that the reciprocal of the true hazard $h(\ell)$ is also linear, at least for lengths of stay 10 weeks or less.

Having shown that the beta-geometric mixture model could describe most of the observed lengths of stay, we now turn to statistical issues of parameter estimation and hypothesis testing. As a first simple procedure, we can try to fit a line through $\hat{h}(\ell)^{-1}$ using least squares, taking into account the variability of our estimates. Ignoring the dependence among the $\hat{h}(\ell)$'s, we find the values $\gamma_0$ and $\gamma_1$ that minimize
Using (3.7), the beta parameters \( \alpha \) and \( \beta \) are estimated as

\[
\lambda \hat{=} \frac{1}{\hat{\gamma}},
\]

(3.12)

\[
\hat{\beta} = 1 + \frac{\hat{\gamma}_o - 1}{\hat{\gamma}},
\]

(3.13)

Application of this approach to the data in Table 3.1 yields \( \hat{\alpha} = 1.00 \) and \( \hat{\beta} = 1.88 \). Later, it will be shown that these estimates are remarkably close to the maximum likelihood estimates of \( \alpha \) and \( \beta \) for the Central Islip data; whether this is true in general is a matter for future research.

While the least squares approach is appealing in its simplicity, a better understood estimation procedure is necessary for hypothesis testing; this is particularly important when considering models involving several covariates. Therefore, we will turn to the method of maximum likelihood for estimating \( \alpha \) and \( \beta \). The probability of our observed data set consists of two parts: the probability of observed lengths of stay for patients actually released, and the probability of

\[
S(\hat{\gamma}_o, \hat{\gamma}_i) = \sum_{k=1}^{10} \frac{\hat{h}(k)^4}{\text{var}(\hat{h}(k))} \left( \frac{1}{\hat{h}(k)} - \hat{\gamma}_o - \hat{\gamma}_k \right)^2
\]

(3.11)
observed lengths of stay for censored patients. Let $l_i$ be the observed length of stay for the $i$th released patient, and $l_j$ be the observed length of stay for the $j$th censored patient (implying that the true length of stay is strictly greater than $l_j$ weeks). Assuming that all lengths of stay follow the beta-geometric mixture (3.5), the likelihood function is given by

$$
L = \prod_i P_r \{ L_i = l_i \} \prod_j P_r \{ L_j > l_j \} \tag{3.14}
$$

However, proceeding to maximize (3.14) or its logarithm over $\alpha$ and $\beta$ by directly substituting in equations (3.5) and (3.6) is not recommended, as the resulting mathematics becomes extremely difficult. Instead, we take advantage of the simple form of the hazard function (3.7), and note that

$$
P_r \{ L_i = l \} = h(l) \frac{l-1}{k=1} \frac{l}{(1 - h(k))} \tag{3.15}
$$

$$
P_r \{ L_i > l \} = \frac{l}{k=1} \frac{l}{(1 - h(k))} \tag{3.16}
$$
Using the above, we can reexpress the likelihood as

\[ L = \prod_{k} \frac{h(k)^{n_1(k)}(1-h(k))^{n_2(k)}}{k} \]  

(3.17)

and the log likelihood is given by

\[ \log L = \sum_{k} \left[ n_1(k) \log h(k) + n_2(k) \log (1-h(k)) \right] \]  

(3.18)

Expression (3.18) is particularly easy to manipulate.

The first derivatives of the log likelihood with respect to \( \alpha \) and \( \beta \) are obtained after substituting (3.7) into (3.18) as

\[ \frac{\partial \log L(\alpha, \beta)}{\partial \alpha} = \frac{1}{\alpha} \sum_{k} \left[ n_1(k) - h(k)(n_1(k) + n_2(k)) \right] \]  

(3.19)

\[ \frac{\partial \log L(\alpha, \beta)}{\partial \beta} = \sum_{k} \left[ \frac{n_2(k)}{\beta + k - 1} - \frac{h(k)(n_1(k) + n_2(k))}{\alpha} \right] \]  

(3.20)
The second derivatives of the log-likelihood are given by

\[
\frac{\partial^2 \log L (\alpha, \beta)}{\partial \alpha^2} = \frac{1}{\alpha^2} \sum_k \left[ n_1(k) (h(k)^2 - 1) + n_2(k) h(k)^2 \right]
\]

(3.21)

\[
\frac{\partial^2 \log L (\alpha, \beta)}{\partial \beta^2} = \frac{1}{\alpha^2} \sum_k \left[ h(k)^2 \left( n_1(k) + n_2(k) \left( 1 - \frac{1}{(1 - h(k))^2} \right) \right) \right]
\]

(3.22)

\[
\frac{\partial^2 \log L (\alpha, \beta)}{\partial \alpha \partial \beta} = \frac{1}{\alpha^2} \sum_k h(k)^2 \left( n_1(k) + n_2(k) \right)
\]

(3.23)

Equations (3.19) through (3.23) suggest two ways to obtain \(\hat{\alpha}\) and \(\hat{\beta}\), the maximum likelihood estimates of \(\alpha\) and \(\beta\). Our first method notes that \(\hat{\alpha}\) and \(\hat{\beta}\) solve the two nonlinear equations \(\partial \log L / \partial \alpha = 0\) and \(\partial \log L / \partial \beta = 0\). Setting (3.19) and (3.20) equal to zero and solving for \(\hat{\alpha}\) yields

\[
\hat{\alpha} = \frac{\sum n_1(k)}{\sum_k \frac{n_2(k)}{\hat{\beta} + k - 1}}
\]

(3.24)
Substituting (3.24) into (3.20) results in an implicit definition of \( \hat{\beta} \):

\[
\sum_{k} \left[ \frac{n_{2}(k)}{\hat{\beta} + k - 1} - \frac{n_{1}(k) + n_{2}(k)}{\sum_{k'} n_{1}(k') + \hat{\beta} + k - 1} \right] = 0 \quad (3.25)
\]

The following algorithm was used to solve (3.25):

(i) Specify \( \delta \), the convergence criterion (convergence is defined in step (iv) below).

(ii) Guess an initial value for \( \hat{\beta}^{(0)} \).

