LIFE DISTRIBUTION MODELS AND INCOMPLETE DATA

by

RICHARD E. BARLOW
and
FRANK PROSCHAN

LEVEL

AD A095068

DTIC FILE COPY

DTIC
ELECTE
FEB 1 7 1981

C

UNIVERSITY OF CALIFORNIA · BERKELEY

81 2 13 010

LIFE DISTRIBUTION MODELS AND INCOMPLETE DATA

by

Richard E. Barlow
Department of Industrial Engineering
and Operations Research
University of California, Berkeley

and

Frank Proschan
Department of Statistics
Florida State University
Tallahassee, Florida

SEPTEMBER 1980   ORC 80-20

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ORC-80-20	2. GOVT ACCESSION NO. AD-A095 068	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) LIFE DISTRIBUTION MODELS AND INCOMPLETE DATA		5. TYPE OF REPORT & PERIOD COVERED Research Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Richard E. Barlow and Frank Proschan		8. CONTRACT OR GRANT NUMBER(s) AFOSR-77-3179
9. PERFORMING ORGANIZATION NAME AND ADDRESS Operations Research Center University of California Berkeley, California 94720		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 2304/A5
11. CONTROLLING OFFICE NAME AND ADDRESS United States Air Force Air Force Office of Scientific Research Bolling Air Force Base, D.C. 20332		12. REPORT DATE September 1980
		13. NUMBER OF PAGES 58
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Likelihood for Incomplete Data
Total Time on Test
Stopping Rules
Bayes Estimators
Weibull Distribution

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

(SEE ABSTRACT)

DD FORM 1473 EDITION OF 1 NOV 65 IS OBSOLETE
1 JAN 73
S/N 0102-014-6601

ACKNOWLEDGEMENTS

```
Accession For
NTIS   GRA&I        ✗
DTIC TAB            ☐
Unannounced         ☐
Justification ___ _ ____
_____ __ ____ _____
By__ _ _ ____
Distribution/
  Avail
Dist
A
```

ABSTRACT

This report represents the second chapter of a book in preparation on
inference and data analysis in reliability and life testing. The point
of view adopted differs from that of most books on the subject in the
following basic respect: *Prior information available to the reliability
analyst is utilized fully in a formal statistical fashion.* Experience
accumulated in helping engineers, quality assurance managers, scientists,
biostatisticians, and others who must make estimates and reach decisions
from either planned experiments or retrospective data has shown us that
the point of view adopted throughout the book has resulted in useful
solutions to real-life problems. By contrast, more classical statistical
methods have often proven inadequate in many practical problems simply
because the data available are insufficient to reach conclusions with a
desired degree of assurance.

The book is intended primarily for actual use by the engineering and
scientific practitioner, rather than for theoretical study and philo-
sophical analysis by the statistician. Thus we omit a philosophical
justification of the methods presented; rather, we rely on the fact that
they have led to useful answers to problems that have arisen in practice.

One final point: Many of the methods and results are original and have
not appeared in the literature. This fact has led us to issue the
chapters as reports under our research grants.

TABLE OF CONTENTS

LIFE DISTRIBUTION MODELS AND INCOMPLETE DATA

by

Richard E. Barlow and Frank Proschan

In this chapter our objective is to introduce additional life distribution models and to discuss methods useful for analyzing failure data, especially *incomplete* data. We show how to express the likelihood functions for general distributions and incomplete data. The likelihood function tends to be fairly flat for incomplete data. For this reason the maximum likelihood estimator may be of limited value. It is therefore especially important in this situation to assess a prior distribution for parameters and plot the posterior distribution or its contours.

Inference based on the exponential model is discussed for general sampling plans. Parameter estimators and credibility intervals are derived for special cases. The Weibull distribution is a very useful model for life distribution studies and also for the analysis of strength data. For these reasons we describe failure mechanisms leading to a Weibull life distribution model. Contour plotting methods for analyzing life data based on a Weibull distribution are also given.

1. LIKELIHOOD

In Chapter 1, we calculated the likelihood for the exponential model under several different sampling plans. In this section we present a unified way of analyzing incomplete data for a large number of failure distribution models. In much of this chapter we *assume* that the failure distribution F is absolutely continuous with density f and failure rate

2

$$(1.1) \qquad\qquad r(x) = \frac{f(x)}{\bar{F}(x)} \ ,$$

where $\bar{F}(x) = 1 - F(x)$. We call

$$(1.2) \qquad\qquad R(x) = \int_0^x r(u)du$$

the *hazard function* associated with F . For general F , define

$$(1.3) \qquad\qquad R(x) = -\ln \bar{F}(x)$$

so that $\bar{F}(x) = \exp [-R(x)]$. Note that when F has a density f ,

$$\frac{d}{dx} [-\ln \bar{F}(x)] = \frac{f(x)}{\bar{F}(x)} = r(x)$$

so that (1.2) and (1.3) agree in this case.

From (1.1) and (1.3) we see that

$$(1.4) \qquad\qquad f(x) = r(x)e^{-R(x)} \ .$$

For a discussion of these fundamental concepts, their inter-relation-
ships and illustrations in the case of well known distributions, see
Barlow and Proschan, Chapter 3, (1975).

Suppose now we observe n independent lifetimes $x_1, x_2, \ldots, x_n$
corresponding to a given failure rate function, r . The joint density
is

$$(1.5) \qquad\qquad \prod_{i=1}^{n} f(x_i) = \left[\prod_{i=1}^{n} r(x_i) \right] e^{- \sum_{i=1}^{n} R(x_i)} \ .$$

The likelihood as a function of the failure rate function for data $D = (x_1, x_2, \ldots, x_n)$ is then

$$(1.6) \qquad L(r(u), u \geq 0 \mid D) = \left[\prod_{i=1}^{n} r(x_i) \right] e^{-\sum_{i=1}^{n} R(x_i)} .$$

1.1 Example: The Time-Transformed Exponential Model

Suppose the survival function is of the form:

$$(1.7) \qquad \bar{F}(x \mid \lambda) = e^{-\lambda R_0(x)}$$

where it is assumed that R_0 is known and differentiable but λ is unknown. By (1.2) we may write:

$$\lambda R_0(x) = \int_0^x \lambda r_0(u) du .$$

It follows that the hazard function and the failure rate function are assumed known up to the parameter λ . Another way to view the model is to consider time x to be transformed by the function $R_0(\cdot)$. Thus in terms of transformed time, the present model coincides with the exponential model of Chapter 1. For this reason (1.7) is called the *time-transformed exponential model*.

Let $x_1, x_2, \ldots, x_n$ be n independent observations given λ from this model. The likelihood is

$$(1.8) \qquad L(\lambda \mid D) = \lambda^n \left[\prod_{i=1}^{n} r_0(x_i) \right] e^{-\lambda \sum_{i=1}^{n} R_0(x_i)} .$$

By Lemma 1.5 of Chapter 1, we conclude that $\sum\limits_{i=1}^{n} R_o(x_i)$ and n is

a *sufficient* statistics for λ . If we use the gamma prior for λ :

$$\pi(\lambda) = \frac{b^a \lambda^{a-1} e^{-b\lambda}}{\Gamma(a)} \, ,$$

we obtain as the posterior density for λ :

$$(1.9) \quad \pi(\lambda \mid D) = \left[b + \sum_{i=1}^{n} R_o(x_i) \right]^{a+n} \lambda^{a+n-1} \frac{\exp\left\{ -\lambda \left[b + \sum_{i=1}^{n} R_o(x_i) \right] \right\}}{\Gamma(a + n)} \, .$$

(See (1.6) of Chapter 1.) Inference proceeds exactly as for the exponential model, except that observation x_i of the exponential model is replaced by its time-transformed value $R_o(x_i)$. This is valid assuming only that $R_o(\cdot)$ is continuous.

1.2 The General Sampling Plan

In many practical life testing situations, the lifetime data collected are *incomplete*. This may be due to the sampling plan itself (recall the sampling plans (a), (b) and (c) of Chapter 1), or due to the unplanned withdrawal of test units during the test. (For example, in a medical experiment, one or more of the subjects may leave town, or suffer an accident, etc.)

We now describe one type of sampling plan. Suppose unit i having lifetime distribution F is observed over an interval of time starting at age 0 and ending at a random or nonrandom age. Termination of observation occurs in either one of the following two ways:

(1) The ith unit is withdrawn or lost from observation at age $\ell_i \geq 0$; ℓ_i may be random or nonrandom.

(2) The ith unit fails at age X_i , where X_i is a random variable.

In addition, we require a technical assumption regarding the "stopping rule"; i.e., a prescription for determining when to stop observation:

(3) Suppose unit lifetime, X , depends on an unknown parameter (or parameters) θ . Observation on a unit may stop before unit lifetime is observed. Let S be a *rule* or *set of instructions* which determines when observation of a unit stops. S may be random and depend on an unknown parameter (or parameters) ϕ . Let $(\tilde{\theta}, \tilde{\phi})$ have a joint prior distribution. Suppose that the rule S and $\tilde{\theta}$, given $\tilde{\phi}$, are judged independent. Also suppose that $\tilde{\theta}$ and $\tilde{\phi}$ are judged independent.

It is important to remark that the "stopping rule" is *not* necessarily the same as the "stopping time."

To understand assumption (3), consider sampling plan (a) of Chapter 1, namely: put n items on life test and stop testing at the kth observed failure. In this case, the stopping rule depends only on k and is clearly independent of life distribution parameters since k is fixed in advance of testing.

Sampling plan (b) of Chapter 1 has stopping rule: stop testing at time t_o . Since t_o is fixed in advance of testing, the stopping rule is again independent of life distribution parameters. Likewise,

6

the stopping rule for sampling plan (c) depends only on k and t_o

and is clearly also independent of life distribution parameters.

For all three sampling plans, the likelihood, up to a constant of

proportionality, depends only on the life distribution model and the

observed data. This proportionality constant depends on the stopping

rule and is $\frac{n!}{(n-k)!}$ for sampling plans (a) and (b) while for sampling

plan (c) it is $\frac{(n-1)!}{(n-k)!(k-1)!}$. Such stopping rules are said to be

noninformative since the posterior distribution in each case depends

only on the life distribution model, the observed data, the prior, and

in these examples, on the independence of unit lifetimes given distri-

bution parameters.

