David S. Newman

BOEING SCIENTIFIC RESEARCH LABORATORIES DOCUMENT D1-82-0660

BOEING

SCIENTIFIC RESEARCH LABORATORIES

Maximum Likelihood Estimation of a U-Shaped Failure Rate Function

T. A. Bray

Gordon B. Crawford

Frank Proschan

MATHEMATICS RESEARCH OCTOBER 1967

D1-82-0660

MAXIMUM LIKELIHOOD ESTIMATION OF A U-SHAPED FAILURE RATE FUNCTION

by

T. A. Bray

Gordon B. Crawford

and

Frank Proschan

Mathematical Note No. 534

Mathematics Research Laboratory

BOEING SCIENTIFIC RESEARCH LABORATORIES

October 1967

Table of Contents

Abstract

A maximum likelihood estimate (MLE) is derived for the failure rate as a function of age based on incomplete data, assuming the failure rate function is initially decreasing and subsequently increasing, with the turning point unknown. An algorithm for computation of the MLE is developed; a small worked example is included to show in detail the steps of the algorithm. A program for machine computation is presented. The MLE is shown to be a consistent estimator.

No further assumption is required as to the mathematical form of the life-length distribution (such as exponential, Wiebull, normal, etc.). Thus the model may be realistically applied in a variety of reliability situations, where the so-called "bathtub shaped" failure rate is appropriate. In the present paper it is applied to the analysis of airplane part failure data. A real life large-scale example is presented in which the failure rate of a constant speed drive unit of a jet airplane is estimated.

Maximum Likelihood Estimation of a U-Shaped Failure Rate Function

1. Introduction

An important problem in reliability analysis is to estimate failure rate as a function of age from a set of observed failure times. Given a failure distribution F with density f, we define the failure rate at any time t for which $F(t) \neq 1$ by

$$(1) \qquad r(t) = f(t)/\overline{F}(t),$$

where $\overline{F}(t) \equiv 1-F(t)$. A knowledge of the failure rate function of components of a system is desirable in predicting system reliability, setting up spare part kits, developing optimal maintenance policies, constructing sound warranty policies, and determining "burn-in" periods for components subject to "infant mortality".

In developing an estimator of the failure rate function it is, of course, desirable to use as much a priori information as is available concerning the *form* of the distribution or its failure rate. For example, from physical or engineering considerations it may be known that any or all of the following phenomena may occur:

(1) The item may be subject to infant mortality; in this case the failure rate is initially high and decreases during the *infant mortality phase*. This might correspond to the presence of defective units in the population which fail early.

(2) Following the infant mortality phase (if it exists), for a relatively long period the item is subject only to chance failures (perhaps as a result of randomly occurring environmental stresses) and thus displays a constant failure rate. This period is known as the *useful life phase.*

(3) Finally the *wearout phase* commences; the item now displays a failure rate increasing with age, reflecting its increasing vulnerability to failure. This may be the result of actual wearing down of material in the unit (e.g., the rubber in a tire), of drift of important parameters from their nominal values (e.g., electrical characteristics of a tube), of deterioration in the functioning of components (e.g., human organs), etc.

Airplane operating data represents a good example of a case in which for most parts some or all of the three phases of the failure rate function are present, but it is not known in advance which phases (if any) are absent.

In this paper we develop the maximum likelihood estimator (MLE) for the failure rate function assuming that the failure rate is initially decreasing and then increasing; we call this a *U-shaped failure rate function.* Note that failure rate functions possessing one or more of the three phases described above *are* initially decreasing and then increasing (neither term is used in the strict sense in this paper), so that if we use the method developed in this paper we obtain an MLE in each case, assuming, of course, that we do not definitely know *a priori* that any given phase is *not* present. A program for machine calculation of the MLE has been developed and is presented in Section 6. A small-scale, simplified problem is solved to show in detail the successive steps involved (Section 3). Mathematical derivation of the MLE is presented in Section 4. Consistency of the MLE is demonstrated in Section 5. The present paper is an expansion of Bray-Crawford-Proschan (1967).

One advantage of the method presented for estimating the failure rate function is that it is not necessary to make a strong, non-verifiable assumption that the form of the distribution is exponential, Weibull, gamma, normal, etc.

2. The Maximum Likelihood Estimate of the Failure Rate Function

Assume that each of n units is observed over some or all of its life. Thus unit i is observed from age $a_i \geq 0$ until it either fails at a random age $Y_i \geq a_i$ or has attained age $b_i \geq a_i$, whichever occurs first $(i=1,\ldots,n)$. We now give a procedure for obtaining a MLE of the failure rate $r(t)$, assuming that r is decreasing on $[0,t_0]$ and increasing on (t_0,∞), where $t_0 (0 \leq t_0 \leq \infty)$ is unknown. To avoid complications, we do not consider the case in which a failure occurs at age 0. Note that $t_0 = 0$ corresponds to the case of increasing failure rate (IFR), while $t_0 = \infty$ corresponds to the case of decreasing failure rate (DFR). In reading the steps of the procedure, it is helpful to see how they are applied in the example of Section 3.

Procedure for Obtaining the MLE

(1) Suppose k $(1 \leq k \leq n)$ failures are actually observed. Let $X_1 \leq X_2 \leq \cdots \leq X_k$ denote the ordered ages at failure; for convenience, define $X_0 \equiv 0$, $X_{k+1} \equiv \max\{b_1,\ldots,b_n,Y_1,\ldots,Y_n\}$ (i.e., the oldest age observed). Let T_i denote the total time observed during (X_{i-1},X_i)

for $i=1,2,\ldots,k+1$; T_i will be referred to as the "exposure time" during the i^{th} interval.

(2) We obtain the MLE $\hat{r}_j(t)$ for the failure rate $r(t)$ on $[0,t_0]$ assuming $r(t)$ decreasing on $[0,t_0]$, where $X_{j-1} \leq t_0 < X_j$, $j=1,\ldots,k+1$; we use only the observations in $[0,t_0]$. (See Marshall and Proschan, 1965.)

(a) Compute T_i^{-1} as the *unconstrained* MLE of the failure rate on (X_{i-1},X_i), assuming it constant on that interval, for $i=1,\ldots,j-1$.

(b) If $T_1^{-1} \geq T_2^{-1} \geq \cdots \geq T_{j-1}^{-1}$, then $\hat{r}_j(t) = T_i^{-1}$ for $X_{i-1} < t \leq X_i$ $(i=1,\ldots,j-1)$ is the MLE on $[0,X_{j-1}]$.

(c) If, on the other hand, a reversal occurs, say $T_i^{-1} < T_{i+1}^{-1}$, then replace each by $[\frac{1}{2}(T_i+T_{i+1})]^{-1}$.

(d) If the new sequence is properly ordered, i.e., $T_1^{-1} \geq \cdots \geq T_{i-1}^{-1} \geq [\frac{1}{2}(T_i+T_{i+1})]^{-1} = [\frac{1}{2}(T_i+T_{i+1})]^{-1} \geq T_{i+1}^{-1} \geq \cdots \geq T_{j-1}^{-1}$, then these represent the failure rates on the successive intervals $(X_{i-1},X_i]$, $i=1,\ldots,j-1$.

(e) If a reversal exists, then replace by appropriate averages. Thus if $[\frac{1}{2}(T_i+T_{i+1})]^{-1} = [\frac{1}{2}(T_i+T_{i+1})]^{-1} < T_{i+2}$, replace each of the three by the average $[\frac{1}{3}(T_i+T_{i+1}+T_{i+2})]^{-1}$. Note that the average replacing reversed failure rates is obtained by taking the number of failures divided by the total exposure time during the interval.

(f) Continue averaging until all reversals are eliminated. The resulting sequence constitutes the MLE $\hat{r}_j(t)$ of the failure rate on the successive intervals $(0,X_1],\ldots,(X_{j-2},X_{j-1}]$.

(g) On$(X_{j-1},t_0]$, the MLE $\hat{r}_j(t) \equiv 0$.

(3) Next obtain the MLE for $r(t)$ on (t_0,∞), assuming it increasing on (t_0,∞), where, as in (2), $X_{j-1} \leq t_0 < X_j$; we use only the observations in (t_0,∞). The procedure is similar to that described in (2) just above, with a few small differences.

(a) Compute T_i^{-1} as the unconstrained MLE of the failure rate on $[X_{i-1},X_i)$, assuming it constant on the interval, for $i=j+1,\ldots,k$. If $X_{k+1} > X_k$, the unconstrained MLE of the failure rate on $[X_k,X_{k+1})$ is $\equiv 0$ (since *no* failures were observed in (X_k,X_{k+1}).)

(b) If $X_{k+1} > X_k$, then a reversal automatically exists since $T_k^{-1} > 0$. Averaging to eliminate the reversal yields $(T_k+T_{k+1})^{-1}$ as the estimate of $r(t)$ on the interval $[X_{k-1},X_{k+1})$. (During that interval *one failure* occurred in a *total exposure time* of $T_k + T_{k+1}$.)

(c) We proceed as in (2) above, averaging to eliminate all reversals. In the present case a reversal occurs if any failure rate is strictly *larger* than the succeeding failure rate. In averaging, we take as the common estimate of the failure rate for each of the intervals entering into the reversal, the number of failures in those intervals divided by the total exposure time in those intervals.

(d) When all reversals have been eliminated, the resulting estimate of failure rate on $[X_j,X_{k+1})$ is the MLE, $\hat{r}_j(t)$.

(e) On(t_0,X_j), $\hat{r}_j(t) \equiv 0$.

(4) Compute the likelihood, $\hat{L}_j$, corresponding to $\hat{r}_j(t)$:

(a) If $X_{k+1} > X_k$, then

$$\hat{L}_j = \left\{\prod_{i=1}^{j-1} \hat{r}_j(X_i)e^{-T_i\hat{r}_j(X_i)}\right\}\hat{r}_j(X_j)\prod_{i=j+1}^{k} \hat{r}_j(X_i)e^{-T_i\hat{r}_j(X_{i-1})}e^{-T_{k+1}\hat{r}_j(X_k)}$$

(2.1) $j=1,\ldots,k$

$$\hat{L}_{k+1} = \prod_{i=1}^{k} \hat{r}_j(X_i)e^{-T_i\hat{r}_j(X_i)} \quad \text{(since } \hat{r}_j(t) = 0 \text{ for } t > X_k\text{)}.$$

(b) If $X_{k+1} = X_k$, then

$$\hat{L}_j = \left\{\prod_{i=1}^{j-1} \hat{r}_j(X_i)e^{-T_i\hat{r}_j(X_i)}\right\}\hat{r}_j(X_j)\prod_{i=j+1}^{k-1} \hat{r}_j(X_i)e^{-T_i\hat{r}_j(X_{i-1})}e^{-T_k\hat{r}_j(X_{k-1})}$$

(2.2) $j=1,\ldots,k-1.$

(5) Suppose $\hat{L}_{j_0} = \max\left\{\hat{L}_1,\ldots,\hat{L}_{k+1}\right\}$. Then $\hat{r}_{j_0}(t)$ is the MLE of $r(t)$. From the MLE of failure rate, we can compute the MLE of the density and the MLE of the distribution function, using the well-known relationships:

(2.3) $$f(t) = r(t)e^{-\int_0^t r(u)du},$$

(2.4) $$F(t) = 1-e^{-\int_0^t r(u)du}.$$

3. Examples

3.1 Small-Scale Worked Example

First we present a small illustrative example to show in detail the steps of the calculation of the MLE of the failure rate. The failure times shown are deliberately selected to be small integers so as to simplify the exposition. The algorithm can be used however for *any* set of data.