(iii) Set

\[
\hat{\alpha}^{(i+1)} = \frac{\sum_{k} n_{1}(k)}{\sum_{k} n_{2}(k)}
\]

(iv) If \( \left| \frac{\hat{\beta}^{(i+1)} - \hat{\beta}^{(1)}}{\hat{\beta}^{(1)}} \right| > \delta \), set \( i + i + 1 \) and go to (ii), otherwise stop.
Once \( \hat{\alpha} \) and \( \hat{\beta} \) are obtained, a consistent estimate of their covariance matrix is given by the inverse of the observed information matrix. The observed information is obtained from (3.21) through (3.23)

\[
I(\hat{\alpha}, \hat{\beta}) = -\begin{pmatrix} \frac{\partial^2 \log L(\alpha, \beta)}{\partial \alpha^2} & \frac{\partial^2 \log L(\alpha, \beta)}{\partial \alpha \partial \beta} \\ \frac{\partial^2 \log L(\alpha, \beta)}{\partial \alpha \partial \beta} & \frac{\partial^2 \log L(\alpha, \beta)}{\partial \beta^2} \end{pmatrix} \tag{3.26}
\]

so our estimate of the covariance matrix is given by

\[
\hat{\Sigma}(\hat{\alpha}, \hat{\beta}) = I(\hat{\alpha}, \hat{\beta})^{-1} \tag{3.27}
\]

This procedure was applied to the data in Table 3.1; the computer program used is shown in Appendix 1 along with a discussion of its performance. The resulting maximum likelihood estimates are \( \hat{\alpha} = 0.94 \) and \( \hat{\beta} = 1.76 \). The estimated covariance matrix is given by

\[
\hat{\Sigma}(0.94, 1.76) = 10^{-3} \times \begin{pmatrix} 1.125 & 2.725 \\ 2.725 & 8.050 \end{pmatrix}
\]
Our first observation is that these estimates are not that different from the estimates obtained from least squares ($\hat{\alpha} = 1.00$, $\hat{\beta} = 1.88$). It could well be that this seemingly complicated model can be fit using linear regression! These results suggest that it would be extremely worthwhile to determine the statistical properties of our least squares method.

It is also interesting to note how highly correlated $\hat{\alpha}$ and $\hat{\beta}$ are; indeed the observed correlation is given by

$$\rho_{\hat{\alpha}\hat{\beta}} = \frac{\text{cov}(\hat{\alpha}, \hat{\beta})}{\sqrt{\text{var}(\hat{\alpha}) \text{var}(\hat{\beta})}} = .91$$

This will be illustrated graphically later – for now, notice that equation (3.23) dictates $\hat{\alpha}$ and $\hat{\beta}$ to be positively correlated, as $\text{cov}(\hat{\alpha}, \hat{\beta})$ will always be positive.

As mentioned earlier, the derivatives of the log likelihood function suggest two ways to estimate parameters. The second approach is the well known Newton-Raphson technique for finding the maximum (or minimum) of a function. Let $\theta' = (\alpha, \beta)'$ be our vector of parameters. The Newton-Raphson method assumes that near the maximum likelihood estimate $\hat{\theta}$, the log likelihood is approximately quadratic:

$$\log L(\theta) \approx \log L(\hat{\theta}) + (\theta - \hat{\theta})' \frac{\partial \log L(\theta)}{\partial \theta} \bigg|_{\theta = \hat{\theta}}$$

$$- \frac{1}{2} (\theta - \hat{\theta})' I(\hat{\theta}) (\theta - \hat{\theta})$$

(3.28)
where $I(\hat{\theta})$ is the information matrix (3.26). To maximize (3.28), one differentiates the log likelihood with respect to $\theta$ and sets the result equal to zero yielding the iterative scheme

$$\hat{\theta}^{(i+1)} = \hat{\theta}^{(i)} + I(\hat{\theta}^{(i)})^{-1} \frac{\partial \log L(\theta)}{\partial \theta} \bigg|_{\theta=\hat{\theta}^{(i)}}$$

(3.29)

Given an initial estimate $\hat{\theta}^{(o)}$, (3.29) can be applied until successive estimates of $\hat{\theta}^{(i)}$ agree. The Newton-Raphson procedure seems more attractive here than in many other applications, as it is easy to obtain $\partial \log L(\theta)/\partial \theta$ and $I(\theta)$ for the beta-geometric model; a simple program is listed in Appendix 2.

The major tradeoff in using the Newton-Raphson method versus the iterative scaling approach discussed earlier is one of accuracy versus speed. If the Newton-Raphson method is given an initial estimate of $\theta$ close to the maximum likelihood estimate $\hat{\theta}$, then the procedure converges rapidly to $\hat{\theta}$. Unfortunately, if the initial guess is far from $\hat{\theta}$, then the method fails to converge. The iterative scaling approach appears to converge from a wide range of starting values. However, this method is very slow. It seems that a hybrid method combining the accuracy of the iterative scaling algorithm with the speed of the Newton-Raphson procedure could be developed; this is a separate issue deserving attention in the future.

Having obtained the maximum likelihood estimates $\hat{\alpha}$ and $\hat{\beta}$, we propose to use the inverse of the observed information matrix to estimate the
covariance matrix; this in turn will be used to make various statistical inferences about the true values $\alpha$ and $\beta$. The asymptotic theory upon which these tests depend assumes that the log likelihood function is quadratic near the maximum likelihood estimate. Noting that $\frac{\partial \log L(\theta)}{\partial \theta} \bigg|_{\hat{\theta}}$ is zero when $\hat{\theta}$ represents the true maximum likelihood estimate, the log likelihood should follow

$$
\log L(\theta) = \log L(\hat{\theta}) - \frac{1}{2} (\theta - \hat{\theta})' \mathbf{I}(\hat{\theta}) (\theta - \hat{\theta})
$$

(3.30)

if the asymptotic theory is true. A contour plot of (3.30) is shown in Figure 3.5; this plot also illustrates the degree to which $\hat{\alpha}$ and $\hat{\beta}$ are correlated.