1.3 Examples of Informative Stopping Rules

Records are routinely kept on failures (partial or otherwise) and

maintenance actions on critical units such as airplane engines. Should

a relatively new type of unit start exhibiting problems earlier than anti-

cipated, this may trigger early withdrawal of units. If this happens

the stopping rule, which is contingent on performance, may also be

informative relative to life distribution parameters. This fact needs

to be considered when calculating the likelihood and analyzing the data.

The second example illustrates another case where assumption (3)

is violated. Suppose lifetime X is exponential with failure rate λ

and the random withdrawal time, W , is also exponential with parameter

ϕ . We observe the minimum of X and W . Furthermore, suppose that X

given λ and W given ϕ are judged independent. Then the likelihood

given an observed failure at x is

$$L(\lambda \, , \, \phi \mid x) = \lambda e^{-\lambda x} e^{-\phi x} \, .$$

If $\tilde{\lambda}$ and $\tilde{\phi}$ are judged a priori independent then the posterior density of $\tilde{\lambda}$ is

$$\pi(\lambda \mid x) \propto \lambda e^{-\lambda x} \pi(\lambda)$$

where π is the prior density for $\tilde{\lambda}$. However, if $\tilde{\lambda}$ and $\tilde{\phi}$ are judged *dependent* with joint prior $\pi(\lambda, \phi)$, then the posterior density is

$$\pi(\lambda \mid x) \propto \lambda e^{-\lambda x} \int_0^\infty e^{-\phi x} \pi(\lambda, \phi) d\phi .$$

The factor $\int_0^\infty e^{-\phi x} \pi(\lambda, \phi) d\phi$, contributed by the stopping rule, depends on λ.

There is an important case *not* covered by the General Sampling Plan—namely when it is known that a unit has failed within some time interval but the exact time of failure is unknown. This case will be covered when we discuss life tables.

The following simple example illustrates the way in which incomplete data can arise.

1.4 Example

Operating data are collected on an airplane part for a fleet of airplanes. A typical age history for several engines is shown in Figure 1.1. The crosses indicate the observed ages at failure. Ordered withdrawal times (nonfailure times) are indicated by short vertical lines. In our example, units 2 and 4 fail at respective times $x_{(1)}$ and $x_{(2)}$ while observation on units 1 and 3 is terminated without failure at times $\ell_{(2)}$ and $\ell_{(1)}$ respectively.
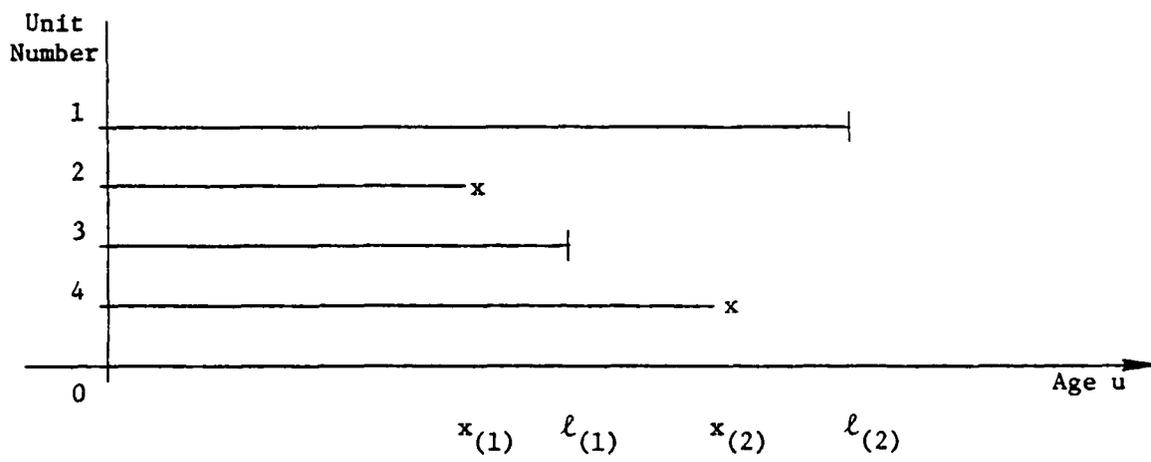
FIGURE 1.1

AGE OF AIRPLANE PART AT FAILURE OR WITHDRAWAL

It is important to note that all data are plotted against the *age axis*. Figure 1.2 illustrates how events may have occurred in calendar time. For example, units 1 and 3 had not failed at the end of the calendar record.

Total Time on Test

As we saw in Chapter 1, the total time on test is an important statistic for the exponential model.

1.5 Definition

The *total time on test* T is the total of the periods of observation of all the units undergoing test. Excluded from this statistic are any periods following death or withdrawal or preceding observation. Specifically, the periods being totalled include only those in which a death or a withdrawal of a unit under observation can be observed.

Let $n(u)$ be the number of units observed to be operating at *age* u . The observed function $n(u)$, $u \geq 0$, for Example 1.4 is displayed in Figure 1.3. From Figure 1.3 we may readily calculate the total time on test $T(t)$ corresponding to any age t , $0 \leq t \leq \ell_{(2)}$:

$$(1.10) \qquad T(t) = \int_0^t n(u)du .$$

For example, for t such that $x_{(2)} < t < \ell_{(2)}$, we obtain from Figure 1.3:

$$T(t) = \int_0^t n(u)du = 4x_{(1)} + 3(\ell_{(1)} - x_{(1)}) + 2(x_{(2)} - \ell_{(1)}) + (t - x_{(2)}) .$$

After simplifying algebraically, we obtain:
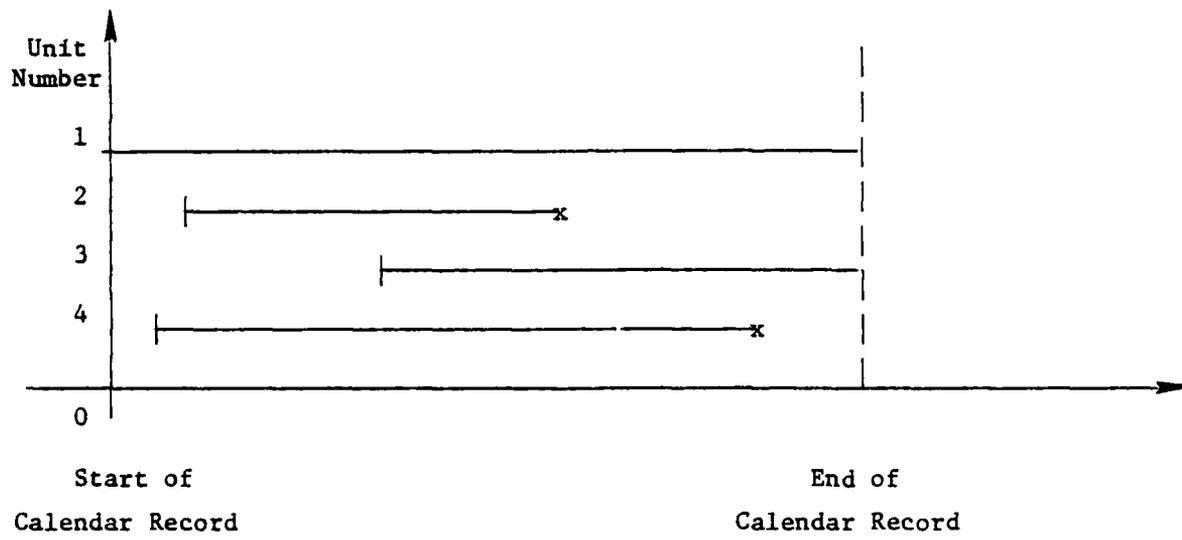
FIGURE 1.2

CALENDAR RECORD FOR AIRPLANE PARTS
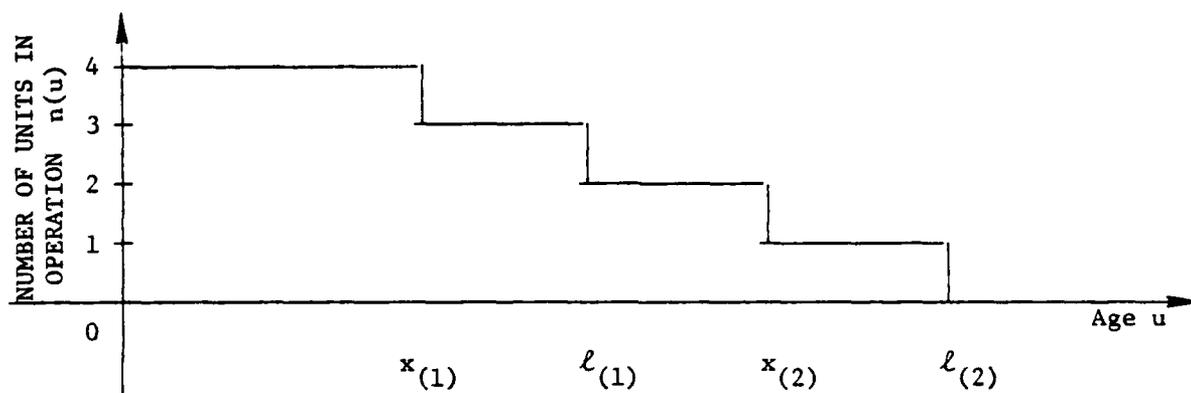
FIGURE 1.3

NUMBER OF UNITS IN OPERATION AS A FUNCTION OF AGE

12

$$(1.11) \qquad T(t) = x_{(1)} + \ell_{(1)} + x_{(2)} + t \ .$$

Note that the resulting expression, given in (1.11), can be obtained directly, since $x_{(1)}$ and $x_{(2)}$ represent the observed lifetimes of the 2 units that are observed to fail, $\ell_{(1)}$ represents the observed age of withdrawal of the unit first withdrawn from observation, and finally t represents the age of the second unit at the instant t specified.

Although in this small example, the directly calculated expression (1.11) for total time on test is simpler, Equation (1.10) is an important identity, since it yields the total time on test accumulated by age t in terms of the (varying) number of units on test at each instant during the interval $[0,t]$ for *any data set in which the ages at death or withdrawal are observed.* Thus it is a general formula applicable in a great variety of problems in which data may be incomplete.

Although $n(u)$ is a step function, the integral representation in (1.10) is advantageous, since it is compact, mathematically tractable, and applicable in a great variety of incomplete data situations. Of course, $\int_0^\infty n(u)du < \infty$ in practical problems since observation ultimately ceases in order to analyze the data in hand.