Operating data are collected on a fleet of airplanes over a fixed period of time. When a part fails, its age (measured in *operating* units of time) at failure is recorded, and the failed part is replaced by a functioning part. The failed part is repaired and becomes available for replacement. A part may be retired or overhauled (so that it becomes as good as new) when it reaches a certain age. Given a collection of data on a given part type (See Fig. 1), obtain the MLE of the failure rate $r(t)$, assuming $r(t)$ initially decreasing and then increasing, with the turning point t_0 assumed unknown.

In Fig. 1, the crosses indicate the observed times of failure. Observation on unit 1 begins when its age is 1 unit of time. Unit 1 experiences failure at age 4. After repair, unit 1 experiences a second failure after 2 more hours of operation, i.e., at age 6. After repair, unit 1 experiences a third failure after 1 more hour of operation, i.e., at age 7. After repair, unit 1 experiences failure requiring overhaul after 1 additional hour of operation, i.e., at age 8. At this point unit 1, as such, disappears.

After unit 1 is overhauled, its further history is recorded under the designation 1' on the second line. Observation begins at age 0, since the unit is now as good as new. Failures of unit 1' occur at ages 5,7, and 9; at age 10 the unit is overhauled. No additional hours are logged on the unit by the time the study ends.

Observation on unit 2 begins at age 0. Successive failures occur at ages 1, 2, 3, 4. At age 12 overhaul occurs.

Observation on unit 3 begins at age 5 and is discontinued at age 7. No failures are observed.

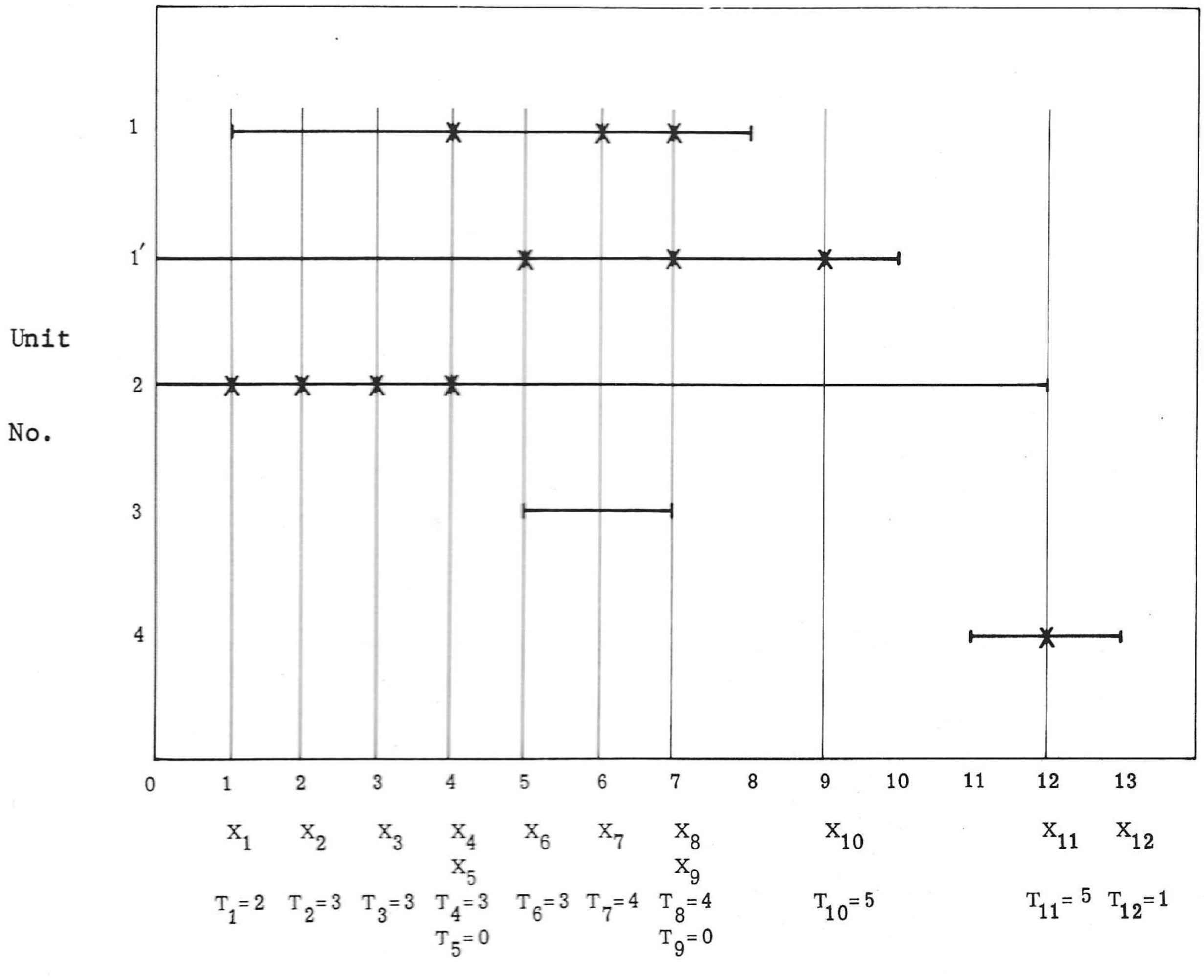

Fig. 1. Operating Data for Airplane Part

X_i = age at i^{th} failure, $i = 1, \ldots, 11$

T_i = total exposure time during $[X_{i-1}, X_i)$

Observation on unit 4 begins at age 11. Failure occurs at age 12. Overhaul occurs at age 13.

Application of Procedure for Obtaining MLE

(1) The ordered actual failure times occur at $X_1 = 1$, $X_2 = 2$, $X_3 = 3$, $X_4 = 4 = X_5$, $X_6 = 5$, $X_7 = 6$, $X_8 = 7 = X_9$, $X_{10} = 9$, and $X_{11} = 12$. In addition, the last observed time point is $X_{12} = 13$ (no failure occurred at this time point). The total exposure time T_1 in $[0,1)$ is 2, resulting from 1 hour observed on unit 1' and 1 hour observed on unit 2. Similarly, we obtain $T_2 = 3$, $T_3 = 3$, $T_4 = 3$, $T_5 = 0$ (since $X_4 = X_5$), $T_6 = 3$, $T_7 = 4$, $T_8 = 4$, $T_9 = 0$, $T_{10} = 5$, $T_{11} = 5$, and $T_{12} = 1$.

(2) We obtain the MLE $\hat{r}_j(t)$ for the failure rate $r(t)$ assuming $r(t)$ decreasing on $[0,t_0]$, where $X_{j-1} \leq t_0 < X_j$, $j=1,2,...,k+1$; we use only the observations on $[0,t_0]$.

(a) Start with $j = 1$. Then on $[0,t_0]$, $\hat{r}_1(t) = 0$. (No failures have been observed in $[0,t_0]$.)

(b) For $j = 2$, $\hat{r}_2(t) = T_1^{-1} = \frac{1}{2}$ on $[0,1]$ and $\hat{r}_2(t) = 0$ on $(1,t_0]$.

(c) For $j = 3$, the initial unconstrained estimate for $\hat{r}_3(t)$ is $T_1^{-1} = \frac{1}{2}$ on $[0,1]$, is $T_2^{-1} = \frac{1}{3}$ on $(1,2]$, and is 0 on $(2,t_0]$. Since $\frac{1}{2} \geq \frac{1}{3} \geq 0$, no reversals exist, and thus the initial unconstrained estimates constitute the MLE on $[0,t_0]$.

(d) We continue in this fashion obtaining $\hat{r}_4(t)$ and $\hat{r}_5(t)$, with no reversals occurring in $[0,t_0]$.

(e) A reversal occurs in computing $\hat{r}_6(t)$. The initial estimates of failure rate on the successive intervals are $\frac{1}{2}, \frac{1}{3}, \frac{1}{3}, \frac{1}{3}, \infty, 0$. Since $T_4^{-1} = \frac{1}{3} < \infty = T_5^{-1}$, we pool, obtaining as the new estimate of failure rate on $(X_3,X_5]$, $2(T_4+T_5)^{-1} = \frac{2}{3}$. The new sequence of trial estimates of the failure rate still exhibits a reversal: $\frac{1}{2}, \frac{1}{3}, \frac{1}{3}, \frac{2}{3}, \frac{2}{3}, 0$. We pool again to obtain a new trial estimate of failure rate on $(X_2,X_5]$, $3(T_3+T_4+T_5)^{-1} = \frac{3}{6}$. The new sequence of trial estimates still exhibits a reversal: $\frac{1}{2}, \frac{1}{3}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, 0$. Pooling again, we obtain $4(T_2+T_3+T_4+T_5)^{-1} = \frac{4}{9}$ as an estimate of the failure rate on $(X_1,X_5]$. The new sequence of failure rate estimates exhibits no reversals: $\frac{1}{2}, \frac{4}{9}, \frac{4}{9}, \frac{4}{9}, \frac{4}{9}, 0$. We conclude that $\hat{r}_6(t) = \frac{1}{2}$ on $[0,1]$, $\hat{r}_6(t) = \frac{4}{9}$ on $(1,4]$, and $\hat{r}_6(t) = 0$ on $(4,t_0]$.

(f) In a similar fashion, we obtain $\hat{r}_j(t)$ on $[0,t_0]$ for $j = 7, 8, 9, 10$.

(3) Next obtain the MLE $\hat{r}_j(t)$ for $r(t)$ on (t_0,∞), assuming it increasing on (t_0,∞), where $X_{j-1} \leq t_0 < X_j$, $j=1,2,\ldots,k+1$; we use only the observations in (t_0,∞).