To check on the quadratic approximation to the log likelihood, two cross-sections of the actual log likelihood are shown in Figure 3.6 and 3.7. In Figure 3.6, $\alpha$ is fixed at its maximum likelihood estimate while $\beta$ is varied; Figure 3.7 fixes $\beta$ at its MLE while $\alpha$ is varied. These curves are not symmetric. The log likelihood function is much steeper for parameter values smaller than the MLE's than for parameter values larger than the MLE's. However, the log likelihood does appear to look quadratic within several standard errors of $\hat{\alpha}$ and $\hat{\beta}$, so hypothesis tests based on the quadratic approximation to the log likelihood should be fairly accurate. Of course, one could also perform likelihood ratio tests, but such tests become cumbersome when dealing with models involving several covariates, an issue that will concern us in Chapter IV.
FIGURE 3.5

Contour Plot of Normal Approximation to the Likelihood Surface

BETA

9.0+
-  
-  
-  
-  
-  
6.0+
-  
-  
-  
-  
-  
-  
3.0+
-  
-  
-  
-  
-  
-  
0+
-  
-  
-  
-  
-  
-3.0+
-  
-  
-  
-  
-  
-6.0+  

+-----------------------+--------------------------+ALPH
-1.5  -.5  .5  1.5  2.5  3.5
FIGURE 3.6

Log-Likelihood versus Beta for Alpha=.94
FIGURE 3.7

Log-Likelihood versus Alpha for Beta=1.76

GLGA
-12000.+
-15000.+
-18000.+
-21000.+
-24000.+
-27000.+

+-------+--------------+---------------------------+ALPHA
  0     1.5    3.0    4.5    6.0    7.5
Before closing this section on the properties of the beta-geometric model, it is important to restate that the usual homogeneous geometric is a special case of the model we have been discussing. Suppose that a given data set did in fact stem from the simple geometric model. What should be expected if these data were overfit using a beta-geometric mixture? Recall from our earlier work that if we write the hazard as

\[ h(l) = \frac{1}{\gamma_0 + \gamma_1 l} \]

then the parameters \( \alpha \) and \( \beta \) are given by

\[ \alpha = \frac{1}{\gamma_1}, \quad \beta = 1 + \frac{\gamma_0 - 1}{\gamma_1} \]

If the true value of \( \gamma_1 \) equals zero (i.e. a simple geometric model), then the estimated value of \( \gamma_1 \) will be either small and positive, or small and negative due to random fluctuation. If \( \hat{\gamma}_1 \) is small and positive, then both \( \alpha \) and \( \beta \) will be large and positive (note that \( \gamma_0 > 1 \) if the simple geometric model is true). However, if \( \hat{\gamma}_1 \) is small and negative, then both \( \alpha \) and \( \beta \) will be large, negative numbers. Our derivation of the beta-geometric mixture model prohibits non-positive values of \( \alpha \) and \( \beta \), yet such values could arise due to chance fluctuations if the data actually followed a simple geometric distribution.

To summarize the work of this chapter, we began by axiomatically deriving a model for length of stay that is intuitively satisfying. We
showed how one can graphically check the appropriateness of this model, and then presented a crude estimation procedure based on least squares. Two techniques were proposed for obtaining maximum likelihood estimates of the model's parameters, and some graphical checks were made regarding the validity of asymptotic statistical assumptions. The model was fit to the length of stay data from Central Islip, and some hypothesis testing issues were discussed. Having examined the properties of this beta-geometric model in some detail, we are now ready to use this model to examine the relationship between length of stay and other variables present in the Central Islip data.
The empirical results of Chapter II provide a starting point for our more formal analysis of the Central Islip data. The log linear regressions and proportional hazards model suggested that both legal status and clinical diagnosis are closely associated with length of stay, while other variables are not. However, given the interest of mental health administrators and researchers in statistics conditioned by sex and age, it was decided to consider also these variables when fitting beta-geometric mixture models.

Age, sex and legal status remain defined as in Chapter II. Based on our earlier work, diagnosis is recategorized into six groups: schizophrenia, alcoholic psychosis, alcoholism, major affective disorders, psychosis associated with cerebral condition, other and unknown. Thus, we have a total of 120 (5 ages x 2 legal statuses x 2 sexes x 6 diagnoses) subpopulations to consider.

Our strategy consists of two steps:

1. Using the methods of Chapter III, obtain the beta-geometric maximum likelihood parameter estimates for each subpopulation.

2. Relying on asymptotic likelihood theory, fit models relating the parameters to the covariates using weighted least squares.

To carry out step 1., a program was written to apply the iterative scaling algorithm to the different subpopulations. Of the 120 groups, 19 contained too few observations to fit the model. The iterative scaling routine converged for 81 of the remaining 101 groups, revealing
that the routine won't work if $\alpha$ and $\beta$ are negative: all of the models fit using alternative methods yielded negative parameter estimates for $\alpha$ and $\beta$, and all of these models proved to be of the "overfit" variety discussed in the end of Chapter III.

These last twenty models were fit using the Newton-Raphson routine. Needless to say, some of these data sets defied many starting values, but eventually estimates were obtained. However, for nine of these models, estimated variances were so large that it seemed foolish to proceed with these subgroups.

We are left with 92 groups for which estimates of $\alpha$ and $\beta$ were found.

Let $\hat{\alpha}_{ijkl}$, $\hat{\beta}_{ijkl}$ = maximum likelihood estimate of $\alpha$ and $\beta$ for age category $i$, ($i = 1, \ldots, 5$), legal status $j$ ($j = 1, 2$), sex $k$ ($k = 1, 2$), and diagnosis $l$ ($l = 1, \ldots, 6$).

We can use weighted least squares to fit models of the form

$$\hat{\alpha}_{ijkl} = \mu + \alpha_{1i} + \alpha_{2j} + \alpha_{3k} + \alpha_{4l} + \alpha_{5(ij)} + \alpha_{6(jk)} + \alpha_{7(ik)} + \alpha_{8(ikl)} + \varepsilon_{ijkl} \tag{4.1}$$

$$\hat{\beta}_{ijkl} = \omega + \beta_{1i} + \beta_{2j} + \beta_{3k} + \beta_{4l} + \beta_{5(ij)} + \beta_{6(jk)} + \beta_{7(ik)} + \beta_{8(ikl)} + \varepsilon_{ijkl} \tag{4.2}$$
where the $u$ and $w$ terms are subject to the usual analysis of variance constraints (e.g. $\sum_{i} u_{1(i)} = \sum_{j} u_{2(j)} = \sum_{k} u_{3(k)} = \sum_{l} u_{4(l)} = 0$).