The Likelihood Function for Incomplete Data

All recorded data are necessarily discrete. Likewise real world life distribution models should also be discrete. Continuous life distribution models are convenient approximations to real world life distributions. However, it is most convenient to define initially the likelihood concept in the context of discrete models.

For our purposes, we find it preferable to define the likelihood concept for the General Sampling Plan in the context of a discrete model. Computation of the likelihood function is an intermediate step between specification of the prior distribution on the space Θ and computation of the posterior distribution on Θ given observed data D.

Suppose temporarily that the life distribution is discrete, i.e., failures can occur only at times $1, 2, \ldots$; similarly, withdrawals can occur only at these time points. Suppose that the probability of failure of a given unit at x is $p(x \mid \theta)$. Suppose k failures are observed at times x_s, $s = 1, \ldots, k$, and m withdrawals are observed at times ℓ_t, $t = 1, \ldots, m$. Failure and withdrawal times need *not* be distinct. All observations are assumed statistically independent, given parameters. Withdrawal times are produced by a stopping rule which is noninformative concerning θ.

For example, the stopping rule might specify that we observe a unit until failure or until withdrawal, whichever comes first, where withdrawal time is specified in advance. For this model, the probability of the observed outcome is:

$$(1.12) \qquad p(D \mid \theta) = \prod_{s=1}^{k} p(x_s \mid \theta) \prod_{t=1}^{m} \bar{P}(\ell_t \mid \theta) ,$$

where $\bar{P}(u_j \mid \theta) \overset{\text{def}}{=} \sum_{i=1}^{\infty} p(u_{j+i} \mid \theta)$ represents the probability that a specified unit fails at age u_{j+1} or later, given the parameter is θ. Note that the first product corresponds to the k failures at respective ages $x_1, \ldots, x_k$, while the second product corresponds to the m withdrawals at respective ages $\ell_1, \ldots, \ell_m$.

Another way to model withdrawal is to suppose there exists a

random withdrawal age W such that $P[W = t] = q(t)$, $t = 1,2, \ldots ,$ with W independent of unit lifetimes and of θ . Under this model, we suppose that we observe

$$\text{minimum } (X,W) = \begin{cases} X & \text{if } X \leq W \\ W & \text{if } X > W . \end{cases}$$

Now for observed data $D = \{x_1, \ldots, x_k , \ell_1, \ldots, \ell_m\}$, the probability of the observed outcome given parameter θ is:

$$(1.13) \quad p(D \mid \theta) = \prod_{t=1}^{m} q(\ell_t) \prod_{s=1}^{k} \bar{Q}(x_s) \prod_{s=1}^{k} p(x_s \mid \theta) \prod_{t=1}^{m} \bar{P}(\ell_t \mid \theta) ,$$

where $\bar{Q}(u_j) \overset{\text{def}}{=} \sum_{i=1}^{\infty} q(u_{j+i})$ represents the probability that $W > u_j$. Note that (1.12) and (1.13) differ only by a factor that does *not* depend on θ . Thus, *relative to calculating the MLE of* θ *, the two models for withdrawals (withdrawal deterministic or withdrawal random) do not differ essentially.*

There are many practical testing situations in which withdrawals occur as a result of chance mechanisms unrelated to the parameter θ of the lifetime distribution. For example, concluding the collection of data at a specified chronological time has the effect of withdrawing from observation those units still alive at that point in time. In Figure 1.2, this phenomenon is illustrated by units 1 and 3. Other chance mechanisms causing withdrawal at a random age result from human errors and accidents. The net effect of the various stopping rules that are unrelated to the value of the parameter θ is summarized in the factor $g(\underline{x},\underline{\ell})$ in the expression for the probability of the observed outcome:

$$(1.14) \quad p(D \mid \theta) = g(\underline{x},\underline{\ell}) \left[\prod_{s=1}^{k} p(x_s \mid \theta) \right] \prod_{t=1}^{m} \bar{P}(\ell_t \mid \theta) .$$

1.6 Definition

The *likelihood*, $L(\theta \mid D)$, is the probability of the observed outcome, $p(D \mid \theta)$, considered as a function of the parameter θ given the data, D. In the case of a continuous model, the corresponding likelihood will have this interpretation relative to a discrete probability approximation.

It follows from (1.14) that

$$(1.15) \qquad L(\theta \mid D) \propto \prod_{s=1}^{k} p(x_s \mid \theta) \prod_{t=1}^{m} \bar{P}(\ell_t \mid \theta) .$$

From Bayes' Theorem (Chapter 1, Theorem 1.2.), it is clear that we need not know $g(\underline{x}, \underline{\ell})$ in order to compute the posterior density of θ.

In this subsection, we have thus far confined our discussion to the case of discrete time life distributions since the basic concepts are easier to grasp in this case. However, in the case of continuous time life distributions, the likelihood concept is equally relevant, and in fact the expression for the likelihood $L(\theta \mid D)$ assumes a rather elegant form if we use $n(u)$, the number on test function. In the continuous case, $p(x \mid \theta)$ is replaced by the probability density element $f(x \mid \theta)$.

1.7 Theorem

Given the failure rate, independent observations are made under the General Sampling Plan. Let $x_1, x_2, \ldots, x_k$ denote the k observed failure ages. Let $n(u)$ denote the number of units under observation at age u, $u \geq 0$, and $r(u)$ denote the failure rate function of the unit at age u. Then the likelihood of the failure rate function $r(u)$,

having observed the data D described above, is given by:

$$L(r(u) , u \geq 0 \mid D)$$

$$(1.16) \quad \propto \begin{cases} \left[\prod_{s=1}^{k} r(x_s) \right] \exp \left[- \int_0^\infty n(u)r(u)du \right] , & k \geq 1 \\ \exp \left[- \int_0^\infty n(u)r(u)du \right] , & k = 0 . \end{cases}$$

<u>Proof</u>:

To justify (1.16), we first note that the underlying random events are the ages at failure or withdrawal. Thus the likelihood of the observed outcome is specified by the likelihood of the failure ages and survivals until withdrawal. By Assumption (3) of the General Sampling Model, we need not include any factor contributed by the stopping rule, since the stopping rule does not depend on the failure rate function $r(\cdot)$.

To calculate the likelihood, we use the fact that given $r(\cdot)$,

$$f(x) = r(x)e^{-\int_0^x r(u)du} .$$

(See (1.4).) Specifically, if a unit is observed from age 0 until it is withdrawn at age ℓ_t without having failed during the interval $[0, \ell_t]$, a factor $\exp \left[- \int_0^{\ell_t} r(u)du \right]$ is contributed to the likelihood. Thus, if *no* units fail during the test (i.e., k = 0) , the likelihood of the observed outcome is proportional to the expression

given in (1.16) for $k = 0$.

On the other hand, if a unit is observed from age 0 until it fails at age x_s , a factor,

$$r(x_s) \, \exp\left[- \int_0^{x_s} r(u)du\right] ,$$

is contributed to the likelihood. The exponential factor corresponds to the survival of the unit during $[0,x_s)$, while $r(x_s)$ represents the rate of failure at age x_s . (Note that if we had retained the differential element "dx" , the corresponding expression $r(x_s)dx$ would approximate an actual probability: the conditional probability of a failure during the interval $(x_s , x_s + dx)$ given survival to age x_s .)

The likelihood expression in (1.16) corresponding to the outcome $k \geq 1$ now is clear. The exponential factor corresponds to the survival intervals of both units that failed under observation and units that were withdrawn before failing:

$$\int_0^\infty n(u)r(u)du = \sum \int_0^{x_s} r(u)du + \sum \int_0^{\ell_t} r(u)du ,$$

where the first sum is taken over units that failed while the second sum is taken over units that were withdrawn. The upper limit "∞" is for simplicity and introduces no technical difficulty, since $n(u) \equiv 0$ after observation ends. ∎

The likelihood (1.16) applies for *any* absolutely continuous life distribution. In the important special case of an exponential life distribution model, $f(x \mid \lambda) = \lambda e^{-\lambda x}$, the likelihood of the observed outcome takes the simpler form:

$$(1.17) \qquad L(\lambda \mid D) \propto \begin{cases} \lambda^k \exp\left[-\lambda \int_0^\infty n(u)du\right] , & k \geq 1 \\[4mm] \exp\left[-\lambda \int_0^\infty n(u)du\right] , & k = 0 . \end{cases}$$

The following theorem is obvious from (1.17).

1.8 Theorem

Assume that the test plan satisfies Assumptions (1), (2) and (3) of the General Sampling Plan. Assume that k failures and the number of units operating at age u , $n(u)$, $u \geq 0$, are observed and that the model is the exponential density $f(x \mid \lambda) = \lambda e^{-\lambda x}$. Then

(a) k and $T = \int_0^\infty n(u)du$ together is a sufficient statistic for λ ;

(b) $\dfrac{k}{T}$ is the MLE for λ .

Note that the MLE, $\dfrac{k}{T}$, for λ represents the *number of observed failures divided by the total time on test*. This conforms with the results obtained for the MLE under all the test plans consid-red in Chapter 1.

The maximum likelihood estimator is the mode of the posterior density corresponding to a uniform prior (over an interval containing the MLE). A uniform prior is often a convenient *reference* prior. Under suitable circumstances, the analyst's actual posterior distribution will be approximately what it would have been had the analyst's prior been uniform. To ignore the departure from uniformity, it is sufficient that the analyst's actual prior density changes gently in the region favored by the data and also that the prior density not too strongly favors some other region. This result is rigorously expressed in the *Principle of Stable Estimation* (see Edwarls, Lindman and Savage (1963)). DeGroot (1970), pages 198-201, refers to this result under the name of *precise measurement*.