(a) Start with $j = 1$. On $[t_0,1)$, $\hat{r}_1(t) = 0$. The initial trial estimate for $\hat{r}_1(t)$ on $[1,2)$ is $T_2^{-1} = \frac{1}{3}$, on $[2,3)$ is $T_3^{-1} = \frac{1}{3}$, on $[3,4)$ is $T_4^{-1} = \frac{1}{3}$, etc. Next we examine these trial estimates on successive intervals to see if a reversal exists: $\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, \infty, \frac{1}{3}, \frac{1}{4}, \frac{1}{4}, \infty, \frac{1}{5}, \frac{1}{5}, 0$. (The estimate 0 corresponds to no failures observed in an exposure time of 1 in the final interval $[12,13]$.)

To eliminate the reversal $\infty > \frac{1}{3}$, we pool to obtain the new estimate $2(T_4+T_5)^{-1} = \frac{2}{3}$ on [3,4).

The new sequence of failure rate estimates, $\frac{1}{3}, \frac{1}{3}, \frac{2}{3}, \frac{2}{3}, \frac{1}{3}, \frac{1}{4}, \frac{1}{4}, \infty, \frac{1}{5}, \frac{1}{5}, 0,$ still contains reversals. We continue to pool to eliminate reversals until we finally obtain the single failure rate estimate $\hat{r}_1(t) = \frac{10}{31}$ over the entire interval [1,13].

(b) Next suppose $j = 2$. Then $\hat{r}_2(t) = 0$ on $[t_0, 2)$. Proceeding as in (a) we obtain the single failure estimate $\hat{r}_2(t) = \frac{9}{28}$ on [2,13].

(c) In a similar fashion we obtain $\hat{r}_3(t) = 0$ on $[t_0, 3)$. Again, we are required to pool on the entire remaining interval to obtain the single failure estimate $\hat{r}_3(t) = \frac{8}{25}$ on [3,13].

The results of steps (2) and (3) are summarized in Table 1.

$\hat{r}_j(t)$

Interval† assumed to contain t_0

j	0,1	1,2	2,3	3,4	4,5	5,6	6,7	7,9	9,12	12,13
1	0	.3226	.3226	.3226	.3226	.3226	.3226	.3226	.3226	.3226
2	.5000	0	.3214	.3214	.3214	.3214	.3214	.3214	.3214	.3214
3	.5000	.3333	0	.3200	.3200	.3200	.3200	.3200	.3200	.3200
4	.5000	.3333	.3333	0	.3182	.3182	.3182	.3182	.3182	.3182
5	.5000	.4444	.4444	.4444	0	.2500	.2500	.2727	.2727	.2727
6	.5000	.4444	.4444	.4444	.3333	0	.2500	.2727	.2727	.2727
7	.5000	.4444	.4444	.4444	.3333	.2500	0	.2727	.2727	.2727
8	.5000	.4444	.4444	.4444	.3636	.3636	.3636	0	.1667	.1667
9	.5000	.4444	.4444	.4444	.3636	.3636	.3636	.2000	0	1.0000
10	.5555	.4444	.4444	.4444	.3636	.3636	.3636	.2000	.2000	0

Table 1

†The i^{th}-interval is open on the left and closed on the right for $i = 1, \ldots, j-1$, and is closed on the left and open on the right for $i = j, \ldots, 10$.

(4) Next we compute the likelihood L_j corresponding to $\hat{r}_j(t)$ for $j = 1, \ldots, 10$. Since some exposure time has occurred beyond the last observed failure, we use formula (2.1). We display the natural logarithm of each likelihood in Table 2.

j	$\ln L_j$	j	$\ln L_j$	j	$\ln L_j$
1	-.2131 + 2	5	-.2061 + 2	9	-.2059 + 2
2	-.2091 + 2	6	-.2032 + 2	10	-.1959 + 2
3	-.2091 + 2	7	-.2032 + 2		
4	-.2091 + 2	8	-.1978 + 2		

Table 2

Note that L_{10} is the largest. Thus the MLE of the failure rate is given by $\hat{r}_{10}(t)$ in Table 1.

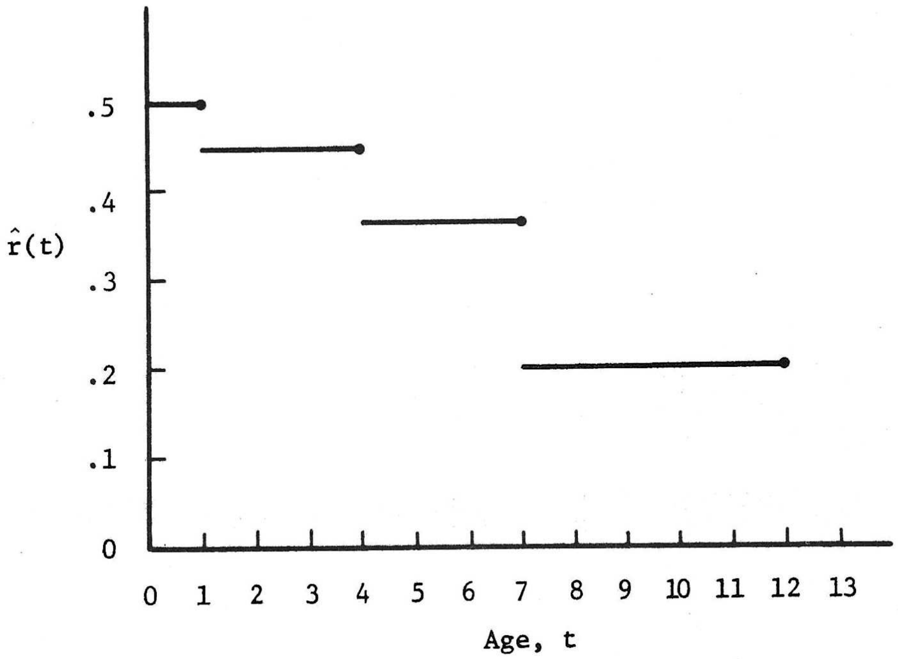

Fig. 2. Maximum Likelihood Estimate of the Failure Rate Function

3.2 Real Life Application

The maximum likelihood estimate of the failure rate of a constant speed drive of a jet airplane was obtained, assuming the failure rate initially decreasing and then increasing. The data was obtained over a one year period (1964), and contains 141 observed failures. Because of the large number of observations, we do not present the raw data, but show the maximum likelihood estimate $\hat{r}(t)$ of the failure rate in Fig. 3, the maximum likelihood estimate $\hat{\bar{F}}(t)$ of the survival probability in Fig. 4, and the maximum likelihood estimate, $\hat{F}_{50}(t) = \dfrac{\hat{F}(t+50) - \hat{F}(t)}{\hat{\bar{F}}(t)}$,