We assume that the $\hat{\alpha}$'s and $\hat{\beta}$'s are normally distributed with known variance. Then the weighted residual sum of squares is given by

$$R\sum{i,j,k,l} \frac{(\hat{\alpha}_{ijkl} - \hat{\alpha}_{ijkl})^2}{\hat{\alpha}_{ijkl}}$$

where $\hat{\alpha}_{ijkl}$ is the weighted least squares estimate of $\alpha_{ijkl}$; a similar formula holds for the $\hat{\beta}$'s. The residual sum of squares follows a $\chi^2$ distribution with degrees of freedom equal to 92 minus the number of parameters estimated. This $\chi^2$ statistic can be used to check the goodness of fit for any proposed model.

Several models were developed for $\alpha$ and $\beta$ - these are summarized in Table 4.1. It is interesting that the best models for $\hat{\alpha}$ and $\hat{\beta}$ contain the same terms - main effects due to clinical diagnosis and legal status, and legal/diagnosis interaction terms. This result isn't terribly surprising, as we saw in Chapter III that $\alpha$ and $\beta$ must be positively correlated.

Another thought regarding the similarity of the best models for $\hat{\alpha}$ and $\hat{\beta}$ - it is possible to simultaneously estimate the $u$ and $w$ terms of (4.1) and (4.2) using generalized least squares; this is desireable given that the $\hat{\alpha}$'s and $\hat{\beta}$'s are correlated. The weighted least squares approach assumes that the $\hat{\alpha}$'s and $\hat{\beta}$'s are mutually independent. Let $\theta$
### TABLE 4.1

**Models for Alpha**

<table>
<thead>
<tr>
<th>Covariates</th>
<th>Degrees of Freedom</th>
<th>Residual Sum of Squares</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>None (fixed alpha)</td>
<td>91</td>
<td>180.6</td>
<td>0.00</td>
</tr>
<tr>
<td>Age only</td>
<td>87</td>
<td>160.7</td>
<td>0.00</td>
</tr>
<tr>
<td>Legal Status only</td>
<td>90</td>
<td>177.0</td>
<td>0.00</td>
</tr>
<tr>
<td>Sex only</td>
<td>90</td>
<td>179.4</td>
<td>0.00</td>
</tr>
<tr>
<td>Diagnosis only</td>
<td>86</td>
<td>84.8</td>
<td>0.52</td>
</tr>
<tr>
<td>Legal and diagnosis</td>
<td>85</td>
<td>78.1</td>
<td>0.69</td>
</tr>
<tr>
<td>All main effects</td>
<td>80</td>
<td>71.5</td>
<td>0.74</td>
</tr>
<tr>
<td>Legal, diagnosis, and interaction</td>
<td>80</td>
<td>62.7</td>
<td>0.92</td>
</tr>
</tbody>
</table>
### Models for Beta

<table>
<thead>
<tr>
<th>Covariates</th>
<th>Degrees of Freedom</th>
<th>Residual Sum of Squares</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>None (fixed beta)</td>
<td>91</td>
<td>206.9</td>
<td>0.00</td>
</tr>
<tr>
<td>Age only</td>
<td>87</td>
<td>184.8</td>
<td>0.00</td>
</tr>
<tr>
<td>Legal Status only</td>
<td>90</td>
<td>206.7</td>
<td>0.00</td>
</tr>
<tr>
<td>Sex only</td>
<td>90</td>
<td>204.9</td>
<td>0.00</td>
</tr>
<tr>
<td>Diagnosis only</td>
<td>86</td>
<td>113.1</td>
<td>0.03</td>
</tr>
<tr>
<td>Legal and diagnosis</td>
<td>85</td>
<td>112.9</td>
<td>0.02</td>
</tr>
<tr>
<td>All main effects</td>
<td>80</td>
<td>103.7</td>
<td>0.04</td>
</tr>
<tr>
<td>Legal, diagnosis, and interaction</td>
<td>80</td>
<td>65.6</td>
<td>0.88</td>
</tr>
</tbody>
</table>
be the observed vector of estimated $\hat{a}, \hat{b}$ pairs, and let $\hat{v}_i$ be the estimated covariance matrix for the $i$th subgroup. Then we could estimate the covariance of $\hat{\theta}$ by

$$\hat{\Sigma} = \begin{pmatrix}
\hat{\sigma}_1^2 & \hat{\sigma}_{12} & \ldots & \hat{\sigma}_{1n} \\
\hat{\sigma}_{21} & \hat{\sigma}_2^2 & \ldots & \hat{\sigma}_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
\hat{\sigma}_{n1} & \hat{\sigma}_{n2} & \ldots & \hat{\sigma}_{nn}
\end{pmatrix}$$

(4.4)

For a matrix of covariates $X$ and a vector of $u$ and $w$ terms $\Psi$ we would specify the model

$$\hat{\Theta} = X\Psi$$

(4.5)

and estimate coefficients using

$$\hat{\Psi} = (X'\hat{\Sigma}^{-1}X)^{-1}X'\hat{\Sigma}^{-1}\hat{\Theta}$$

(4.6)

Goodness of fit would be checked using the statistic

$$\bar{Z}^2 = (\hat{\Theta} - X\hat{\Psi})'\hat{\Sigma}^{-1}(\hat{\Theta} - X\hat{\Psi})$$

(4.7)
where $Z^2$ follows a $\chi^2$ distribution with the appropriate degrees of freedom if (4.5) is true. Though we will not pursue this approach here, the block diagonal structure of (4.4) can be exploited to produce efficient computer codes for (4.6) and (4.7).

Table 4.2 presents the weighted least squares estimates of the $u$ and $w$ terms along with their standard errors. It is clear from this table that the interaction terms are very important. It is not so clear from this table how legal status and diagnosis affect length of stay.

For interpretive purposes, it is easier to examine the predicted values of $\alpha$ and $\beta$. These are shown in Table 4.3 along with their standard errors. With the notable exception of involuntarily admitted patients suffering from major affective disorders, all estimates could give rise to valid beta mixing distributions. As an example, Figure 4.1 plots the two distributions obtained for schizophrenics. Patients diagnosed as schizophrenic who are voluntarily admitted face the possibility of extremely long lengths of stay - this is due to the high probability mass associated with low probabilities of release. However, these same patients seem to face a better chance at an immediate release than involuntarily admitted schizophrenics. Involuntarily admitted schizophrenics have less variable lengths of stay; this is clear from the relative tightness of the beta mixture for these patients.