1.9 Example

The exact likelihood can be calculated explicitly for specified stopping rules. Suppose that withdrawal times are determined in advance. Then the likelihood is

$$(1.18) \qquad L(r(u) , u \geq 0 \mid D) = \left[\prod_{s=1}^{k} n\left(x_s^-\right) r(x_s) \right] e^{-\int_0^\infty n(u) r(u) du}$$

where $n\left(x_s^-\right)$ is the number surviving just prior to the observed failure at age x_s. To see this consider the airplane engine data in Example 1.4. Using Figure 1.3 as a guide, the likelihood will have the following factors:

1. For the interval $[0, x_{(1)}]$ we have the contribution

$$4r(x_{(1)}) \exp \left[- \int_0^{x_{(1)}} 4r(u)du \right]$$

corresponding to the probability that all 4 units survive to $x_{(1)}^-$ and the first failure occurs at $x_{(1)}$;

2. For the interval $(x_{(1)}, \ell_{(1)}]$ we have the contribution

$$\exp \left[- \int_{x_{(1)}}^{\ell_{(1)}} 3r(u)du \right]$$

corresponding to the conditional probability that the remaining 3 survive this interval;

3. For the interval $(\ell_{(1)}, x_{(2)}]$ we have the contribution

$$2r(x_{(2)}) \exp \left[- \int_{\ell_{(1)}}^{x_{(2)}} 2r(u)du \right]$$

corresponding to the conditional probability that the remaining 2 units survive to $x_{(2)}^-$ and a failure occurs at $x_{(2)}$;

4. For the interval $(x_{(2)}, \ell_{(2)}]$ we have the contribution

$$\exp \left[- \int_{x_{(2)}}^{\ell_{(2)}} r(u)du \right]$$

corresponding to the conditional probability that the remaining unit survives to age $\ell_{(2)}$. Multiplying together these conditional probabilities, we obtain a likelihood having the form shown in (1.18).

2. PARAMETER ESTIMATORS AND CREDIBLE INTERVALS

In the previous section we saw how to calculate the likelihood function for general life distributions. This is required in order to calculate the posterior distribution. Calculation and possibly graphical display of the posterior density would conceivably complete our data analysis.

If we assume a life density $p(x \mid \theta)$ and $\pi(\theta)$ is the prior, then $p(x,\theta) = p(x \mid \theta)\pi(\theta)$ is the joint density and $p(x) = \int_{\textcircled{H}} p(x \mid \theta)\pi(\theta)d\theta$ is the marginal or *predictive* density. Given data D and the posterior density $\pi(\theta \mid D)$, the predictive density is

$$p(x \mid D) = \int_{\textcircled{H}} p(x \mid \theta)\pi(\theta \mid D)d\theta .$$

If asked to give the probability of survival until time t , we would calculate

$$P(X > t \mid D) = \int_t^\infty p(x \mid D)dx .$$

2.1 Example

For the exponential density $\lambda e^{-\lambda x}$, k observed failures, T total time on test, and the General Sampling Plan, the likelihood is proportional to $\lambda^k e^{-\lambda T}$. For the natural conjugate prior,

$$\pi(\lambda) = \frac{b^a \lambda^{a-1} e^{-b\lambda}}{\Gamma(a)} ,$$

22

the posterior density is

$$\pi(\lambda \mid k , T) = (b + T)^{a+k} \lambda^{a+k-1} e^{-(b+T)\lambda} \Big/ \Gamma(a + k) .$$

In this case the probability of survival until time t is

(2.1)
$$P(X > t \mid k , T) = \int_0^\infty e^{-\lambda t} \pi(\lambda \mid k , T) d\lambda$$

$$= \left(\frac{b + T}{b + t + T}\right)^{a+k} .$$

Bayes Estimators

We will need the following notation:

$$E[\tilde{\theta}] = \int_\Theta \theta \pi(\theta) d\theta$$

and

$$E[\tilde{\theta} \mid D] = \int_\Theta \theta \pi(\theta \mid D) d\theta .$$

Of course, $E[\tilde{\theta}]$ is the mean of the prior distribution while $E[\tilde{\theta} \mid D]$ is the mean of the posterior distribution.

We wish to select a single value as representing our "best" estimator of the unknown parameter θ . To define the best estimator we must specify a criterion of goodness (or equivalently, of poorness).

Statisticians, being pessimistic by nature, measure the poorness of an estimator $\hat{\theta}$ by the expected "loss" resulting from their estimator $\hat{\theta}$. One very popular loss function is *squared error loss*: specifically, having observed data D and determined the posterior density $\pi(\theta \mid D)$, the expected squared error loss is given by:

$$(2.2) \qquad E[(\tilde{\theta} - \hat{\theta})^2 \mid D] ;$$

the expectation is calculated with respect to the posterior density $\pi(\theta \mid D)$. We choose a point estimator $\hat{\theta}$ so as to minimize the expected squared error loss in (2.2); i.e., we choose $\hat{\theta}$ to satisfy:

$$(2.3) \qquad \underset{a}{\text{minimum}} \; E[(\tilde{\theta} - a)^2 \mid D] = E[(\tilde{\theta} - \hat{\theta})^2 \mid D] .$$

To find the minimizing value $\hat{\theta}$, we add and subtract $E(\tilde{\theta} \mid D)$ in the loss function to obtain:

$$E[(\tilde{\theta} - a)^2 \mid D] = E[(\tilde{\theta} - E(\tilde{\theta} \mid D))^2 \mid D] + [E(\tilde{\theta} \mid D) - a]^2 .$$

Since we wish to minimize the right hand side, we set $a = E(\tilde{\theta} \mid D)$, which then represents the solution to (2.3). The resulting estimator, $E(\tilde{\theta} \mid D)$, the mean of the posterior, is called the *Bayes estimator with respect to squared error loss*.

2.2 Theorem

The Bayes estimator of a parameter θ with respect to squared

loss is the mean $E(\tilde{\theta} \mid D)$ of the posterior density.

Another loss function in popular use is the *absolute value loss function*:

$$(2.4) \qquad\qquad E[\,|\tilde{\theta} - \hat{\theta}|\;\mid D]\;.$$

To find the minimizing estimator using this criterion, we choose $\hat{\theta}$ to satisfy:

$$(2.5) \qquad \underset{a}{\text{minimum}}\; E[\,|\tilde{\theta} - a|\;\mid D] = E[\,|\tilde{\theta} - \hat{\theta}|\;\mid D]\;.$$

It is easy to show:

2.3 Theorem

The *Bayes estimator* of a parameter θ with respect to the absolute value loss function is the median of the posterior density. Specifically, the estimator $\hat{\theta}$ satisfies:

$$(2.6) \qquad \int_{-\infty}^{\hat{\theta}} \pi(\theta \mid D)d\theta = \int_{\hat{\theta}}^{\infty} \pi(\theta \mid D)d\theta = \frac{1}{2}\;.$$

The proof is left to the Exercises.

Of course, the prior density and the loss function enter crucially in determining a "best" estimator. However, no matter what criterion is used, all the information concerning the unknown parameter θ is contained in the posterior density. Thus, a graph of $\pi(\theta \mid D)$ is more informative than any single parameter of the posterior density, whether it be the mean, the median, the mode, a quartile, etc.

2.4 Example

Assume that lifetime is governed by the exponential model, $\frac{1}{\theta} e^{-\frac{x}{\theta}}$.
Suppose we conjecture that $E[\tilde{\theta} \mid k , T]$, for sampling plan with
k , T sufficient, is linear in T for fixed k . It turns out that
such a linear relationship holds if and only if we use as our prior the
natural conjugate prior:

$$\pi(\theta) = \frac{b^a \theta^{-(a+1)} e^{-\frac{b}{\theta}}}{\Gamma(a)} .$$

(See Section 3 of Chapter 1 for a discussion of this natural conjugate
prior and Diaconis and Ylvisaker (1979) for a proof of this result and
for more general results of this kind.) The corresponding Bayes
estimator with respect to squared error loss is:

$$(2.7) \qquad\qquad E[\tilde{\theta} \mid k , T] = \frac{(b + T)}{(a + k - 1)} .$$

However, the natural conjugate prior would not be appropriate if we
believed, for example, that θ could assume values only in two disjoint
intervals. Under this belief, a bimodal prior density would be more
natural, and the corresponding estimator $E[\tilde{\theta} \mid D]$ would very likely
be difficult to obtain in closed form such as in (2.7). However
$E[\tilde{\theta} \mid D]$ could be computed by numerical integration.

There are many other functions of unknown parameters for which we may
want the Bayes estimator with respect to squared error loss. For example,
we may wish to estimate the probability of survival until age t for the
exponential model; i.e., estimate

$$(2.8) \qquad g(\theta) = \exp\left[-\frac{t}{\theta}\right] .$$

It is easy to show in this case that

$$(2.9) \qquad \hat{g} = E\left[\exp\left[-\frac{t}{\theta}\right] \mid D\right]$$

is the Bayes estimator. If $\pi(\theta)$ is the natural conjugate prior, then it is easy to verify that

$$\hat{g} = \left[\frac{b + T}{b + t + T}\right]^{a+k} ,$$

i.e., this is the Bayes estimator of the probability of survival to age t given total time on test T and k observed failures. Note that $\hat{g} = \left[\frac{b + T}{b + t + T}\right]^{a+k}$ is precisely the marginal probability of survival until time t.

Credible Intervals

As we have seen, Bayes estimators correspond to certain functions of the posterior distribution such as the mean, the mode, etc. A credible set or interval is another way of presenting a partial description of the posterior distribution.

Specifically, we choose a set C on the positive axis (since we are dealing with lifetime) such that

$$(2.10) \qquad \int_C \pi(\theta \mid D)d\theta = 1 - \alpha .$$

Such a set C is called a *Bayesian* $(1 - \alpha)100$ percent *credible set*

(or credible interval if C is an interval) for θ .

Obviously, the set C is not uniquely determined. It would seem

desirable to choose the set C to be as small (e.g., least length,

area, volume) as possible. To achieve this, we seek a constant

$c_{1-\alpha}$ and a corresponding set C such that:

$$(2.11) \qquad C = \{\theta \mid \pi(\theta \mid D) \geq c_{1-\alpha}\}$$

and

$$(2.12) \qquad \int_C \pi(\theta \mid D)d\theta = 1 - \alpha .$$

A set C satisfying (2.11) and (2.12) is called a *highest posterior*

density credible set [Box and Tiao (1973)]. In general, C would

have to be determined numerically with the aid of a computer.

For the exponential model $\lambda e^{-\lambda x}$, we have seen that the natural

conjugate prior (Table 3.1, Chapter 1) is the gamma density. Since

the gamma density is a generalization of the chi-square density, we

recall the definition of the latter so that we can make use of it to

determine credible intervals for the failure rate of the exponential.