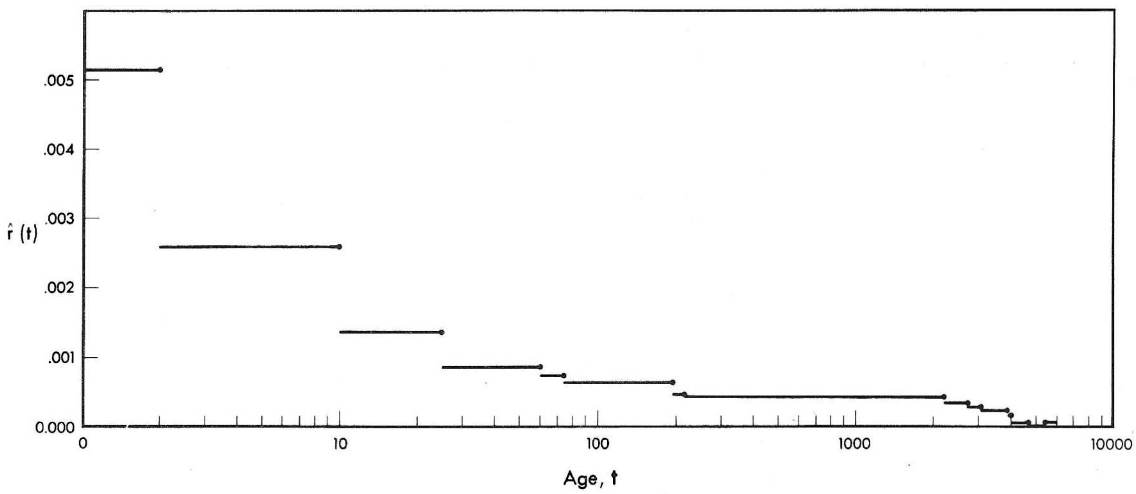

Fig. 3. MLE of $\hat{r}(t)$, the failure rate function

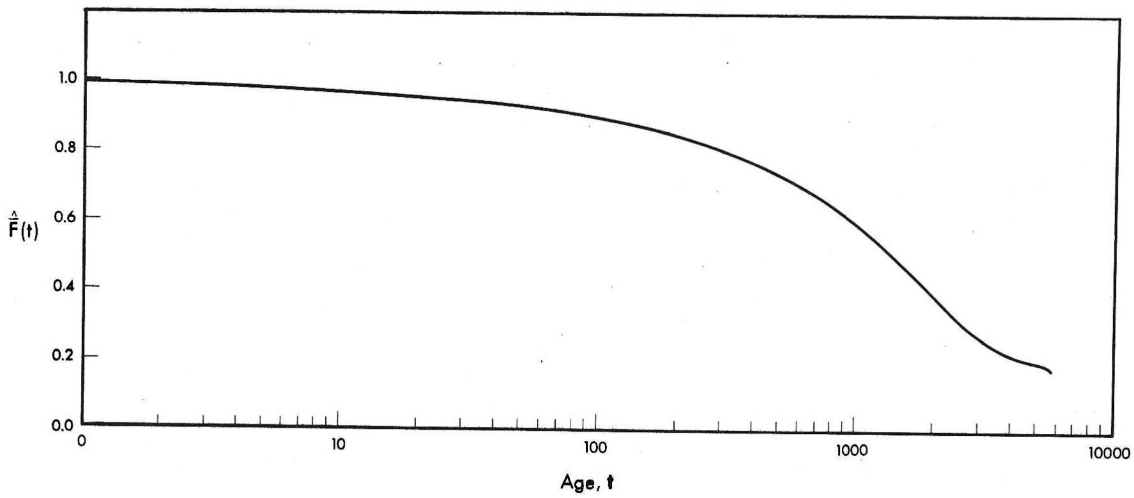

Fig. 4. MLE of $\hat{F}(t)$, the survival probability function

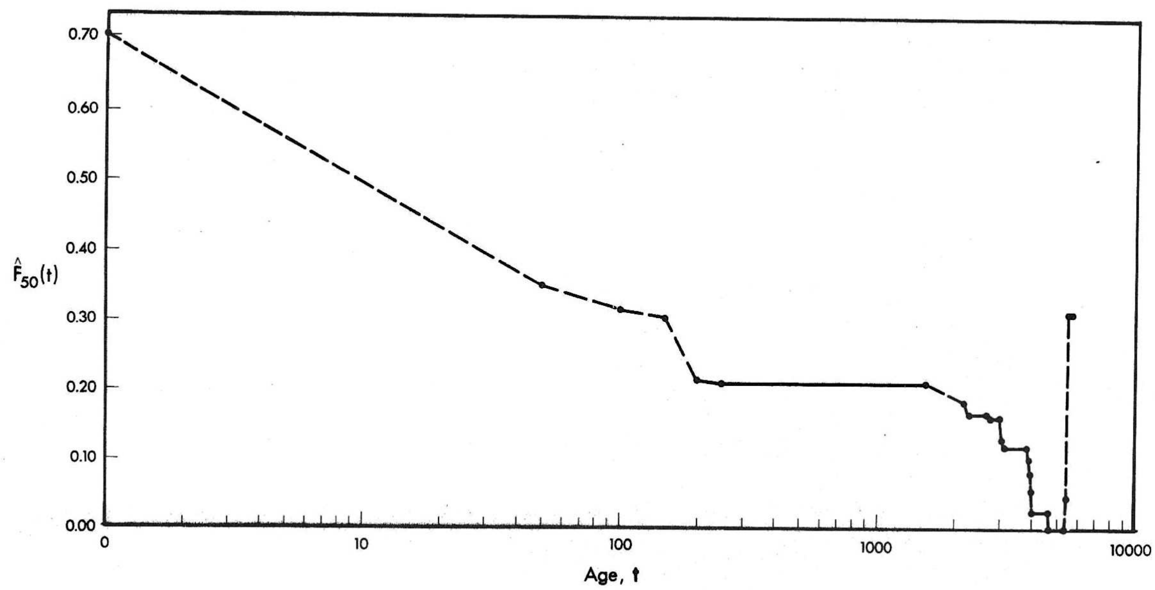

Fig. 5. MLE of $\hat{F}_{50}(t)$, of the conditional probability of failure in the next 50 hours given survival until age t

of the conditional probability of failure in the next 50 hours, given that the unit is alive at age t, in Fig. 5.

4. Derivation of the Maximum Likelihood Estimate

4.1 The Case $X_k < X_{k+1}$

Lemma 4.1 Assume $r(t)$ is decreasing on $[0,t_0]$ and increasing on (t_0,∞). Then $\hat{r}(t)$ is constant between successive failures; specifically $\hat{r}(t)$ has the form (4.3).

Proof Note that the likelihood expression is composed of factors of the following types:

$$(4.1) \qquad r(Y_i)e^{-\int_{a_i}^{Y_i} r(u)\,du},$$

$$(4.2) \qquad e^{-\int_{a_i}^{b_i} r(u)\,du}.$$

(4.1) represents the likelihood of a unit failing at age Y_i given that it is alive at age a_i, while (4.2) represents the likelihood of a unit surviving until age b_i given that it is alive at age a_i.

It follows that

$$(4.3) \qquad r^*(t) = \begin{cases} r_i & \text{for } X_{i-1} < t \le X_i, \quad 1 \le i \le j-1 \quad \text{for } j > 1, \\ 0 & \text{for } X_{j-1} < t < X_j, \\ r_i & \text{for } X_i \le t < X_{i+1}, \quad j \le i \le k \quad \text{for } j < k+1 \end{cases}$$

minimizes $\int_a^b r(u)\,du$ for $0 \le a < b \le X_{k+1}$ among all $r(t)$ satisfying the assumptions that $r(t)$ is decreasing on $[0,t_0]$, increasing on

(t_0,∞), and $= r_i$ at X_i for $i = 1,\ldots,k$. This follows from the fact that $r^*(t)$ is as small as possible at every value of t, $0 \leq t \leq X_{k+1}$, consistent with the assumptions.||

Next we must determine the set of values $r_i = r(X_i)$, $i = 1,\ldots,k$, which maximizes the likelihood. To do so we first express the likelihood L_j in terms of the $r_1,\ldots,r_k$ and $T_1,\ldots,T_{k+1}$, assuming the turning point t_o satisfies $X_{j-1} \leq t_0 < X_j$, where j may be $1,\ldots,k+1$. Note that if $t_0 \geq X_{k+1}$, the likelihood reduces to the value for the case $X_k \leq t_0 < X_{k+1}$, namely L_{k+1}. This is a consequence of the fact that in both cases the failure rate must be taken 0 on the interval (X_k,∞) in order to maximize the likelihood.

<u>Lemma 4.2</u> L_j is given by

$$(4.4) \qquad L_j = \left\{\prod_{i=1}^{j-1} r_i e^{-r_i T_i}\right\} r_j \prod_{i=j+1}^{k} r_i e^{-r_{i-1}T_i} e^{-r_k T_{k+1}}$$

for $j = 1,\ldots,k$;

$$(4.5) \qquad L_{k+1} = \prod_{i=1}^{k} r_i e^{-r_i T_i}$$

where products of the form $\prod_{i=1}^{0}$, $\prod_{k+1}^{k}$ are, of course, vacuous.

<u>Proof</u> Consider a typical interval $(X_{i-1},X_i]$, where say $i \leq j - 1$. Suppose in this interval exposure periods of duration $\tau_1,\tau_2,\ldots,\tau_m$ have been observed, one of which (say the h^{th}) ends in failure at time X_i. The corresponding likelihood of factors are $e^{-r_i\tau_1}, e^{-r_i\tau_2},\ldots,e^{-r_i\tau_{h-1}}$, $r_i e^{-r_i\tau_h}, e^{-r_i\tau_{h+1}},\ldots,e^{-r_i\tau_m}$, with product $r_i e^{-r_i T_i}$, since by definition $T_i = \tau_1+\ldots+\tau_m$.

The other factors in (4.4) and (4.5) may be obtained in a similar fashion.||

Next we note that to maximize L_j given by (4.4) subject to both $r_1 \geq r_2 \geq \ldots r_{j-1} \geq 0$ and $0 \leq r_j \leq r_{j+1} \leq \ldots \leq r_k$, it suffices

(a) to maximize $\prod_{i=1}^{j-1} r_i e^{-r_i T_i}$ subject to $r_1 \geq r_2 \geq \ldots \geq r_{j-1} \geq 0$,

and separately

(b) to maximize $r_j \prod_{i=j+1}^{k} r_i e^{-r_{i-1} T_i} e^{-r_k T_{k+1}}$ subject to $r_j \leq r_{j+1} \leq \ldots \leq r_k$.

Consider problem (a). Let $\bar{r}_1 \geq \ldots \geq \bar{r}_{j-1}$ denote the maximizing set. Then we will need:

<u>Lemma 4.3</u> Let $r_1^*,\ldots,r_{j-1}^*$ denote the set maximizing $\prod_{i=1}^{j-1} r_i e^{-r_i T_i}$ subject only to $r_i \geq 0$, $i = 1,\ldots,j-1$. If for some i $(1 \leq i \leq j-1)$, $r_i^* < r_{i+1}^*$, then $\bar{r}_i = \bar{r}_{i+1}$.

<u>Proof</u> It suffices to show that $\bar{r}_i > \bar{r}_{i+1}$ implies $r_i^* \geq r^*_{i+1}$.

Assume then that $\bar{r}_i > \bar{r}_{i+1}$. It is easy to verify that $r_i e^{-r_i T_i}$ has a unique maximum $r_i^* = T_i^{-1}$, and that $r_i e^{-r_i T_i}$ is increasing for $0 \leq r_i \leq r_i^*$ and decreasing for $r_i^* \leq r_i$.

Now $\bar{r}_1 \geq \bar{r}_2 \geq \ldots \geq \bar{r}_i > \bar{r}_{i+1} \geq \ldots \geq \bar{r}_{j-1}$. Suppose $r_i^* < \bar{r}_i$. Let $r_i' = \max(r_i^*, \bar{r}_{i+1})$. Then $\bar{r}_1 \geq \bar{r}_2 \geq \ldots \geq \bar{r}_{i-1} > r_i' \geq \bar{r}_{i+1} \geq \ldots \geq \bar{r}_{j-1}$. Since $\bar{r}_i > r_i'$, then $\bar{r}_i e^{-\bar{r}_i T_i} < r_i' e^{-r_i' T_i}$, which is a contradiction, since it would imply that $\bar{r}_1 \geq \bar{r}_2 \geq \ldots \geq \bar{r}_{j-1}$ is not the maximizing set. Thus $r_i^* \geq \bar{r}_i$. Similarly we may show $r_{i+1}^* \leq \bar{r}_{i+1}$. It follows that $r_i^* \geq r_{i+1}^*$.||

Thus if there is a reversal in the unrestricted maximum likelihood estimates, say $r_i^* < r_{i+1}^*$, then $\bar{r}_i = \bar{r}_{i+1}$. In this event, set $r_i = r_{i+1}$ in $\prod_{i=1}^{j-1} r_i e^{-r_i T_i}$ and solve the unrestricted problem, where now $r_i = r_{i+1}$. We now obtain as the new trial estimates $r_i^{**} = r_{i+1}^{**} = [\frac{1}{2}(T_i+T_{i+1})]^{-1}$. It is clear that repeated applications of Lemma 4.3 (and its analogue on $[t_0,\infty)$) results in the Procedure described in Section 2 for obtaining the maximum likelihood $\hat{L}_j$ assuming $X_{j-1} \leq t_0 < X_j$, $j=1,\ldots,k+1$, or $X_{k+1} \leq t_0$.

It is apparent that to obtain the MLE assuming t_0 unknown, i.e., to solve the original problem, we simply compare the $\hat{L}_j, j=1,\ldots,k+1$, and find the largest; it corresponds to the maximum likelihood, as described in Section 2.

4.2 The Case $X_k = X_{k+1}$

In this case, distributions placing mass at X_k must be considered. In fact since they dominate those placing no mass at X_k (with respect to the magnitude of the likelihood), we need consider *only* distributions placing mass at X_k. Confining ourselves to such distributions, we may show by an argument similar to that of Lemma 4.1 that the MLE has constant failure rate between successive observed failures.

To write down the likelihood L_j assuming $X_{j-1} \leq t_0 < X_j$, we define as before $r_i = r(X_i)$, $i=1,\ldots,k-1$. Then

$$(4.6)\quad L_j = \left\{\prod_{i=1}^{j-1} r_i e^{-T_i r_i}\right\} r_j \prod_{i=j+1}^{k-1} r_i e^{-T_i r_{i-1}} e^{-T_k r_{k-1}} \quad \text{for } j=1,\ldots,k-1.$$

The justification of (4.6) is similar to that of (4.4) and (4.5), except that in the likelihood expression we set equal to 1 the conditional probability of a unit failing at time X_k given that it has survived until time X_k without failure. Note that the expression in (4.6) coincides with that of (4.4) with $k - 1$ in place of k.