Alternatively, consider patients suffering from alcoholic psychoses; their beta mixtures are shown in Figure 4.2. Both voluntary and involuntary admissions face an initial drastic either/or syndrome: release comes immediately, or release perhaps never occurs until death! After the first few weeks in the hospital, individuals with release
## TABLE 4.2

Parameter Estimates from Weighted Least Squares

<table>
<thead>
<tr>
<th>Covariate</th>
<th>( \hat{u} )</th>
<th>se(( \hat{u} ))</th>
<th>( \hat{w} )</th>
<th>se(( \hat{w} ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>.601</td>
<td>.097</td>
<td>.425</td>
<td>.374</td>
</tr>
<tr>
<td>Voluntary</td>
<td>.119</td>
<td>.097</td>
<td>.993</td>
<td>.374</td>
</tr>
<tr>
<td>Schizophrenic</td>
<td>.927</td>
<td>.193</td>
<td>3.525</td>
<td>.683</td>
</tr>
<tr>
<td>Alc. Psychosis</td>
<td>-.039</td>
<td>.121</td>
<td>-.293</td>
<td>.375</td>
</tr>
<tr>
<td>Alcoholism</td>
<td>-.063</td>
<td>.116</td>
<td>.140</td>
<td>.404</td>
</tr>
<tr>
<td>Psych. Cereb.</td>
<td>-.208</td>
<td>.239</td>
<td>-.089</td>
<td>.445</td>
</tr>
<tr>
<td>Schiz &amp; Vol.</td>
<td>-.763</td>
<td>.193</td>
<td>-2.795</td>
<td>.683</td>
</tr>
<tr>
<td>Alcpsy &amp; Vol.</td>
<td>-.201</td>
<td>.121</td>
<td>-.960</td>
<td>.375</td>
</tr>
<tr>
<td>Alcohol &amp; Vol.</td>
<td>-.243</td>
<td>.116</td>
<td>-1.110</td>
<td>.404</td>
</tr>
<tr>
<td>Maj.A.D.&amp; Vol.</td>
<td>1.276</td>
<td>.388</td>
<td>7.412</td>
<td>1.753</td>
</tr>
<tr>
<td>Psycer &amp; Vol.</td>
<td>.163</td>
<td>.239</td>
<td>-.794</td>
<td>.445</td>
</tr>
</tbody>
</table>
### TABLE 4.3

Estimated Values of Alpha, Beta, and Standard Errors

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>Voluntary Admission</th>
<th>Involuntary Admission</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Alpha (st.err.)</td>
<td></td>
</tr>
<tr>
<td>Schizophrenic</td>
<td>.884 (.118)</td>
<td>2.171 (.392)</td>
</tr>
<tr>
<td>Alcoholic Psy.</td>
<td>.480 (.042)</td>
<td>.644 (.171)</td>
</tr>
<tr>
<td>Alcoholism</td>
<td>.414 (.036)</td>
<td>.662 (.152)</td>
</tr>
<tr>
<td>Maj.Aff.Dis.</td>
<td>1.143 (.770)</td>
<td>-1.647 (.505)</td>
</tr>
<tr>
<td>Psych. Cereb.</td>
<td>.675 (.531)</td>
<td>.110 (.053)</td>
</tr>
<tr>
<td>Other/Unknown</td>
<td>.726 (.060)</td>
<td>.949 (.082)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Beta (st.err.)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Schizophrenic</td>
<td>2.149 (.407)</td>
<td>5.752 (1.339)</td>
</tr>
<tr>
<td>Alcoholic Psy.</td>
<td>.165 (.020)</td>
<td>.099 (.059)</td>
</tr>
<tr>
<td>Alcoholism</td>
<td>.448 (.059)</td>
<td>.682 (.368)</td>
</tr>
<tr>
<td>Maj.Aff.Dis.</td>
<td>4.696 (3.585)</td>
<td>-12.115 (2.178)</td>
</tr>
<tr>
<td>Psych. Cereb.</td>
<td>.536 (.582)</td>
<td>.136 (.117)</td>
</tr>
<tr>
<td>Other/Unknown</td>
<td>.517 (.065)</td>
<td>2.036 (.253)</td>
</tr>
</tbody>
</table>
FIGURE 4.1
Beta Priors:
Schizophrenia

FIGURE 4.2
Beta Priors:
Alcoholic Psychosis
probabilities close to 1 have already been let go. Those remaining must face very long lengths of stay, with voluntary admissions experiencing lengthier visits than involuntary admissions.

For Table 4.3, it seems clear that the behavior of involuntarily admitted patients with major affective disorders cannot be well described by a beta-geometric mixture. To check this, maximum likelihood estimates of \( \alpha \) and \( \beta \) for involuntary admissions with major affective disorders were obtained using the Newton-Raphson procedure. The estimates are:

\[
\hat{\alpha} = -3.99, \quad \hat{\beta} = -23.80, \quad \hat{\lambda} = \begin{pmatrix} 3.49 & 17.22 \\ 17.22 & 87.00 \end{pmatrix}
\]

Both \( \hat{\alpha} \) and \( \hat{\beta} \) are significantly less than zero, confirming that a beta mixture interpretation is impossible for this subgroup. Rather, we note that for involuntary admissions with major affective disorders, the hazard function

\[
h(l) = \frac{-3.99}{-3.99 - 23.80 + l - 1}
\]

is increasing with time spent in the hospital. Indeed, if we solve \( \hat{h}(l) = 1 \) for \( l \), we find that according to this model, patients are
guaranteed to be released from the hospital within 25 weeks after admission.

That the hazard function for involuntary admissions with major affective disorders is increasing is clear from Figure 4.3. A simple linear weighted least squares fit to these data yields acceptable results (residual sum of squares = 5.17 compared to a \( \chi^2 \) with eight degrees of freedom). The estimated model then becomes

\[
\hat{h}(l) = 1.272 + 0.0083 l \\
(0.0131) (0.0029)
\]

where the numbers in parentheses are the estimated standard errors. Of course, this model is only valid for \( l \leq 10 \).