2.5 Definition

A random variable $\chi^2(n)$ having density:

$$(2.13) \qquad f_{\chi^2(n)}(x) = \frac{x^{\frac{n}{2}-1} \exp\left[-\frac{x}{2}\right]}{2^{\frac{n}{2}}\Gamma\left[\frac{n}{2}\right]} \quad \text{for} \quad x \geq 0 , \ n = 1,2, \ldots ,$$

is called a *chi-square random variable* with n *degrees of freedom* (d.f.).

A table of percentage points of the chi-square distribution may be found in Pearson and Hartley (1958). In addition, chi-square programs are available for more extensive calculations using electronic computers and programmable calculators.

It is easy to verify that the χ^2 random variable with 2n d.f. is distributed as $2(Y_1 + Y_2 + \dots + Y_n)$, where $Y_1, Y_2, \dots, Y_n$ are independent, exponentially distributed random variables with mean one. Thus, we obtain the following result useful in computing credibility intervals for the failure rate of the exponential model with corresponding natural conjugate prior.

2.6 Theorem

Let k failures and total time on test T be observed under sampling assumptions (1), (2) and (3), (Section 1) for the exponential model $\lambda e^{-\lambda x}$. Let $\tilde{\lambda}$ have posterior density corresponding to the natural conjugate prior

$$\pi(\lambda) = \frac{b^a \lambda^{a-1} e^{-b\lambda}}{\Gamma(a)} \ ,$$

with a an integer. Then

$$(2.14) \qquad P\left[\frac{\chi^2_{\frac{\alpha}{2}}[2(a + k)]}{2(b + T)} \leq \tilde{\lambda} \leq \frac{\chi^2_{1-\frac{\alpha}{2}}[2(a + k)]}{2(b + T)} \ \middle| \ D \right] = 1 - \alpha \ ,$$

where $\chi^2_\beta(n)$ is the 100β percentage point of a chi-square distribution with n d.f.; i.e., $\displaystyle\int_0^{\chi^2_\beta(n)} f_{\chi^2(n)}(x)dx = \beta$.

<u>Remark</u>:

Because of the lack of symmetry of the χ^2 density, the interval in (2.14) is *not* the highest posterior density credible interval.

<u>Proof</u>:

It is easy to verify that $(b + T)\tilde{\lambda}$ given the data has a gamma density,

$$\frac{\lambda^{a+k-1} e^{-\lambda}}{\Gamma(a + k)} \, ,$$

corresponding to the density of $Y_1 + \ldots + Y_{a+k}$, where the Y's are independent unit exponential random variables. Hence

$$2\tilde{\lambda}(b + T) \underset{st}{=} 2(Y_1 + \ldots + Y_{a+k}) \, ,$$

where $\underset{st}{=}$ denotes stochastic equality; i.e., $2\tilde{\lambda}(b + T)$ has a chi-square density with $2(a + k)$ d.f. ∎

2.7 Corollary

For $2(a + k)$ large (say $2(a + k) > 30$) , the normal approximation provides the approximate credibility statement

$$(2.15) \quad P\left[\frac{(a + k) + (a + k)^{\frac{1}{2}} z_{\frac{\alpha}{2}}}{b + T} \leq \tilde{\lambda} \leq \frac{(a + k) + (a + k)^{\frac{1}{2}} z_{1-\frac{\alpha}{2}}}{b + T} \; \Big| \; D\right] \doteq 1 - \alpha \, ,$$

where z_α satisfies $\displaystyle\int_{-\infty}^{z_\alpha} \phi(u)du = \alpha$ and $\phi(u) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$ is the normal density with mean 0 and variance 1.

30

Proof:

Since the $\chi^2(2n)$ random variable can be written as

$$\chi^2(2n) = 2(Y_1 + Y_2 + \dots + Y_n)$$

where $Y_1, Y_2, \dots, Y_n$ are independent unit exponentials, the Central Limit Theorem [e.g., Hoel, Port and Stone (1971)] applies. Note that $E\chi^2(2n) = 2n$ and $\text{Var}\,[\chi^2(2n)] = 4n$. Thus,

$$\frac{\chi^2(2n) - 2n}{2\sqrt{n}}$$

is approximately normal with mean 0 and variance 1 by the Central Limit Theorem. ■

2.8 Corollary

Let k failures and T total time on test be observed under the General Sampling Plan assumptions (1), (2) and (3) (see Section 1), for the exponential model $\frac{1}{\theta} e^{-\frac{x}{\theta}}$. Let $\tilde{\theta}$ have the natural conjugate prior with integer a, then

$$(2.16) \qquad P\left[\frac{2(b + T)}{\chi^2_{1-\frac{\alpha}{2}}[2(a + k)]} \le \tilde{\theta} \le \frac{2(b + T)}{\chi^2_{\frac{\alpha}{2}}[2(a + k)]} \;\middle|\; D\right] = 1 - \alpha .$$

Proof:

Since $\tilde{\theta}$ has the natural conjugate prior distribution for the model $\theta^{-1} e^{-\frac{x}{\theta}}$, then $\tilde{\lambda} = \frac{1}{\theta}$ has the natural conjugate prior for the model

$\lambda e^{-\lambda x}$. (2.16) follows from (2.14). ∎

2.9 Example

We use the data in Exercise 7, Chapter 1, to illustrate the calculation of the credibility limits of (2.16). In this case, n = 13 , k = 10 , and T = 275 weeks. Figure 2.1 displays prior and posterior 95% credibility limits for selected values of the natural conjugate prior parameters for the exponential model $\frac{1}{\theta} e^{-\frac{x}{\theta}}$. For the improper prior $\pi(\theta) = \frac{1}{\theta}$ corresponding to a = b = 0 we cannot calculate a 95% interval. However, the posterior is proper in this case and we can calculate a 95% credible interval. This is the same as the usual sample theory 95% confidence interval. It is, however, wider than posterior credible intervals corresponding to proper priors. The credible interval corresponding to the prior with a = 8 , b = 120 was computed using Corollary 2.7. For this relatively small number of failures (k = 10) , the choice of a prior is crucial and should be considered rather carefully.

2.10 Example: Credible Limits on the Failure Rate Average

In many life test data collections only the *numbers* of failures in specified time intervals are recorded, and not the more informative *times* of failure. For example, as part of its quality assurance program, a company producing semiconductors, routinely selects components from a product batch and subjects them to an accelerated life test. Specifically, early failure is induced by operating the components in a high temperature environment. At the end of each week, components are examined and those which do not meet specifications are designated as failed.
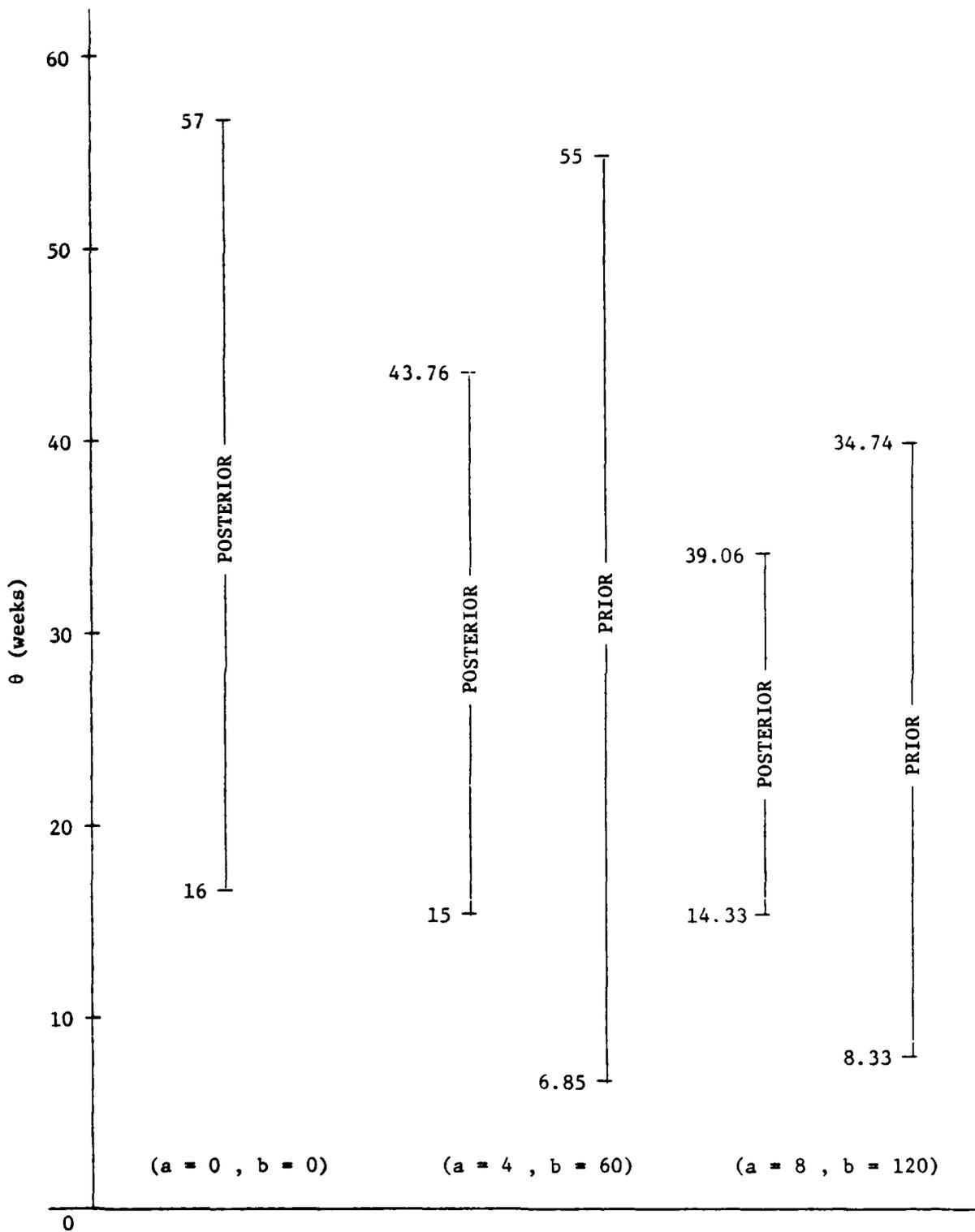
FIGURE 2.1

COMPARISON OF 95% CREDIBLE INTERVALS FOR $\tilde{\theta}$ (k = 10 , T = 275 weeks)

In Table 2.1, the results of testing $n = 840$ semiconductors are recorded for a 3 week period.