As in Section 4.1, we maximize the likelihood expression in (4.6) subject to $r_1 \geq \ldots \geq r_{j-1}$ *and* $r_j \leq \ldots \leq r_{k-1}$ by separately maximizing

(a) $\prod_{i=1}^{j-1} r_i e^{-T_i r_i}$ subject to $r_1 \geq \ldots \geq r_{j-1}$, and

(b) $r_j \prod_{i=1+1}^{k-1} r_i e^{-T_i r_{i-1}} e^{-T_k r_{k-1}}$ subject to $r_j \leq \ldots \leq r_{k-1}$.

The solution is as given in Section 4.1, where of course $k + 1$ is replaced by k.

5. Consistency of the Estimate

In this section we prove that the maximum likelihood estimate of the failure rate described in Section 2 is consistent. I.e., as the number of observations increases to infinity, the estimate converges to the true value with probability one.

Since the proof of consistency is rather long and involved, we give a sketch of the main ideas used.

Method of Proof If the MLE $\hat{r}_n(\cdot)$ converges weakly to some U-shaped limit $\tilde{r}(\cdot)$, then we first show (Lemma 5.2), using the methods of Marshall and Proschan (1965), that $\tilde{r}(\cdot)$ is a.s. equal to $r(\cdot)$ at those points where $\tilde{r}(\cdot)$ and $r(\cdot)$ are both increasing or both decreasing.

The demonstration depends on the form of the estimate; the fact that $\hat{r}_n(\cdot)$ is the MLE is incidental in the demonstration.

Next we must show that $\tilde{r}(\cdot)$ and $r(\cdot)$ cannot have different turning points. If we assume that they do, we arrive at a contradiction by constructing the estimate $\check{r}_n(\cdot)$, the MLE given the true turning point, and then showing that with probability 1 it is asymptotically more likely than $\hat{r}_n(\cdot)$, (which is the MLE over a larger class of U-shaped failure rates).

The proof that $\check{r}_n(\cdot)$ is more likely than $\hat{r}_n(\cdot)$ uses a uniform bound on the failure rate over the interval of possible disagreement of $\tilde{r}(\cdot)$ and $r(\cdot)$. The existence of this bound requires a separate argument if $\tilde{r}(\cdot)$ is monotone instead of strictly U-shaped. (See Lemma 5.5.)

Definitions

We use the notation $\hat{r}_n(\cdot)$ for the MLE of the U-shaped failure rate, based on a sample of size n, as described in Section 2.

Given a U-shaped failure rate $r(\cdot)$ we shall say that y is a *point of decrease* (*increase*) of $r(\cdot)$ if in every neighborhood of y there exists $u, v, u < y < v$ such that $r(u) > r(v)$; $(r(u) < r(v))$.

A point t is said to be a *turning point* for $r(\cdot)$ if $r(\cdot)$ has no points of decrease greater than t and no points of increase less than t.

We say that a sequence of functions $f_n(\cdot)$ *converges* to a function $f(\cdot)$ *in the weak sense* if $f_n(x)$ converges to $f(x)$ for all points of continuity of $f(\cdot)$.

Corresponding to a sample of size n, let $F_n(\cdot)$ be the empirical distribution function, i.e., $F_n(y) = \frac{1}{n}\cdot$(number of failures observed to have occurred by time y). If $F_n(\cdot)$ converges in the weak sense to some limit $F_0(\cdot)$, note that $F(\cdot)$ will not in general be the true distribution function; this is a consequence of the truncation of the data.

Given an interval $I = (x,y)$ and a hazard rate $\bar{r}(\cdot)$, we define a *measure of likelihood of* $\bar{r}(\cdot)$ *on* I as follows:

$$\mathcal{L}_I(\bar{r}(\cdot),n) = \frac{1}{nF_n(I)}\left(\sum \ln \bar{r}(X_i) - \int_x^y N_n(t)\bar{r}(t)\,dt\right)$$

where $N_n(t)$ is the number of units on test at time t. Here and in the sequel the sum will be understood to extend over all i such that X_i is in I.

Theorem 5.1 If the empirical distribution function $F_n(\cdot)$ converges a.s. to an absolutely continuous distribution function $F_0(\cdot)$, and if the true failure rate $r(\cdot)$ is strictly positive on the support of $F_0(\cdot)$, then on $(0,\infty)$ $\hat{r}_n(\cdot)$ a.s. converges in the weak sense to $r(\cdot)$ on the support of $F_0(\cdot)$, except possibly at some one turning point of $r(\cdot)$.

Proof Assume for the sake of simplicity that the support of $F_0(\cdot)$ is $[0,\infty)$. If $t^{(n)}$ is a turning point of $\hat{r}_n(\cdot)$, then there exists a subsequence $t^{(m(n))}$ which converges to some t_0 in $[0,\infty]$.

If $\hat{r}_{m(n)}(\cdot)$ is a sequence of U-shaped functions assuming values in $[0,\infty]$ with turning points converging to a limit, then it follows as in the Helly weak compactness theorem that $\hat{r}_{m(n)}(\cdot)$ has a subsequence which converges to a U-shaped function $\tilde{r}(\cdot)$ at points of continuity of the latter; moreover, t_0 is a turning point of $\tilde{r}(\cdot)$.

To prove consistency of the estimate $\hat{r}_n(\cdot)$ it will suffice to show that $\tilde{r}(\cdot) = r(\cdot)$ at the continuity points of $r(\cdot)$, for it then will follow that every subsequence of $\hat{r}_n(\cdot)$ has a subsequence which converges to $r(\cdot)$ in the weak sense. Hence $\hat{r}_n(\cdot)$ will be convergent to $r(\cdot)$ in the weak sense (cf. Kelley (1955), p. 74, statement (c)).

Let $\hat{r}_{\ell(n)}(\cdot)$ be a subsequence which converges in the weak sense to $\tilde{r}(\cdot)$, and whose turning points converge to t_0. In the sequel we will be concerned only with this subsequence; for notational convenience we will refer to this fixed subsequence as $\hat{r}_n(\cdot)$, and the respective turning points as t_n; thus

$$\hat{r}_n(\cdot) \overset{w}{\to} \tilde{r}(\cdot),$$
$$t_n \to t_0.$$

Let t_t be a turning point of $r(\cdot)$.

Corresponding to the sample of size $\ell(n)$ (from which the MLE $\hat{r}_n(\cdot)$ was calculated), let the empirical distribution function be denoted by $F_n(\cdot)$, and let $\check{r}_n(\cdot)$ denote the MLE of the hazard rate given the turning point t_t. Then $F_n(\cdot)$ converges to $F_0(\cdot)$, with $F_0(\cdot)$ absolutely continuous, and it follows from Marshall-Proschan (1965) that $\check{r}_n(\cdot)$ converges in the weak sense to $r(\cdot)$ everywhere on the support of $F_0(\cdot)$, except possibly at t_t.

The proof of Marshall and Proschan is stated for the case where the data is uncensored; however, the arguments are applicable in the present situation if we replace the $T_{n,i}$ of their proof by the $T_{n,i}$ of Lemma 5.2 below.

It follows from Lemma 5.2 that if a is the supremum of the points of decrease of $\tilde{r}(\cdot)$ less than $(t_0 \wedge t_t)$, and $\lim \hat{r}_n(a^+) < \infty$, then $\tilde{r}(a^-) = r(a^-)$ a.s. and $\tilde{r}(a^+) = r(a^+)$ a.s. (In particular, $\lim \hat{r}_n(a^+) < \infty$ for $a > 0$.) Thus, regardless of whether a is a point of continuous decrease or a jump point of $\tilde{r}(\cdot)$, it is a.s. a point of decrease for $r(\cdot)$.

Since a is a point of decrease of $\tilde{r}(\cdot)$, it follows that there must be a sequence s_n, where s_n is a jump point of $\hat{r}_n(\cdot)$, such that s_n converges to a. If $t^* < a$, then for arbitrary $\varepsilon > 0$ we must have $t^* < s_n < a + \varepsilon$ for n sufficiently large. In this case $\hat{r}_n(t^*)$ is an average which cannot extend to the right of s_n; hence for $\varepsilon > 0$ and sufficiently large n, $\hat{r}_n(t^*)$ is an average which cannot extend to the right of $a + \varepsilon$.

Similarly, since $\check{r}_n(\cdot)$ is consistent, $\check{r}_n(t^*)$ is an average which for sufficiently large n a.s. cannot extend to the right of $a + \delta$, for arbitrary small $\delta > 0$. Thus it can be shown that $\hat{r}_n(t^*)$ and $\check{r}_n(t^*)$ have a.s. for large n the same functional form, and are therefore equal for all $t \leq t^*$.

Two consequences are immediate:

i) $r(\cdot) \overset{W}{=} \tilde{r}(\cdot)$ a.s. on $(0,a)$

ii) $\mathcal{L}_{(0,a-\varepsilon]}(\check{r}_n(\cdot),n) = \mathcal{L}_{(0,a-\varepsilon]}(\hat{r}_n(\cdot),n)$ with probability 1 for sufficiently large n.

Similarly, if b is the infimum of the points of decrease of $\tilde{r}(\cdot)$ greater than $(t_0 \vee t_t)$, $b < \infty$, then

iii) $r(\cdot) \overset{W}{=} \tilde{r}(\cdot)$ a.s. on (b,∞)

iv) $\mathcal{L}_{[b+\varepsilon,\infty)}(\check{r}_n(\cdot),n) = \mathcal{L}_{[b+\varepsilon,\infty)}(\hat{r}_n(\cdot),n)$ a.s. with probability 1 for sufficiently large n.

The case $a = 0$ or $b = \infty$ is considered in Lemma 5.5.

It follows from Lemma 5.2(a) that on the open interval (a,b) $\tilde{r}(\cdot)$ can have at most two jump points (at t_0 and t_t).

The proof is completed by assuming that $a < b$ and $\tilde{r}(\cdot)$ is not equal in the weak sense to $r(\cdot)$ on (a,b) and then proving the contradictory statement that $\check{r}_n(\cdot)$ is, for large n, more likely than $\hat{r}_n(\cdot)$.

Now, $\hat{r}_n(\cdot)$ and $\tilde{r}(\cdot)$ are uniformly bounded on (a,b), $0 < a \leq b < \infty$. From the fact that $\hat{r}_n(\cdot)$ is unimodal and is consistent off (a,b), it can be seen that for small $\varepsilon > 0$, $\hat{r}_n(\cdot)$ and $\tilde{r}(\cdot)$ are a.s. uniformly bounded on the interval $I = (a-\varepsilon, b+\varepsilon)$.

Similarly, on I $\check{r}_n(\cdot)$ is a.s. uniformly bounded above and below. Therefore, in the limit $\mathcal{L}_I(\check{r}_n(\cdot),n)$ and $\mathcal{L}_I(r(\cdot),n)$ are almost surely equal.