Why is it that involuntarily admitted patients with major affective disorders face an increasing hazard function? These patients are subject to dynamics different from what we studied in Chapter III. It may be that patients of this type face a specific course of treatment that is much less variable than treatments for other patients, though this cannot be verified from our data.

Our original objective was to determine the relationship between length of stay and various covariates. Having shown that legal status and diagnosis are strongly related to length of stay, the question arises as to which groups of patients stay longer than other groups. To answer this question, we computed the probability of staying greater than four weeks for our twelve diagnosis by legal status subgroups. The
FIGURE 4.3

Empirical Hazard for Involuntary Admits with Major Affective Disorders
results are shown in Table 4.4.

Patients with major affective disorders appear to have the highest chance of staying longer than one month; while alcoholic psychotics have the shortest observed lengths of stay. The distinction between voluntary and involuntary admissions has the strongest effect for patients suffering from psychoses associated with cerebral conditions, with involuntary admissions staying longer than voluntary admissions. Patients in the "other/unknown" diagnostic category also differ in their lengths of stay according to legal status. One could break down this category by finer diagnoses to investigate why this is so, but it seems clear that in general, involuntary admissions will correspond to more serious cases, requiring longer treatment periods.

To summarize this chapter, models describing length of stay were constructed in two stages. First, maximum likelihood estimates for subpopulations defined by various covariate structures were obtained using the methods of Chapter III. Then, structural models were fit to these estimates using weighted least squares. Clinical diagnosis, legal status and diagnosis/legal interaction terms satisfactorily describe the estimated parameters. These models were interpreted by studying the resultant beta mixture densities. In one instance, the mixture interpretation was inappropriate; this subpopulation (involuntary admissions diagnosed with major affective disorders) exhibited a significantly increasing hazard function.

The results of Chapter III and IV suggest many areas for future study. These ideas will be briefly discussed in the next chapter.
TABLE 4.4

Probability of Staying Longer than Four Weeks in the Hospital

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>Voluntary Admission</th>
<th>Involuntary Admission</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schizophrenic</td>
<td>.39</td>
<td>.34</td>
</tr>
<tr>
<td>Alcoholic Psy.</td>
<td>.13</td>
<td>.05</td>
</tr>
<tr>
<td>Alcoholism</td>
<td>.31</td>
<td>.25</td>
</tr>
<tr>
<td>Maj. Aff. Dis.</td>
<td>.50</td>
<td>.56</td>
</tr>
<tr>
<td>Psych. Cereb.</td>
<td>.20</td>
<td>.46</td>
</tr>
<tr>
<td>Other/Unknown</td>
<td>.18</td>
<td>.35</td>
</tr>
</tbody>
</table>
V. TOPICS FOR FUTURE RESEARCH

The analysis in Chapter IV and V assumes that length of stay can be modeled by a beta mixture of geometric distributions. Although the beta distribution is quite flexible in the number of shapes it can assume (as evidenced by Figures 4.1 and 4.2), the reason for using the beta distribution was mathematical tractability. Suppose we asked a more general question: assuming that length of stay has the mass function

\[ P_r \{ L = l \} = \int_0^1 \rho (1-\rho)^{l-1} f(\rho) \, d\rho \quad l=1, 2, \ldots \quad (5.1) \]

how could one estimate \( f(\rho) \) from observed data? This question was pursued somewhat during the course of our analysis, and some tentative results will be discussed here.

We employed two quite different procedures to try and estimate the prior mixing density \( f(\rho) \). Both of these procedures make the discrete approximation

\[ \int_0^1 \rho (1-\rho)^{l-1} f(\rho) \, d\rho \approx \sum_{k=1}^{K} q_k p_k (1-p_k)^{l-1} \quad (5.2) \]

subject to \( 0 < p_1 < p_2 < \ldots < p_K \); \( q_k > 0 \); \( \sum_{k=1}^{K} q_k = 1 \). Here the \( p_k \)'s are fixed numbers between 0 and 1, and \( q_k = \Pr\{ \rho = p_k \} \). Given values
of $P_{\ell}$, the empirical mass function for length of stay, one could try to estimate the $q_k$ by minimizing

$$G(q) = \sum_{\ell} \phi \left( P_{\ell} - \sum_{k=1}^{K} q_k \rho_k (1 - \rho_k) \ell \right)$$  \hspace{1cm} (5.3)

subject to $q_k > 0$; $\sum_{k=1}^{K} q_k = 1$, where $\phi(\cdot)$ is a weighting function.

One could also use the survivor distribution instead of the mass function for length of stay, since

$$P_r \{ L > \ell \} = \int_0^1 (1 - \rho)^\ell f(\rho) d\rho$$  \hspace{1cm} (5.4)

The advantage of the survivor distribution is that it can easily take censoring into account via the Kaplan-Meier estimation procedure (see Kalbfleisch and Prentice, 1980, p.10-16). If we let $\hat{F}(\ell)$ be the Kaplan-Meier estimate of the probability that length of stay is greater than $\ell$, then we can reformulate (5.3) as minimize

$$G_1(q) = \sum_{\ell} \phi \left( \hat{F}(\ell) - \sum_{k=1}^{K} q_k (1 - \rho_k) \ell \right)$$  \hspace{1cm} (5.5)

subject to $q_k > 0$; $\sum_{k=1}^{K} q_k = 1$. 
To use (5.5), we took \( \phi(\cdot) \) to be the absolute value function. Thus, our problem is to find those values \( q_k \) which minimize the sum of the absolute deviations between the observed and predicted survivor function. The solution to this problem can be obtained using the simplex method of linear programming.

We computed Kaplan-Meier survivor estimates for lengths of stay up to 25 weeks for each of the five samples referred to in Chapter II. Taking \( K = 20 \), we let \( p_k = 0.025, 0.075, 0.125, \ldots, 0.975 \). The resulting values of \( q_k \) are shown in Table 5.1. These five estimated priors are striking in their similarity, particularly for values of \( p_k \) less than 0.275.