TABLE 2.1

SEMICONDUCTOR LIFE TEST DATA

Time Interval, hours	Observed Number of Failures
0 - 168	9
168 - 336	6
336 - 504	13

Let $N(t)$ be the number of failures in $[0,t]$ and $p \overset{def}{=} F(t)$, the probability of failure, in the interval. Then

$$(2.17) \qquad P[N(t) = k \mid p] = \binom{n}{k} p^k (1 - p)^{n-k} .$$

From Table 2.1 it is evident that failures are, relatively speaking, rare events. Thus, a Poisson distribution approximation to the binomial may be justified; i.e., $P[N(t) = k \mid p] \overset{\cdot}{=} \frac{[np]^k}{k!} e^{-np}$. Note also that the MLE for p in the first interval $[0,168]$ is $\hat{p} \overset{\cdot}{=} 0.01$. This suggests the approximation

$$p = 1 - \exp\left[- \int_0^t r(u)du\right] \overset{\cdot}{=} \int_0^t r(u)du .$$

Let $\lambda_1 = \frac{1}{t} \int_0^t r(u)du$, the *failure rate average* in the first time interval.

(If the failure rate is constant, then of course so also is the failure

rate average.) Using these approximations we obtain for the interval $(0, t]$;

$$L(\lambda \mid k , t) \doteq \frac{(n\lambda t)^k}{k!} e^{-n\lambda t} .$$

The natural conjugate prior is easily seen to be:

$$\pi(\lambda) = \frac{b^a \lambda^{a-1}}{\Gamma(a)} e^{-b\lambda} .$$

It follows that the posterior is given by:

$$\pi(\lambda \mid k , t) = \frac{(b + nt)^{a+k} \lambda^{a+k-1}}{\Gamma(a + k)} e^{-(b+nt)\lambda} .$$

For example, suppose we choose as our hyper-parameters for the prior distribution $a = 4$, and $\frac{a}{b} = 10^{-3}$ or $b = 4 \times 10^3$. Using Theorem 2.6, we can compute 95% credible limits on the failure rate average. For the first time interval of 168 hours, we have

$$P \left\{ 8.70 \times 10^{-5} = \frac{\chi^2_{.025}[2(a + k)]}{2(nt + b)} \leq \tilde{\lambda}_1 \leq \frac{\chi^2_{.975}[2(a + k)]}{2(nt + b)} = \right.$$
$$\left. 1.44 \times 10^{-4} \mid n , k , a , b \right\} = 0.95 .$$

If we judge failure rate parameters for distinct time intervals to be independent, then it follows that they are also independent with respect to the joint posterior distribution. Using this fact a joint posterior distribution for failure rates for all three time intervals can be easily calculated.

3. THE WEIBULL DISTRIBUTION

Whenever possible, the choice of a life distribution model should be based on the underlying failure mechanisms. Simple structures composed of statistically independent components have been used to derive life distribution models valid when the number of structural components is very large.

Suppose a structure of n components fails as soon as k components fail. For example, each of the strands whose lifetimes under stress are recorded in Table 1.1, Chapter 1, is composed of approximately 270 filaments bound together by an epoxy. If strand failure coincides with the kth failure of filaments, then structure lifetime $\zeta_{k,n}$ is the kth smallest of n lifetimes. If also component lifetimes are judged identically distributed and independent, then there are only two possible limiting structure life distributions in the sense that there exist sequences of normalizing constants $\{a_n\}_{n=1}^{\infty}$, $\{\lambda_n\}_{n=1}^{\infty}$ such that for all real x :

$$\lim_{n \to \infty} P\{\lambda_n(\zeta_{k,n} - a_n) \leq x\}$$

exists. The limit is either

$$(3.1) \qquad \frac{1}{(k-1)!} \int_0^{[\lambda(x-a)]^\alpha} e^{-u} u^{k-1} du , \quad \alpha , \lambda > 0 , x > a \geq 0 ,$$

or

$$(3.2) \qquad \frac{1}{(k-1)!} \int_0^{\exp[\lambda(x-a)]} e^{-u} u^{k-1} du , \quad -\infty < x < \infty$$

$$-\infty < a < \infty , \lambda > 0 .$$

[Smirnov (1952)]. In both cases a is a location parameter and λ is a scale parameter while α and k are shape parameters.

If k = 1 , then (3.1) becomes

(3.1') $W(x \mid a , \lambda , \alpha) = 1 - \exp \{-[\lambda(x - a)]^{\alpha}\}$, $x \geq a \geq 0$,

the *Weibull distribution*, and (3.2) becomes

(3.2') $\Lambda(x \mid a , \lambda) = 1 - \exp \{-e^{\lambda(x-a)}\}$, $-\infty < x < \infty$.

Thus, if X is the structure lifetime, then either X or exp (X) has a Weibull distribution. The failure rate for the Weibull distribution of (3.1') is

$$r_W(x) = \alpha\lambda^{\alpha}(x - a)^{\alpha-1} \quad \text{for} \quad x \geq a ,$$

and 0 elsewhere. In the second case it is

(3.4) $r_{\Lambda}(x) = \lambda \exp [\lambda(x - a)]$.

For all parameter values, (3.4) is increasing in x . Hence, if we wish to allow the possibility that the failure rate may be decreasing we must choose the Weibull model, (3.1'), with $\alpha < 1$.

The Weibull model appears to furnish an adequate fit for some strand lifetime data with estimated values of α less than 4. On the other hand, it has been empirically observed that for strength data, estimates for α using the Weibull model are often large (> 27 in some cases). This suggests that (3.2') may provide a better model for strand strength data.

Inference for the Weibull Distribution

The Weibull life distribution model has three parameters: a , λ , and α . The parameter a > 0 is a threshold value for lifetime; before time a we expect to see no failures. If there is no physical reason to justify a positive threshold value, the analyst should use the two parameter Weibull model. The most simple model compatible with prior knowledge concerning physical processes will often provide the most insight. The Weibull density is

$$(3.5) \qquad f(x \mid a , \alpha , \lambda) = \alpha\lambda^{\alpha}(x - a)^{\alpha-1}e^{-[\lambda(x-a)]^{\alpha}}$$

for $x \geq a$ and 0 elsewhere.

Usually we wish to quantify our uncertainty about a particular aspect of the life distribution, such as the probability of surviving x hours. For the three parameter Weibull model, this is given by:

$$(3.6) \qquad \bar{F}(x \mid a , \lambda , \alpha) = \exp \{-[\lambda(x - a)]^{\alpha}\} .$$

It is clearly sufficient to assess our uncertainty concerning a , λ , and α .

Suppose data are obtained under the General Sampling Plan (Section 1). Let $x_1, x_2, \ldots, x_k$ denote the unordered observed failure ages and n(u) the number surviving until age u . Then by Theorem 1.6 in Section 1, the likelihood is given by:

$$L(a , \alpha , \lambda \mid D)$$

$$\propto a^{k}\lambda^{k\alpha}\left[\prod_{i=1}^{k} (x_i - a)\right]^{\alpha-1} \exp \left\{-\lambda^{\alpha}\left[\int_{a}^{\infty} \alpha n(u)(u - a)^{\alpha-1}du\right]\right\}$$

for $a \leq x_i$ and α , $\lambda > 0$. Suppose there are m withdrawals and we pool observed failure and loss times and relabel them as:

$$0 \equiv t_{(0)} \leq t_{(1)} \leq t_{(2)} \leq \cdots \leq t_{(k+m)} \leq t .$$

Then, for $a \leq x_i$, $i = 1,2, \ldots, k$, we have

$$\int_a^\infty n(u)(u - a)^{\alpha-1} du$$

$$= \sum_{i=1}^{k+m} (n - i + 1) \int_{t_{(i-1)}}^{t_{(i)}} (u - a)^{\alpha-1} du + (n - k - m) \int_{t_{(k+m)}}^{t} (u - a)^{\alpha-1} du .$$

Observation is confined to the age interval $[0,t]$.

Two important deductions can be made from (3.7):

1. The only sufficient statistic for all three parameters (or for α and λ alone when $a = 0$) is the *entire* data set.

2. No natural conjugate family of priors is available for all three parameters (or for α and λ alone when $a = 0$). Consequently, the posterior distribution must be computed using numerical integration. [see Diaconis and Ylvisaker (1979)].

For most statistical investigations, a and perhaps also α would be considered nuisance parameters. By matching our joint prior density on a , λ and α with the likelihood (3.7), we can calculate the posterior density, $\pi(a , \lambda , \alpha \mid D)$. For example, if a is considered a nuisance parameter, then we would calculate the marginal

density on λ and α as:

$$\pi(\alpha , \lambda \mid D) = \int\limits_{0}^{\infty} \pi(a , \alpha , \lambda \mid D) da .$$

Credibility Regions for Two Parameter Models

Let $\pi(\alpha , \lambda \mid D)$ be the posterior density for a two parameter model such as the Weibull model above with scale parameter λ and shape parameter α . To find the so-called "highest posterior density" credibility region for α and λ simultaneously (Section 2), we find a constant $c(\beta)$ by sequential search such that:

(3.8) $$R = [(\alpha , \lambda) \mid \pi(\alpha , \lambda \mid D) \geq c(\beta)]$$

and

$$\iint\limits_{R} \pi(\alpha , \lambda \mid D) d\alpha d\lambda = \beta .$$

The region R defined above is a $\beta(100)$ percent credibility region for α and λ . For unimodal densities such regions are bounded by a single closed curve C which does not intersect itself (i.e., a "simply connected region").

To illustrate the use of Weibull credibility regions we have computed credibility regions corresponding to the data in Tables 3.1 and 3.2. Twenty-one pressure vessels were put on life test at 68% of their ultimate mean burst stress. A pressure vessel is filled with a gas or liquid and provides a source of mechanical energy. They are used on space satellites and other space vehicles. After 13,488 hours of testing, 5 failures were

TABLE 3.1

ORDERED FAILURE AGES OF PRESSURE
VESSELS LIFE TESTED AT 68% OF MEAN RUPTURE
STRENGTH (n = 21 , OBSERVATION TO 13,488 HOURS).

Number of Failure	Age at Failure (hours)
1	4000
2	5376
3	7320
4	8616
5	9120

TABLE 3.2

ORDERED FAILURE AGES OF PRESSURE VESSELS
LIFE TESTED AT 68% OF MEAN RUPTURE STRENGTH
(FAILURES BETWEEN 13,488 HOURS AND 20,568 HOURS)

Number of Failure	Age at Failure (hours)
1	14400
2	16104
3	20231
4	20233

recorded. After an additional 7080 hours of testing, an additional 4
failures were recorded.