Since $\tilde{r}(\cdot) \overset{W}{=} r(\cdot)$ a.s. off $J = (a,b)$, the result of Lemma 5.4 holds with J replaced by I. If $a \neq b$ it then follows from Lemma 5.3 that with probability 1, $r(\cdot)$ is, in the limit, more likely than $\hat{r}_n(\cdot)$ on I, and for large n $\hat{r}_n(\cdot)$ and $\check{r}_n(\cdot)$ are equal off I. Thus, if $a \neq b$, $\check{r}_n(\cdot)$ is for large n a.s. more likely than $\hat{r}_n(\cdot)$, which is a contradiction. Hence we must conclude that $a = b = t_0 = t_t$, and with the possible exception of this point $\tilde{r}(\cdot)$ is equal to $r(\cdot)$ in the weak sense with probability 1.||

Lemma 5.2 Let t be a point of increase or a point of decrease of $\tilde{r}(\cdot)$. Then, with probability 1,

a) $r(t^-) = \tilde{r}(t^-)$, $r(t^+) = \tilde{r}(t^+)$ if $t \in (0, t_0 \wedge t_t)$,

b) $r(t^-) = r(t^+) = \tilde{r}(t^-) = \tilde{r}(t^+)$ if $t \in (t_0 \wedge t_t, t_0 \vee t_t)$,

c) $r(t^-) = \tilde{r}(t^-)$, $r(t^+) = \tilde{r}(t^+)$ if $t \in (t_0 \vee t_t, \infty)$.

Proof First we will assume $t_0 > t > t_t$; the case $t_0 < t < t_t$ may be treated similarly. We show $r(t^+) \geq \tilde{r}(t^-)$ and $\tilde{r}(t^+) \geq r(t^-)$.

For the first inequality, take $t^\dagger$ and t^* such that $t_t < t^\dagger < t^* < t$. Then, for sufficiently large n, $\hat{r}_n(\cdot)$ is decreasing at t^*, and has a jump point which converges to t. Therefore, for small $\varepsilon > 0$ and sufficiently large n,

$$\frac{1}{\hat{r}_n(t^*)} = \min_{v: t^* < x_v \leq t+\varepsilon} \ \max_{u: 0 \leq x_u \leq t^*} \frac{1}{v-u} \sum_{i=u}^{v-1} T_{n,i}.$$

The upper limit on the range of v follows from the fact that, for large n, $\hat{r}_n(\cdot)$ has a jump point in $(t^*,(t+\varepsilon)\wedge t_n)$.

Here and in the sequel, let

$$Z_{n,i} = \int_{x_i}^{x_{i+1}} N_n(t)r(t)dt.$$

The $Z_{n,i}$ are independent exponential random variables with mean 1.

$T_{n,i}$ may be written:

$$T_{n,i} = \int_{x_i}^{x_{i+1}} N_n(t)dt.$$

Then, letting $k(t)$ be the largest k such that $X_k \leq t$:

$$\frac{1}{\hat{r}_n(t^*)} \geq \min_{v:t^*<x_v\leq t+\varepsilon} \frac{1}{v-k(t^\dagger)} \sum_{i=k(t^\dagger)}^{v-1} T_{n,i}.$$

For this range of i:

$$T_{n,i} \geq Z_{n,i}/r(t+\varepsilon).$$

Therefore

$$\frac{1}{\hat{r}_n(t^*)} \geq \frac{1}{r(t+\varepsilon)} \cdot \min_{v:t^*<x_v\leq t+\varepsilon} \frac{1}{v-k(t^\dagger)} \sum_{i=k(t^\dagger)}^{v-1} Z_{n,i},$$

and the result $r(t+\varepsilon) \geq \tilde{r}(t^*)$ a.s. follows from a form of the strong law of large numbers as in Marshall-Proschan (1965).

To prove the inequality $\tilde{r}(t^+) \geq r(t^-)$ a.s., the procedure is the same. Take $t < t^* < t^\dagger < t_0$; then for arbitrary $\varepsilon > 0$ and large n, we have

$$\frac{1}{\hat{r}_n(t^*)} = \min_{v:t^*<x_v\le t_n} \max_{u:t-\varepsilon\le x_u\le t^*} \frac{1}{v-u}\sum_{i=u}^{v-1} T_{n,i},$$

$$\frac{1}{\hat{r}_n(t^*)} \le \max_{u:t-\varepsilon\le x_u\le t^*} \frac{1}{k(t^\dagger)-u} \sum_{i=u}^{k(t^\dagger)-1} T_{n,i}.$$

For this range of i, $T_{n,i} \le Z_{n,i}/r(t-\varepsilon)$, and the result follows as before.

We briefly outline a proof for the case $t < t_0 \wedge t_t$; the proof in the case $t > t_0 \vee t_t$ is similar. To show $r(t^-) \ge \tilde{r}(t^-)$, take $t^\dagger < t^* < t$. Then

$$\frac{1}{\hat{r}_n(t^*)} = \min_{v:t^*<x_v\le t+\varepsilon} \max_{u:0\le x_u\le t^*} \frac{1}{v-u}\sum_{i=u}^{v-1} T_{n,i}$$

$$\ge \min_{v:t^*<x_v\le t+\varepsilon} \frac{1}{v-t} \sum_{i=k(t)}^{v-1} Z_{n,i}/r(t^\dagger).$$

Hence $r(t^\dagger) \ge \tilde{r}(t^*)$ a.s. To show $r(t^-) \le \tilde{r}(t^-)$, take $t^* < t^\dagger < t$. Then

$$\frac{1}{\hat{r}_n(t^*)} \le \max_{u:x_u\le t^*} \frac{1}{t^\dagger-u} \sum_{i=u}^{k(t^\dagger)-1} Z_{n,i}/r(t^\dagger).$$

Therefore $r(t^\dagger) \le \tilde{r}(t^*)$ a.s., and the result $r(t^-) = \tilde{r}(t^-)$ follows. The proof that $r(t^+) = \tilde{r}(t^+)$ a.s. is similar.||

Lemma 5.3 Let $I = (a-\varepsilon, b+\varepsilon)$. Then

$$\lim_n \mathcal{L}_I(\hat{r}_n(\cdot),n) = -\infty \quad \text{a.s., or}$$

$$\lim_n\left[\mathcal{L}_I(\hat{r}_n(\cdot),n) - \mathcal{L}_I(\tilde{r}(\cdot),n)\right] = 0 \quad \text{a.s.}$$

<u>Proof</u> $\hat{r}_n(\cdot)$ converges to $\tilde{r}(\cdot)$ a.e. with respect to $F(\cdot)$ on I. It follows from Egoroff's theorem that this convergence is almost uniform on I; that is, given $\varepsilon > 0$ there exists a set $E \subset I$, $F(E) < \varepsilon$, such that $\hat{r}_n(\cdot)$ converges uniformly to $\tilde{r}(\cdot)$ on $I \sim E$.

Now:

$$\mathcal{L}_I(\hat{r}_n(\cdot),n) - \mathcal{L}_I(\tilde{r}(\cdot),n) =$$

$$\frac{F_n(E)}{F_n(I)}\left[\mathcal{L}_E(\hat{r}_n(\cdot),n) - \mathcal{L}_E(\tilde{r}(\cdot),n)\right]$$

$$+ \frac{F_n(I\sim E)}{F_n(I)}\left[\mathcal{L}_{I\sim E}(\hat{r}_n(\cdot),n) - \mathcal{L}_{I\sim E}(\tilde{r}(\cdot),n)\right].$$

Recall that $\tilde{r}(\cdot)$ assumes its minimum value between a and b and is piecewise constant there. Hence either $\tilde{r}(\cdot)$ is bounded away from zero, or (since the $\hat{r}_n(\cdot)$ are bounded above) $\lim_n \mathcal{L}_{I\sim E}(\hat{r}_n(\cdot),n) = -\infty$. In the former case $\tilde{r}(\cdot)/\hat{r}_n(\cdot)$ converges uniformly to 1. Hence

$$\lim_n \left| \mathcal{L}_{I\sim E}(\hat{r}_n(\cdot),n) - \mathcal{L}_{I\sim E}(\tilde{r}(\cdot),n) \right| =$$

$$\lim_n \left| \frac{1}{\ell(n)F_n\{I\sim E\}} \left\{ \sum \lg \frac{\tilde{r}(x_i)}{\hat{r}_n(x_i)} + \int_{I\sim E} \left(1 - \frac{\tilde{r}(t)}{\hat{r}_n(t)}\right)\hat{r}_n(t)N_n(t)\,dt \right\} \right| \leq$$

$$\lim_n \left[\max_{s\in I\sim E} \left| \lg \frac{\tilde{r}(s)}{\hat{r}_n(s)} \right| + \max_{s\in I\sim E} \left| 1 - \frac{\tilde{r}(s)}{\hat{r}_n(s)} \right| \right] = 0.$$

Using the uniform bound on $\hat{r}_n(\cdot)$ and $\tilde{r}(\cdot)$ mentioned in the hypothesis and the fact that $r(\cdot)$ is bounded away from zero, the contribution of the likelihood over E may be made arbitrarily small.||

Lemma 5.4 Let $J = (a,b)$. Then

$$\lim_n \left[\mathcal{L}_J(r(\cdot),n) - \mathcal{L}_J(\tilde{r}(\cdot),n)\right] > 0 \quad \text{a.s.}$$

Proof As mentioned above, it follows from Lemma 1 that $\tilde{r}(\cdot)$ has at most two points of increase or decrease in J. Thus, unless $\tilde{r}(\cdot) \overset{W}{=} r(\cdot)$, there is at least one subinterval of J with positive $F(\cdot)$ measure whereon $\tilde{r}(\cdot)$ is constant and not equal in the weak sense to $r(\cdot)$. Let M be such a subinterval on which $\tilde{r}(\cdot) = \tilde{r}$, say.