The minimum absolute deviations approach was the first method used to estimate the prior mixing distribution. Our second approach, suggested by Sam Gutmann, is rather different. Suppose again that we use our discrete approximation (5.2). If we momentarily take a Bayesian view of our problem, we could obtain the posterior distribution for \( p_k \) given any observed length of stay using Bayes' rule

\[
q_k \mid \ell = \frac{q_k p_k (1-p_k)^{\ell-1}}{\sum_j q_j p_j (1-p_j)^{\ell-1}}
\]  

(5.6)

In fact, one could compute \( q_k \mid \ell \) for every individual in the sample. It stands to reason that a good estimate of the prior distribution of \( p \) should be close to the posterior distribution that would be obtained from
### TABLE 5.1

Minimum Absolute Deviations Estimates of $q_k$

<table>
<thead>
<tr>
<th>$p_k$</th>
<th>Sample 1</th>
<th>Sample 2</th>
<th>Sample 3</th>
<th>Sample 4</th>
<th>Sample 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>.025</td>
<td>.080</td>
<td>.104</td>
<td>.126</td>
<td>.101</td>
<td>.117</td>
</tr>
<tr>
<td>.075</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>.125</td>
<td>.217</td>
<td>.025</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>.175</td>
<td>.341</td>
<td>.454</td>
<td>.455</td>
<td>.308</td>
<td>.345</td>
</tr>
<tr>
<td>.225</td>
<td>.000</td>
<td>.000</td>
<td>.133</td>
<td>.231</td>
<td>.172</td>
</tr>
<tr>
<td>.275</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>.325</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>.375</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>.425</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>.475</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>.525</td>
<td>.000</td>
<td>.063</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>.575</td>
<td>.000</td>
<td>.252</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>.625</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>.675</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.117</td>
</tr>
<tr>
<td>.725</td>
<td>.204</td>
<td>.000</td>
<td>.077</td>
<td>.360</td>
<td>.243</td>
</tr>
<tr>
<td>.775</td>
<td>.158</td>
<td>.000</td>
<td>.209</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>.825</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>.875</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>.925</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>.975</td>
<td>.000</td>
<td>.102</td>
<td>.000</td>
<td>.000</td>
<td>.007</td>
</tr>
</tbody>
</table>
the correct prior. Thus, to estimate the prior, we will average the posteriors. Letting $m_\ell$ be the number of individuals with length of stay equal to $\ell$, we estimate the prior as the values of $q_k$ which solve the equation

$$q_k = \frac{\sum_\ell m_\ell q_{k|\ell}}{\sum_\ell m_\ell}$$

(5.7)

Equation (5.7) has no obvious solutions other than degenerate ones ($q_k = 1$ for any $k$), so the following procedure was used to estimate $q_k$.

i) Set $i = 0$. Assume $q_k^{(0)} = .05$ for $k = 1, 2, \ldots, K$. (=20).

ii) Set $q_k^{(i+1)}$ equal to the right hand side of (5.7), where (5.7) is computed using $q_k^{(i)}$.

iii) If $i < 10000$, go to ii).

This rather liberal algorithm was used in an attempt to ensure convergence, as we do not yet know whether or not this scheme converges in theory.

The $q_k$'s resulting from this procedure are shown in Table 5.2. Qualitatively speaking, these estimates are not that different from those in Table 5.1; in both tables, most of the estimated prior probability
### TABLE 5.2

Iterative Bayes Estimates of $q_k$

<table>
<thead>
<tr>
<th>$p_k$</th>
<th>Sample 1</th>
<th>Sample 2</th>
<th>Sample 3</th>
<th>Sample 4</th>
<th>Sample 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>.025</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>.075</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>.125</td>
<td>.090</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>.175</td>
<td>.559</td>
<td>.604</td>
<td>.668</td>
<td>.587</td>
<td>.581</td>
</tr>
<tr>
<td>.225</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>.275</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>.325</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>.375</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>.425</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>.475</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>.525</td>
<td>.000</td>
<td>.001</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>.575</td>
<td>.000</td>
<td>.028</td>
<td>.000</td>
<td>.000</td>
<td>.011</td>
</tr>
<tr>
<td>.625</td>
<td>.000</td>
<td>.129</td>
<td>.000</td>
<td>.007</td>
<td>.126</td>
</tr>
<tr>
<td>.675</td>
<td>.000</td>
<td>.114</td>
<td>.000</td>
<td>.205</td>
<td>.164</td>
</tr>
<tr>
<td>.725</td>
<td>.038</td>
<td>.039</td>
<td>.151</td>
<td>.194</td>
<td>.068</td>
</tr>
<tr>
<td>.775</td>
<td>.271</td>
<td>.009</td>
<td>.181</td>
<td>.007</td>
<td>.021</td>
</tr>
<tr>
<td>.825</td>
<td>.042</td>
<td>.003</td>
<td>.000</td>
<td>.000</td>
<td>.008</td>
</tr>
<tr>
<td>.875</td>
<td>.000</td>
<td>.002</td>
<td>.000</td>
<td>.000</td>
<td>.005</td>
</tr>
<tr>
<td>.925</td>
<td>.000</td>
<td>.004</td>
<td>.000</td>
<td>.000</td>
<td>.006</td>
</tr>
<tr>
<td>.975</td>
<td>.000</td>
<td>.068</td>
<td>.000</td>
<td>.000</td>
<td>.010</td>
</tr>
</tbody>
</table>
falls below \( p = .225 \). It is remarkable that these two different procedures yield such similar results.

However, we must temper our enthusiasm towards these methods. We do not know whether these methods, despite their agreeable results, are producing meaningful results. For example, we applied both methods to lengths of stay corresponding to the expected frequencies resulting from a uniform prior distribution. Neither method produced a set of \( q_k \)'s resembling a uniform distribution, so it is not clear that our approximation (5.2) is accurate. We hope to clarify some of these issues regarding estimation of the mixture distribution in our future research.

A second point concerns the use of the beta distribution as a mixing prior. Our example in Chapter III showed that the beta parameters estimated by weighted least squares were very close to the maximum likelihood estimates. This suggests that good estimates for this complicated model can easily be obtained via least squares. We would like to derive the statistical properties of this estimation procedure, and compare them to the properties of the maximum likelihood method.