Figure 3.1 displays credibility contours for α and λ after
13,488 hours of testing and again after 20,568 hours of testing. The
posterior densities were computed relative to uniform priors. The
posterior density computed after 20,568 hours could also be inter-
preted as the result of using the *posterior* (calculated on the basis of
Table 3.1 and a flat prior) as the *new prior* for the data in Table 3.2.
A qualitative measure of the information gained by an additional year
of testing can be deduced by comparing the initial (dark) contours
and the tighter (light) contours in Figure 3.1.

To predict pressure vessel life at the 68% stress level, we can
numerically compute

$$P[X > t \mid D] = \int_0^\infty \int_0^\infty e^{-(\lambda t)^\alpha} \pi(\alpha , \lambda \mid D) d\alpha d\lambda$$

where $\pi(\alpha , \lambda \mid D)$ must be numerically computed using the given data,
D .

If the mean life

$$\theta = \frac{\Gamma\left(1 + \frac{1}{\alpha}\right)}{\lambda}$$

or the standard deviation of life are of interest, their posterior
densities can be computed by making a change of variable and integrating
out the nuisance parameter. For example, if $a = 0$ in the Weibull
model and we are interested in the mean life, θ , we can use the
Weibull density in terms of α and θ
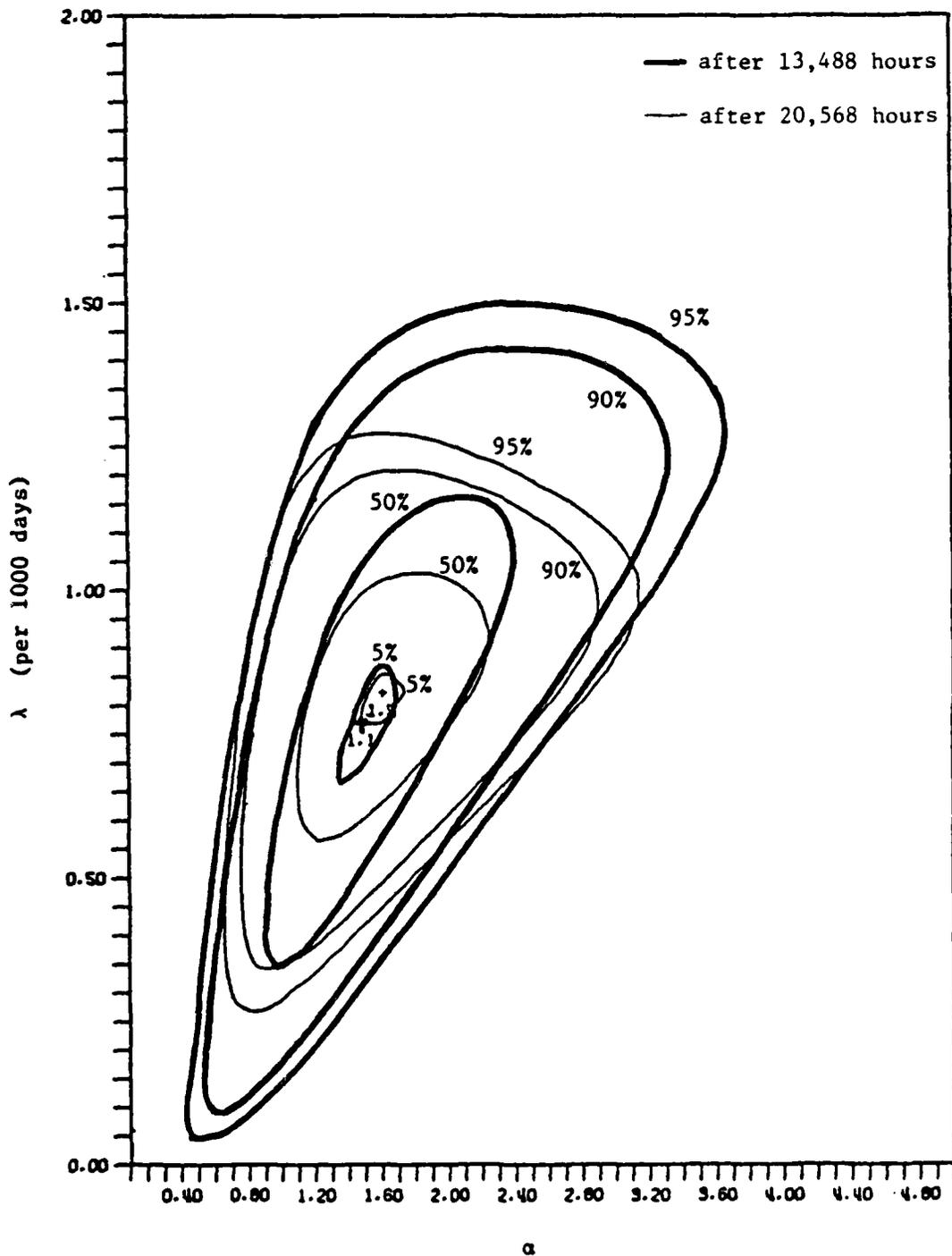
FIGURE 3.1

HIGHEST PROBABILITY DENSITY CONTOURS FOR α AND λ FOR
KEVLAR/EPOXY PRESSURE VESSEL LIFE TEST DATA. THE PRESSURE
VESSELS WERE TESTED AT 68% STRESS LEVEL.

$$f(x \mid \alpha, \theta) = \alpha \left[\frac{\Gamma\left(1 + \frac{1}{\alpha}\right)}{\theta} \right]^{\alpha} x^{\alpha-1} \exp \left\{ -\left[\frac{\Gamma\left(1 + \frac{1}{\alpha}\right)}{\theta} x \right]^{\alpha} \right\}$$

to compute the joint posterior density $\pi(\alpha, \theta \mid D)$. The prior for α and λ must be replaced by the induced prior for α and θ. This may be accomplished by a change of variable and by computing the appropriate Jacobian. The marginal posterior density of θ is then

$$\pi(\theta \mid D) = \int_0^\infty \pi(\alpha, \theta \mid D) d\alpha .$$

This can then be used to obtain credibility intervals on θ.

4. EXERCISES

1. Twenty-one pressure vessels are subjected to a static stress
    equivalent to 68% of the mean rupture stress. Testing begins
    June 22, 1977. Failures as of November 1, 1979 are recorded
    in Table 4.1 in hours.

    <div align="center">

    TABLE 4.1

    PRESSURE VESSEL FAILURE TIMES IN HOURS

    4000

    5376

    7320

    8616

    9120

    14400

    16104

    </div>

    Use a time transformed exponential model with $\alpha = 1.3$ (see
    Example 1.1). Plot the posterior density of λ^{α} relative
    to a uniform prior. Let $R_o(x) = x^{\alpha}$, what is the MLE?
    The mean life for this model is

    $$\theta = \Gamma(1 + 1/\alpha)/\lambda^{1/\alpha} .$$

    What is the mode for the corresponding posterior density for θ ?

2. The probability densities (1.12) and (1.13) were calculated
    for the case where unit i is observed until either it fails
    at age x_i or is withdrawn from observation at age ℓ_i . Suppose
    we have only *ordered* observations $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(k)}$
    and $\ell_{(1)} \leq \ell_{(2)} \leq \cdots \leq \ell_{(m)}$; i.e., we no longer know which
    unit fails at age $x_{(1)}$ or was withdrawn from observation at

age $\ell_{(i)}$. Recalculate the likelihood (1.12) and (1.13) for this case. [Hint: see Example 1.9.]

3. Suppose only the ordered failure ages $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(k)}$ and the ordered ages at withdrawal $\ell_{(1)} \leq \ell_{(2)} \leq \cdots \leq \ell_{(m)}$ are given. Furthermore, suppose withdrawal age W is random with withdrawal rate $\rho(\cdot)$, the analogue of failure rate. Show that the exact expression for the likelihood (1.16) is now

$$L(r(u) , \rho(u) , u \geq 0 \mid D)$$

$$= \prod_{t=1}^{m} n\left(\ell_{(t)}^{-}\right)\rho(\ell_{(t)}) \exp\left[-\int_{0}^{\infty} n(u)\rho(u)du\right]$$

$$\cdot \prod_{s=1}^{k} n\left(x_{(s)}^{-}\right)r(x_{(s)}) \exp\left[-\int_{0}^{\infty} n(u)r(u)du\right] .$$

where $n(u^{-})$ is the number surviving just prior to age u .

4. Let the failure rate function be

$$r(x) = \begin{cases} \lambda_1 , & 0 \leq x \leq t_1 \\ \lambda_2 , & x \geq t_1 . \end{cases}$$

Given $n(u)$, the number surviving until time $u \geq 0$ and observed failure times

$$x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(k)} ,$$

use (1.16) to compute the likelihood up to a constant factor independent of λ_1 and λ_2 . Assume noninformative stopping

rules are in effect and that t_1 is known.

5. Define the pth percentile, x_p , as the solution to

$$\int_0^{x_p} f(x \mid \theta)dx = p .$$

For the exponential model $f(x \mid \theta) = \frac{1}{\theta} \exp\left(-\frac{x}{\theta}\right)$ and the natural conjugate prior on θ , find the Bayes estimator for x_p corresponding to squared error loss. Assume that k failures and T total time on test have been observed. If, for example, $p = 0.01$ then x_p might qualify as the warranty period. The warranty would then be valid for $x_{.01}$ hours (months, years, etc.)

6. Let $x_p(\theta)$ be the pth percentile for the exponential model $f(x \mid \theta) = \frac{1}{\theta} \exp\left(-\frac{x}{\theta}\right)$ as described in Exercise 5. For the data in Exercise 7, Chapter 1, compute an upper 99% credible interval on x_p . Use the same natural conjugate prior specified in Exercise 7, Chapter 1.

7. Show that $\underset{a}{\text{minimum }} E_\pi[|\tilde{\theta} - a|]$ is attained when a satisfies

$$\int_{-\infty}^{a} \pi(\theta)d\theta = \frac{1}{2} ;$$

i.e., a is the median for π . [The same result holds for the posterior density. See (2.6)].