Then, letting $m = \ell(n)F_n(M)$, we have

$$L(n) = \mathcal{L}_M(r(\cdot),n) - \mathcal{L}_M(\tilde{r}(\cdot),n) =$$

$$= \frac{1}{m}\sum \ln r(x_i) - Z_{n,i} - \ln \tilde{r} + \tilde{r}\int_{x_i}^{x_{i+1}} N_n(t)\,dt.$$

Differentiating with respect to $\tilde{r}$ and using the value which minimizes $L(n)$ for each n, we have:

$$L(n) \geq \ln\left(\frac{1}{n}\sum\int_{x_i}^{x_{i+1}} N_n(t)\,dt\right) - \frac{1}{n}\sum \ln[1/r(x_i)] - \frac{1}{n}\sum(Z_{n,i}-1).$$

We may write

$$\int_{x_i}^{x_{i+1}} N_n(t)\,dt = Z_{n,i}/r_i, \quad \min_{t\varepsilon(x_i,x_{i+1})} r(t) \leq r_i \leq \max_{t\varepsilon(x_i,x_{i+1})} r(t) \quad .$$

Thus

$$\begin{aligned} L(n) &\geq \ln \frac{1}{m}\sum Z_{n,i}/r_i - \frac{1}{m}\sum \ln[1/r(x_i)] - \frac{1}{m}\sum(Z_{n,i}-1) \\ &= \ln\left(1 + \left\{\frac{(Z_{n,i}-1)/r_i}{\sum 1/r_i}\right\}\right) + \ln\left(\frac{1/r_i}{\sum 1/r(x_i)}\right) \\ &\quad + \left[\ln \frac{1}{m}\sum 1/r(x_i) - \frac{1}{m}\sum \ln[1/r(x_i)]\right] - \left\{\frac{1}{m}\sum(Z_{n,i}-1)\right\}. \end{aligned}$$

As in Marshall-Proschan (1965), the quantities in braces tend a.s. to zero by the Hájek-Rényi inequality. By the same inequality, the difference inside the square brackets can be shown to have a positive lower bound unless $r(\cdot)$ is constant on M. Also $\sum 1/r_i \geq \sum 1/r(x_i) - \frac{1}{\min\limits_{x_i \text{ in } M} r(x_i)}$. Therefore,

$$\ln\left(\frac{\sum 1/r_i}{\sum 1/r(x_i)}\right) \geq \ln\left(1 - \frac{\frac{1}{\min r(x_i)}}{\sum 1/r(x_i)}\right),$$

which tends to zero a.s.

Hence $L(n)$ has a positive lower bound unless $r(\cdot)$ is constant on M.

There remains only the case where $r(\cdot)$ is constant on M. If $r(\cdot)$ and $\tilde{r}(\cdot)$ are constant on M, then the difference in the measures of likelihood over M may be written as

$$\left(\frac{\tilde{r}}{r} - 1\right)\left(\frac{1}{n}\sum Z_{n,i}\right) + \ln r - \ln \tilde{r},$$

which converges a.s. to

$$f(\tilde{r}) = \left(\frac{\tilde{r}}{r} - 1\right) + \ln r - \ln \tilde{r}.$$

It may be seen that $f(\tilde{r})$ has a unique minimum at $\tilde{r} = r$, and that $f(r) = 0$. Therefore unless $\tilde{r}(\cdot) \overset{W}{=} r(\cdot)$ on M, we have

$$\lim\left[\mathcal{L}_M(r(\cdot),n) - \mathcal{L}_M(\tilde{r}(\cdot),n)\right] > 0.$$

(Alternatively, if $r(\cdot)$ is constant over $(t_0 \wedge t_t, t_0 \vee t_t)$, then we may redefine t_t to be equal to t_0 and it follows that $r(\cdot) \overset{W}{=} \tilde{r}(\cdot)$, except possibly at t_0.)||

Lemma 5.5 If $a = 0$, then for small $\varepsilon > 0$,

$$\lim_n \left[\mathcal{L}_{(0,\varepsilon]}(\check{r}_n(\cdot),n) - \mathcal{L}_{(0,\varepsilon]}(\hat{r}_n(\cdot),n)\right] \geq 0 \quad \text{a.s.}$$

If $b = \infty$, then for N sufficiently large,

$$\lim_n \left[\mathcal{L}_{[N,+\infty)}(\check{r}_n(\cdot),n) - \mathcal{L}_{[N,+\infty)}(\hat{r}_n(\cdot),n)\right] \geq 0 \quad \text{a.s}$$

Hence the conclusion to the main theorem follows as before.

Proof For the case $a = 0$ we begin by assuming $t_t > 0$. Recall that the only difficulty in applying the methods of Lemmas 5.3 and 5.4 occurs when $\lim \hat{r}_n(0^+) = \infty$. In this case take $\varepsilon > 0$, $\varepsilon < t_t$, and if $t_0 > 0$, take $\varepsilon < t_0$. Then $\tilde{r}(\cdot)$ is constant over all such possible values of ε.

Let s_n be a sequence of jump points of $\hat{r}_n(\cdot)$ which converges to 0. (If no such sequence exists, it is easy to demonstrate that $\hat{r}_n(0^+)$ is uniformly bounded.)

If $t^* < s_n$, then

$$\hat{r}_n(t^*) = \max_{t^*<v\leq s_n} \min_{u\leq t^*} \left[\frac{1}{v-u}\sum_{i=u}^{v-1} T_{n,i}\right]^{-1},$$

$$\check{r}_n(t^*) = \max_{t^*<v\leq t_t} \min_{u\leq t^*} \left[\frac{1}{v-u}\sum_{i=u}^{v-1} T_{n,i}\right]^{-1}.$$

Therefore $\check{r}_n(t^*) \geq \hat{r}_n(t^*)$.

Define

$$\hat{\hat{r}}_n(t) = \begin{cases} \tilde{r}(t) & t \le s_n \\ \hat{r}_n(t) & t > s_n \end{cases}$$

$$\check{\check{r}}_n(t) = \begin{cases} \tilde{r}(t) & t \le s_n \\ \check{r}_n(t) & t > s_n \end{cases} .$$

Then, on $(0,\varepsilon]$, $\hat{\hat{r}}_n(\cdot) \overset{w}{\to} \tilde{r}(\cdot)$, $\check{\check{r}}_n(\cdot) \overset{w}{\to} r(\cdot)$ a.s., and the $\hat{\hat{r}}_n(\cdot)$ are uniformly bounded. It follows as in Lemma 5.3 that

$$\lim \left[\mathcal{L}_{(0,\varepsilon]}(\tilde{r}_n(\cdot),n) - \mathcal{L}_{(0,\varepsilon]}(\hat{\hat{r}}_n(\cdot),n)\right] \to 0 \quad \text{a.s.},$$

and as in Lemma 3 that

$$\lim \left[\mathcal{L}_{(0,\varepsilon]}(r(\cdot),n) - \mathcal{L}_{(0,\varepsilon]}(\tilde{r}(\cdot),n)\right] \ge 0 \quad \text{a.s.}$$

It can be easily shown, as in the proof of (5.1) below, that

$$\lim \left[\mathcal{L}_{(0,\varepsilon]}(\check{\check{r}}_n(\cdot),n) - \mathcal{L}_{(0,\varepsilon]}(\check{r}_n(\cdot),n)\right] \ge 0 \quad \text{a.s.},$$

and, since $\check{r}_n(\cdot)/r(\cdot)$ is bounded below, then

$$\lim \left[\mathcal{L}_{(0,\varepsilon]}(\check{r}_n(\cdot),n) - \mathcal{L}_{(0,\varepsilon]}(r(\cdot),n)\right] \ge 0 \quad \text{a.s.}$$

Therefore,

$$\lim \left[\mathcal{L}_{(0,\varepsilon]}(\check{\check{r}}_n(\cdot),n) - \mathcal{L}_{(0,\varepsilon]}(r(\cdot),n)\right] \ge 0 \quad \text{a.s.}$$

Combining these limits, we obtain

$$\lim \left[\mathcal{L}_{(0,\varepsilon]}(\check{\check{r}}_n(\cdot),n) - \mathcal{L}_{(0,\varepsilon]}(\hat{\hat{r}}_n(\cdot),n)\right] \ge 0 \quad \text{a.s.}$$

The result follows if we can show

$$\lim \Delta\mathcal{L}(n) =$$

$$\lim\ (\mathcal{L}_{(0,\varepsilon]}(\check{\check{r}}_n(\cdot),n) - \mathcal{L}_{(0,\varepsilon]}(\hat{\hat{r}}_n(\cdot),n)) - (\mathcal{L}_{(0,\varepsilon]}(\check{r}(\cdot),n) - \mathcal{L}_{(0,\varepsilon]}(\hat{r}(\cdot),n))$$

$$\leq 0 \quad \text{a.s.}$$

Now, by definition

$$\Delta\mathcal{L}(n) = \frac{1}{nF_n(\varepsilon)} \left\{ - \sum_{i:x_i \leq s_n} (\ln \check{r}(x_i) - \ln \hat{r}(x_i)) + \left(\int_0^{s_n} rN_n dt - \int_0^{s_n} \tilde{r}N_n dt \right) - \left(\int_0^{s_n} \check{r}N_n dt - \int_0^{s_n} \hat{r}N_n dt \right) \right\}.$$

As mentioned above, $(\ln \check{r}_n(t) - \ln \hat{r}_n(t))$ is non-negative on $(0,s_n)$. $\int_0^{s_n} rN_n dt$ is a sum of $nF_n(s_n)$ independent exponential random variables, and therefore is a.s. $o(nF_n(\varepsilon))$. $\int_0^{s_n} \hat{r}N_n dt = nF_n(s_n)$ and therefore is $o(nF_n(\varepsilon))$ a.s. Hence $\lim_n \Delta\mathcal{L}(n) \leq 0$ a.s.

The case $t_t = 0$ may be treated similarly. Instead of $\check{r}_n(\cdot)$ we consider $\bar{r}_n(\cdot)$, the MLE of the U-shaped failure rate, with turning point s_n.

If $s_n < \varepsilon$, $\bar{r}_n(\cdot)$ is more likely on (ε,∞) than is $\check{r}_n(\cdot)$, and the proof of Theorem 5.1 proceeds as before.

The case $b = \infty$ is similar. The fact that we must consider an infinite interval (N,∞) offers no difficulty in the argument of Lemma 5.3, since convergence there is with regard to the finite measure $F(\cdot)$.||

6. Computer Program

A computer program has been written which will perform the operations described in earlier sections to estimate the failure rate, the probability distribution, and the conditional probability of failure in the next N units of time where N is an input parameter. The program is written in the FORTRAN programming language and is broken into four separate subroutines as follows:

1. The main subroutine which controls the over-all computation and calculates the final results. A user need call only this subroutine and provide it with the required input data.
2. A subroutine for computing the total exposure.
3. A subroutine for adjusting the failure rate estimate to eliminate reversals.
4. A subroutine for computing the likelihood.

Each subroutine includes comments giving the names of the pertinent quantities used by that subroutine.