Finally, our data seem to be well described by the beta-geometric model, at least for lengths of stay 10 weeks or less. An important issue is the degree to which the beta-geometric model can be thrown off due to one or two influential covariates, or one or two extreme lengths of stay. Thus, we intend to study the robustness of the beta-geometric model using both analytical and simulation methods.
REFERENCES


APPENDIX 1

Iterative Scaling Algorithm

The algorithm described in Chapter III works well for small values of the convergence criterion $\delta$. However, the algorithm is particularly slow to converge when $\delta$ is extremely small. Figure A1 shows the number of iterations required for different starting values, and different values of $\delta$. It should be noted that the algorithm converged to the maximum likelihood estimates for $\delta = .0001$ from all starting values considered. At the other extreme, using $\delta = .01$ resulted in convergence to $\hat{\alpha}$ and $\hat{\beta}$ only for starting values $\beta^{(0)}$ close to the true estimate. The performance for $\delta = .001$ was between these two extremes.
FIGURE A-1

Number of Iterations to Convergence for Iterative Scaling Algorithm

A <=> Delta=.01
B <=> Delta=.001
C <=> Delta=.0001
FORTRAN LISTING FOR ITERATIVE SCALING ALGORITHM

**DIMENSION N1(10), N2(10)**
DATA N1 /2488, 1082, 572, 435, 333, 224, 201, 175, 155, 128/
DATA N2 /4477, 3373, 2781, 2328, 1984, 1751, 1544, 1363, 1201, 1068/
J=0
DISPLAY "INPUT STARTING GUESS FOR BETA AND DELTA CRITERION"
ACCEPT BNEW, DELT
SN=0
DO 10 I=1, 10
SN=SN+N1(I)
10 CONTINUE

**C**
**C**
**C**
**C**
**C**
**C**

**COMPUTE ALPHA**

BOTTOM=0
BOLD=BNEW
J=J+1
DO 20 I=1, 10
BOTTOM=BOTTOM+N2(I)/(BOLD+I-1)
20 CONTINUE
A=SN/BOTTOM

**C**
**C**
**C**
**C**
**C**

**COMPUTE NEW ESTIMATE OF BETA**

TERM=0
DO 30 I=1, 10
TERM=TERM+(N1(I)+N2(I))/(A+BOLD+I-1)
30 CONTINUE
BNEW=BOLD*BOTTOM/TERM
DISPLAY "ITER=", J, "ALPHA=", A, "BETA=", BNEW
IF (ABS(BNEW-BOLD)/ABS(BOLD).GE.DELT) GO TO 5

**C**
**C**
**C**
**C**
**C**

**COMPUTE LOG-LIKELIHOOD**

GLOG=0
DO 40 I=1, 10
H=A/(A+BNEW+I-1)
GLOG=GLOG+N1(I)*ALOG(H)+N2(I)*ALOG(1.-H)
40 CONTINUE
DISPLAY "LOG-LIKELIHOOD=", GLOG
STOP
END
FORTRAN LISTING FOR NEWTON-RAPHSON ALGORITHM

**DIMENSION N1(10),N2(10)**

**DATA N1 /2488,1082,572,435,333,224,2C1,75,155,128/**
**DATA N2 /4477,3373,2781,2328,1984,1751,1544,1363,1201,1068/**

**J=0**

DISPLAY "ENTER GUESS FOR ALPHA AND BETA, AND SPECIFY DELTA"

ACCEPT ANEW,BNEW,DELTA

**C C DA=FIRST DERIVATIVE WRT ALPHA**

**C DB=FIRST DERIVATIVE WRT BETA**

**C DAA=SECOND DERIVATIVE WRT ALPHA**

**C DAB=CROSS DERIVATIVE**

**C DBB=SECOND DERIVATIVE WRT BETA**

**C VA=VARIANCE(ALPHA)**

**C VB=VARIANCE(BETA)**

**C VAB=COVARIANCE(ALPHA,BETA)**

**C RA=PROPOSED CORRECTION FOR ALPHA**

**C RB=PROPOSED CORRECTION FOR BETA**

**C BEGIN LOOP TO COMPUTE LOG LIKELIHOOD AND DERIVATIVES**

**30 AOLD=ANEW**

**BOLD=BNEW**

**DA=0**

**DB=0**

**DAA=0**

**DBB=0**

**DAB=0**

**GLOG=0**

**J=J+1**

**DO 10 I=1,10**

**C C H=ALPHA/(ALPHA+BETA+T-1) IS THE HAZARD FUNCTION**

**C**

**H=AOLD/(AOLD+BOLD+I-1)**

**DA=DA+N1(I)/AOLD-(N1(I)+N2(I))*H/AOLD**

**DB=DB+N2(I)/(BOLD+I-1)-(N1(I)+N2(I))*H/AOLD**

**DAA=DAA+N1(I)*(H*H-1)/(AOLD**2)+N2(I)*H*H/(AOLD**2)**

**DBB=DBB+N1(I)*H*H/(AOLD**2)+N2(I)*H*H*(1-1./(1.-H)**2)/(AOLD**2)**

**DAB=DAB+(N1(I)+N2(I))*H*H/(AOLD**2)**

**GLOG=GLOG+N1(I)*ALOG(H)+N2(I)*ALOG(1.-H)**

**10 CONTINUE**

**C C INVERT THE INFORMATION MATRIX TO FORM THE COVARIANCE MATRIX**

**C**

**DET=DAA*DBB-DAB*DAB**

**VA=-DBB/DET**

**VAB=DAB/DET**

**VB=-DAA/DET**
FORM CORRECTIONS

RA = VA * DA + VAB * DB
RB = VB * DB + VAB * DA

UPDATE ESTIMATES

ANEW = AOLD + RA
BNEW = BOLD + RB
DELTA1 = ABS((AOLD - ANEW) / AOLD)
DELTA2 = ABS((BOLD - BNEW) / BOLD)
DELT = DELTA1
IF (DELTA2 .GT. DELTA1) DELT = DELTA2
IF (DELT .GE. DELTA) GO TO 30
DISPLAY "ALPHA = ", ANEW
DISPLAY "BETA = ", BNEW
DISPLAY "DAA = ", DAA
DISPLAY "DAB = ", DAB
DISPLAY "DBB = ", DBB
DISPLAY "VAR(ALPHA) ", VA
DISPLAY "COV(ALPHA, BETA) ", VAB
DISPLAY "VAR(BETA) ", VB
DISPLAY "LOG LIKELIHOOD ", GLOG
DISPLAY "NUMBER OF ITERATIONS ", J
STOP
END