8. Using the data in Table 2.1, calculate 95% credibility intervals on

$$\lambda_1 = \frac{1}{168} \int\limits_0^{168} r(u)\,du \ ,$$

$$\lambda_2 = \frac{1}{168} \int\limits_{168}^{336} r(u)\,du \ ,$$

and

$$\lambda_3 = \frac{1}{168} \int\limits_{336}^{504} r(u)\,du \ .$$

Use natural conjugate priors with $a = 4$ and $b = 4 \times 10^3$ for both λ_1, λ_2 and λ_3. Suppose we judge them independent. Initially, 840 semiconductors were put on life test.

10. Let $F(x \mid \alpha , \lambda) = 1 - \exp [-(\lambda x)^\alpha]$. Assume α is known and let $\gamma = \lambda^\alpha$. Let

$$\pi(\gamma) = \frac{b^a \gamma^{a-1} e^{-b\gamma}}{\Gamma(a)} \ .$$

Calculate the Bayes estimator with respect to squared error loss of the probability of survival until age x given k failures and $n(u)$, $n \geq 0$ where $n(u)$ is the number of units surviving until age u.

11. Assume the exponential model $\lambda e^{-\lambda x}$ and suppose k failures and T total time on test have been observed. Using the natural conjugate prior $\pi(\lambda) = \frac{b^a \lambda^{a-1} e^{-b\lambda}}{\Gamma(a)}$, compute the posterior. Now suppose that the resulting posterior is used as a prior with

48

the *same* data, i.e., no additional data are collected to compute a new posterior. Furthermore, suppose that this process is repeated infinitely often with the *same* data.

(a) What will be our final conclusion?

(b) What is the fallacy in this inference procedure?

12. Suppose n units are put on life test. The stopping rule is as follows. If the first failure occurs before time t_1, stop at the time of first failure. Otherwise continue testing until time $t_2 > t_1$. Compute the likelihood for the exponential model. Is the stopping rule noninformative? Why?

13. Consider the following life test for $n + 1$ units, each with life distribution $F(\cdot \mid \theta)$. A unit is selected at random and life tested to failure, say at time x. This failure time is then used to provide the stopping rule: test n units to time x. Is this stopping rule informative relative to θ?

5. NOTES AND REFERENCES

Section 1

In the General Sampling Plan we needed to assume that any stopping rules used were noninformative concerning the failure distribution. The need for this assumption was pointed out by Raiffa and Schlaiffer (1961). Examples of informative stopping rules were given by Roberts (1967) in the context of two stage sampling of biological populations to estimate population size (so-called capture-recapture sampling).

Section 2: Unbiasedness

The posterior mean is a Bayes estimator of a parameter, say θ, with respect to squared error loss. It is also a function of the data. An estimator, $\hat{\theta}(D)$, is called *unbiased* in the sample theory sense if

$$E_F[\hat{\theta}(D) \mid \theta] = \theta$$

for each $\theta \in \Theta$. No Bayes estimator (based on a corresponding proper prior) can be unbiased in the sample theory sense [Bickel and Blackwell (1967)].

Most unbiased estimators are in fact inadmissible in the sample theory sense with respect to squared error loss. For example, $\hat{\theta}(D) = \frac{T}{k}$ is a sample theory unbiased estimator for the mean of the density $\frac{1}{\theta} e^{-\frac{x}{\theta}}$ under sample plan (a) of Chapter 1. However it is inadmissible in the sense that there exists another $c\hat{\theta}(D)$ with $c \neq 1$ such that, for all θ

$$E_F[[c\hat{\theta}(D) - \theta]^2 \mid \theta] < E_F[[\hat{\theta}(D) - \theta]^2 \mid \theta] .$$

50

To find this c , consider $Y = \dfrac{\hat{\theta}(D)}{\theta}$ and note $EY = 1$. Then we
need only find c such that

$$E_F[(cY - 1)^2 \mid \theta]$$

is minimum. This occurs for $c_0 = \dfrac{EY}{EY^2}$ which is clearly *not* 1. Hence
$\hat{\theta}(D)$ is sample theory inadmissible. Sample theory unbiasedness is
not a viable criterion.

For large k , $\hat{\theta}(D) = \dfrac{T}{k}$ will be approximately the same as our Bayes
estimator. However, $\dfrac{T}{k}$ is not recommended for small k .

Since tables of the chi-square distribution have in the past been
more accessible than tables of the gamma distribution, we have given
the chi-square special treatment. However with modern computing facili-
ties, we really only need to use the more general gamma distribution.

Confidence Intervals

A $(1 - \alpha)100\%$ confidence interval in the sample theory sense is one
such that if the experiment is repeated infinitely often (and the interval
recomputed each time) then $(1 - \alpha)100\%$ of the time the interval will
cover the fixed unknown true parameter θ . Since confidence intervals
do not produce a probability distribution on the parameter space for θ ,
they cannot provide the basis for action in the decision theory sense;
i.e., a decision maker cannot use a sample theory confidence interval to
compute an expected utility function which can then be maximized over his
set of possible decisions.

If for $\lambda e^{-\lambda x}$ we choose the *improper* prior, $\pi(\lambda) = \dfrac{1}{\lambda}$, then the
chi-square $(1 - \alpha)100\%$ credible intervals and the sample theory

$(1 - \alpha)100\%$ confidence intervals agree. Unfortunately, such improper credible intervals can be shown to violate certain rules of logical behavior. Lindley (personal communication) provides the following simple illustration of this fact for the exponential model $\lambda e^{-\lambda x}$. Suppose n units are put on test and we stop at the first failure, so that $T = nX_{(1)}$. Now T given λ also has density $\lambda e^{-\lambda x}$ so that $\frac{(\ln 2)}{T}$ is a 50% improper upper credible limit on λ; i.e.,

$$(5.1) \qquad P\left[\tilde{\lambda} \leq \frac{(\ln 2)}{T} \mid T, \pi(\lambda) = \frac{1}{\lambda}\right] = 0.50 .$$

Suppose now that T is observed and we accept the probability statement (5.1). Consider the following hypothetical bet.

(i) If $\tilde{\lambda} < \frac{(\ln 2)}{T}$ we lose the amount e^{-T};

(ii) If $\tilde{\lambda} \geq \frac{(\ln 2)}{T}$ we win e^{-T}.

We can then pretend that the true λ is somehow revealed and bets are paid off. If we believe statement (5.1), then given T such a bet is certainly fair.

Now let us compute our expected gain before T is observed (preposterior analysis). This is easily seen to be (conditional on λ)

$$- \int_0^{\frac{\ln 2}{\lambda}} \lambda e^{-\lambda t} e^{-t} dt + \int_{\frac{\ln 2}{\lambda}}^{\infty} \lambda e^{-\lambda t} e^{-t} dt = \frac{\lambda}{1 + \lambda}\left[2^{-\frac{1}{\lambda}} - 1\right]$$

which is negative for all $\lambda > 0$. Note that this is what we *subjectively*

52

expect, since as (improper) Bayesians, every probability (and presumably even an improper prior) is subjective.

The contradiction lies in the observation that

1. Conditional on λ and prior to observing T , our expected winnings are negative for all λ ;

2. Conditional on T , our expected loss is zero $\left(\text{using the improper prior } \pi(\lambda) = \frac{1}{\lambda}\right)$.

The source of the contradiction is that we have not measured our uncertainty for all events by probability. For example, we have assigned the value ∞ to the event $\bar{\lambda} < \lambda_o$ for all $\lambda_o > 0$; i.e.,

$$\int_0^{\lambda_o} \pi(\lambda)d\lambda = \int_0^{\lambda_o} \frac{1}{\lambda} d\lambda = \infty .$$ See Appendix B for a proof that for any set of uncertainty statements that are not probabilistically based (relative to *proper* distributions), a system of bets can be constructed which will result in the certain loss of money. A bet consists of paying $pz < z$ dollars to participate with the understanding that if an event E occurs you win z dollars and otherwise you win nothing.

Section 3

The Weibull distribution is one of several extreme value distributions. See Barlow and Proschan (1975), Chapter 8, for a more advanced discussion of extreme value distributions.

REFERENCES

[1] Barlow, R. E. and F. Proschan, STATISTICAL THEORY OF RELIABILITY
     AND LIFE TESTING, Holt, Rinehart and Winston, Inc., New York,
     (1975).

[2] Barlow, R. E. and F. Proschan, "Inference for the Exponential Life
     Distribution," (Chapter I), ORC 79-16, (1979).

[3] Bickel, P. J. and D. Blackwell, "A Note on Bayes Estimates," Annals
     of Mathematical Statistics, Vol. 38, pp. 1907-1911, (1967).

[4] Box, G. E. P. and T. C. Tiao, BAYESIAN INFERENCE IN STATISTICAL
     ANALYSIS, Addison-Wesley Publishing Company, Reading, Massachusetts,
     (1973).

[5] De Groot, M. H., OPTIMAL STATISTICAL DECISIONS, McGraw-Hill, New York,
     (1970).

[6] Diaconis, R. and D. Ylvisaker, "Conjugate Priors for Exponential
     Families," Annals of Statistics, Vol. 7, pp. 269-281, (1979).

[7] Edwards, W., H. Lindman and L. J. Savage, "Bayesian Statistical
     Inference for Psychological Research," Psychological Review,
     Vol. 70, pp. 193-242, (1963).

[8] Hoel, P. G., S. C. Port and C. J. Stone, INTRODUCTION TO PROBABILITY
     THEORY, Houghton Mifflin Company, Boston, Massachusetts, (1971).

[9] Lindley, D. V., "The Bayesian Approach," Scandinavian Journal of
     Statistics, Vol. 5, pp. 1-26, (1978).

[10] Pearson, E. S. and H. O. Hartley, BIOMETRIKA TABLES FOR STATISTICIANS,
     Vol. 1, The University Press, Cambridge, England, (1958).

[11] Raiffa, H. and R. Schlaiffer, APPLIED STATISTICAL DECISION THEORY,
     Harvard Business School, Boston, Massachusetts, (1961).

[12] Roberts, H. V., "Informative Stopping Rules and Inference about
     Population Size," Journal of American Statistical Association,
     Vol. 62, pp. 763-775, (1967).

[13] Smirnov, N. V., "Limit Distributions for the Terms of a Variational
     Series," Transaction of the Mathematical Society Series, Vol. 1,
     pp. 1-64.

N

DAT