Listings of the subroutines are given on the following pages.

```
C     MLE OF DFR-IFR--SUBROUTINE FALRAT
C
C     INPUT REQUIRED BY SUBROUTINE FALRAT
C
C     B(I) = AGE OF UNIT I AT BEGINNING OF OBSERVATION
C     A(I,J) = AGE OF UNIT I AT ACTION J
C     E(I) = AGE OF UNIT I AT END OF OBSERVATION
C     NU = NUMBER OF UNITS
C     NA(I) = NUMBER OF ACTIONS FOR UNIT I
C     FALINT=CONDITIONAL PROBABILITY OF FAILURE INTERVAL LENGTH
C     KMAX = LENGTH OF ARRAY G(I)
C
C     OUTPUT FROM SUBROUTINE FALRAT
C
C     NTA = TOTAL NUMBER OF ACTIONS + 1
C     X(I) = ORDERED SET OF ACTIONS
C     RH(I) = ESTIMATED FAILURE RATE IN INTERVAL X(I)-X(I-1)
C     FML = LOG(LIKLIHOOD OF RH(I))
C     FB(I) = MLE OF CAPITAL F BAR OF X(I)
C     K=NO. OF G(I)
C     G(I)=PROBABILITY OF FAILURE IN NEXT FALINT INTERVAL
C
      SUBROUTINE FALRAT(B,A,E,NU,NA,NTA,X,RH,FMLMX,FB,FALINT,K,KMAX,G)
C
      DIMENSION B(1),A(1),E(1),NA(1),X(1),RH(1)
     1,FB(1),G(1)
      DIMENSION T(5000),TI(5000)
C
      CALL TOTEXP(B,A,E,NU,NA,NTA,X,T,TI)
      WRITE(6,2000)(I,T(I),TI(I),I=1,NTA)
 2000 FORMAT(//4H1  I,5X,1HT,10X,3H1/T/(1X,I3,2X,F5.0,2X,F11.8))
      RH(1)=0.
      CALL DFRIFR(T,TI,2,NTA,RH,1,NTA)
      CALL LIKHOD(T,RH,NTA,1,FMLMX)
      ITPS=1
      LM=NTA-2
      DO 20 ITP=2,LM
      ITPM=ITP-1
      ITPP=ITP+1
      CALL DFRIFR(T,TI,1,ITPM,RH,2,NTA)
      RH(ITP)=0.
      CALL DFRIFR(T,TI,ITPP,NTA,RH,1,NTA)
      CALL LIKHOD(T,RH,NTA,ITP,FML)
      IF(FML-FMLMX)20,20,10
   10 FMLMX=FML
      ITPS=ITP
   20 CONTINUE
      ITP=NTA-1
      CALL DFRIFR(T,TI,1,NTA-2,RH,2,NTA)
      RH(ITP)=0.
      RH(NTA)=1.0/T(NTA)
      CALL LIKHOD(T,RH,NTA,ITP,FML)
      IF(FML-FMLMX)26,26,24
   24 FMLMX=FML
```

```
      ITPS=ITP
   26 CALL DFRIFR(T,TI,1,NTA-1,RH,2,NTA)
      RH(NTA)=0.
      CALL LIKHOD(T,RH,NTA,NTA,FML)
      IF(FML-FMLMX)40,40,30
   30 FMLMX=FML
      GO TO 80
   40 IF(ITPS-1)60,50,60
   50 RH(1)=0.
      CALL DFRIFR(T,TI,2,NTA,RH,1,NTA)
      GO TO 80
   60 ITPM=ITPS-1
      ITPP=ITPS+1
      CALL DFRIFR(T,TI,1,ITPM,RH,2,NTA)
      RH(ITPS)=0.
      CALL DFRIFR(T,TI,ITPP,NTA,RH,1,NTA)
      IF(ITPS-NTA+1)80,70,80
   70 RH(NTA)=1.0/T(NTA)
C
C     DISTRIBUTION
C
   80 SM=X(1)*RH(1)
      FB(1)=EXP(-SM)
      DO 90 I=2,NTA
      SM=SM+RH(I)*(X(I)-X(I-1))
   90 FB(I)=EXP(-SM)
C
C     CONDITIONAL PROBABILITIES
C
      K=1
      FBT=1.0
      PR=0.
      FAL=FALINT
      SM=0.
  100 IF(X(1)-FAL)140,140,110
  110 SM=SM+RH(1)*(FAL-PR)
      FBTP=EXP(-SM)
      G(K)=(FBT-FBTP)/FBT
      IF(K-KMAX)130,120,120
  120 STOP
  130 K=K+1
      FBT=FBTP
      PR=FAL
      FAL=FAL+FALINT
      GO TO 100
  140 SM=SM+RH(1)*(X(1)-PR)
      PR=X(1)
      DO 190 I=2,NTA
  150 IF(X(I)-FAL)160,160,170
  160 SM=SM+RH(I)*(X(I)-PR)
      PR=X(I)
      GO TO 190
  170 SM=SM+RH(I)*(FAL-PR)
      FBTP=EXP(-SM)
```

```
      G(K)=(FBT-FBTP)/FBT
      IF(K-KMAX)180,120,120
  180 K=K+1
      FBT=FBTP
      PR=FAL
      FAL=FAL+FALINT
      GO TO 150
  190 CONTINUE
      K=K-1
      RETURN
      END
```

000119

```
C       MLE OF DFR-IFR--SUBROUTINE TOTEXP(B,A,E,NU,NA,NTA,X,T,TI)
C
C       B(I) = AGE OF UNIT I AT BEGINNING OF OBSERVATION
C       A(I,J) = AGE OF UNIT I AT ACTION J
C       E(I) = AGE OF UNIT I AT END OF OBSERVATION
C       NU = NUMBER OF UNITS
C       NA(I) = NUMBER OF ACTIONS FOR UNIT I
C       NTA = TOTAL NUMBER OF ACTIONS + 1
C       X(I) = ORDERED SET OF ACTIONS
C       T(K) = TOTAL EXPOSURE TIME IN INTERVAL X(K)-X(K-1)
C       TI(K) = 1.0/T(K)
C
        SUBROUTINE TOTEXP(B,A,E,NU,NA,NTA,X,T,TI)
C
        DIMENSION B(1),E(1),NA(1),T(1),TI(1),X(1)
        DIMENSION A(500,10)
C
        K=0
        DO 15 I=1,NU
        LM=NA(I)
        IF(LM)15,15,5
      5 DO 10 J=1,LM
        K=K+1
     10 X(K)=A(I,J)
     15 CONTINUE
        NTA=K+1
        FMX=0.
        DO 17 I=1,NU
        IF(E(I)-FMX)17,17,16
     16 FMX=E(I)
     17 CONTINUE
        X(NTA)=FMX
        CALL SORT(X,NTA)
        K=1
        T(1)=0.
        DO 50 I=1,NU
        IF(B(I)-X(1))20,50,50
     20 IF(E(I)-X(1))30,40,40
     30 T(1)=T(1)+E(I)-B(I)
        GO TO 50
     40 T(1)=T(1)+X(1)-B(I)
     50 CONTINUE
        IF(T(1))52,52,54
     52 TI(1)=1.0E+35
        GO TO 56
     54 TI(1)=1.0/T(1)
     56 DO 160 K=2,NTA
        T(K)=0.
        TINT=X(K)-X(K-1)
        DO 130 I=1,NU
        IF(E(I)-X(K-1))130,130,60
     60 IF(B(I)-X(K))65,130,130
     65 IF(E(I)-X(K))70,100,100
```

```
 70 IF(B(I)-X(K-1))80,90,90
 80 T(K)=T(K)+E(I)-X(K-1)
    GO TO 130
 90 T(K)=T(K)+E(I)-B(I)
    GO TO 130
100 IF(B(I)-X(K-1))110,110,120
110 T(K)=T(K)+TINT
    GO TO 130
120 T(K)=T(K)+X(K)-B(I)
130 CONTINUE
    IF(T(K))140,140,150
140 TI(K)=1.0E+35
    GO TO 160
150 TI(K)=1.0/T(K)
160 CONTINUE
    TI(NTA)=0.
    RETURN
    END
```

```
C     MLE OF DFR-IFR--SUBROUTINE DFRIFR(T,TI,II,IF,RH,ID,NTA)
C
C     T(I) = TOTAL EXPOSURE TIME IN INTERVAL X(I)-X(I-1)
C     TI(I) = 1.0/T(I)
C     II = INITIAL INTERVAL X(I)
C     IF = FINAL INTERVAL X(I)
C     RH(I) = ESTIMATED FAILURE RATE IN INTERVAL X(I)-X(I-1)
C     ID = 1 FOR INCREASING FAILURE RATE ESTIMATE
C     ID = 2 FOR DECREASING FAILURE RATE ESTIMATE
C     NTA = TOTAL NUMBER OF ACTIONS + 1
C
      SUBROUTINE DFRIFR(T,TI,II,IF,RH,ID,NTA)
C
      DIMENSION T(1),TI(1),RH(1)
C
      DO 10 I=II,IF
   10 RH(I)=TI(I)
      IF(II-IF)30,20,20
   20 RETURN
   30 I=II
      LBL=0
   40 K=I
   50 I=I+1
      GO TO(60,90),ID
   60 IF(RH(I)-RH(K))100,70,70
   70 IF(I-IF)40,80,80
   80 IF(LBL)30,20,30
   90 IF(RH(I)-RH(K))70,70,100
  100 LBL=1
      SUM=0.
      DO 110 J=K,I
  110 SUM=SUM+T(J)
      FIK=I-K+1
      IF(I-NTA)115,112,112
  112 FIK=FIK-1
  115 RHJ=FIK/SUM
      DO 120 J=K,I
  120 RH(J)=RHJ
      IF(I-IF)50,80,80
      END
```

```
C     MLE OF DFR-IFR--SUBROUTINE LIKHOD(T,RH,NTA,ITP,FLK)
C
C     T(I) = TOTAL EXPOSURE TIME IN INTERVAL X(I)-X(I-1)
C     RH(I) = ESTIMATED FAILURE RATE IN INTERVAL X(I)-X(I-1)
C     NTA = TOTAL NUMBER OF ACTIONS + 1
C     ITP = INTERVAL IN WHICH TURNING POINT IS ASSUMED
C     FLK = LIKLIHOOD
C
      SUBROUTINE LIKHOD(T,RH,NTA,ITP,FLK)
C
      DIMENSION T(1),RH(1)
C
      FLK=0.
      LM=NTA-2
      DO 30 I=1,LM
      IF(I-ITP)10,20,20
   10 FLK=FLK+ALOG(RH(I))-T(I)*RH(I)
      GO TO 30
   20 IF(RH(I+1))30,30,25
   25 FLK=FLK+ALOG(RH(I+1))-RH(I+1)*T(I+1)
   30 CONTINUE
      FLK=FLK-T(NTA)*RH(NTA)
      RETURN
      END
```

000025

REFERENCES

T. A. Bray, G. B. Crawford, and F. Proschan (1967), Maximum likelihood estimation of a failure rate function initially decreasing and then increasing. *Proceedings of the 1967 Spring Seminar on Reliability, Maintainability*, IEEE.

J. L. Kelley (1955), *General Topology*, Van Nostrand, New York.

A. W. Marshall and F. Proschan (1965), Maximum likelihood estimation for distributions with monotone failure rate, *Ann. Math. Statist.* 36, pp. 69-77.