MAXIMUM LIKELIHOOD ESTIMATION FOR DISTRIBUTIONS WITH MONOTONE FAILURE RATE

ALBERT W. MARSHALL AND FRANK PROSCHAN

Boeing Scientific Research Laboratories

1. Introduction. Given a set of observations $X_1, \cdots, X_n$ from a common distribution function F, it is natural in the absence of additional information to estimate F by the usual empirical distribution function. However, one would not use this estimator if there were at hand sufficient *a priori* information about the distribution F, e.g., that F is a member of a given parameteric class such as the normal. In this paper, we examine an intermediate case, the case that F is known to have monotone failure rate. Using the idea of maximum likelihood, Grenander [10] derives an estimator for F in the case of increasing failure rate which itself has increasing failure rate. We discuss this case and also obtain estimators in the case of decreasing failure rate. We show that these estimators are consistent.

2. Properties of IFR distributions, and formulation of the problem. The failure rate (or hazard rate) r of a distribution F having derivative f is defined by $r(x) = f(x)/[1 - F(x)]$, for $F(x) < 1$. It is easy to verify that if r is increasing, then $R(x) \equiv -\log[1 - F(x)]$ is convex on the support of F, an interval. (Throughout this paper we write "increasing" for "nondecreasing" and "decreasing" for "nonincreasing.") Whether f exists or not, we say that F has increasing failure rate (IFR) if the support of F is of the form $[\alpha, \beta]$, $-\infty \leqq \alpha \leqq \beta \leqq \infty$, and if R is convex on $[\alpha, \beta)$. The importance of the IFR property and its applications to life testing and reliability are discussed in [3], [4].

If F is IFR and $F(z) < 1$, then F is absolutely continuous on $(-\infty, z)$. To see this, choose $\epsilon > 0$ and points $\alpha_1 < \beta_1 < \alpha_2 < \beta_2 < \cdots < \alpha_m < \beta_m \leqq z$ satisfying $\sum_1^m (\beta_i - \alpha_i) < \epsilon/r^+(z)$, where $r^+(z) = \lim_{\delta \downarrow 0} [R(z + \delta) - R(z)]/\delta$ exists finitely since R is convex. Then

$$\sum_1^m |R(\beta_i) - R(\alpha_i)| = \sum_1^m \frac{R(\beta_i) - R(\alpha_i)}{\beta_i - \alpha_i} (\beta_i - \alpha_i) \leqq r^+(z) \sum_1^m (\beta_i - \alpha_i) \leqq \epsilon.$$

Thus R is absolutely continuous on $(-\infty, z)$, and the result follows. Note, however, that F may have a jump at the right-hand endpoint of its interval of support.

For convenience, if F is IFR we define $r(x) = \infty$ for all x such that $F(x) = 1$. Note that for any distribution F and any x for which r is finite on $(-\infty, x)$, we have

$$(2.1) \qquad 1 - F(x) = \exp[-R(x)] = \exp[-\textstyle\int_{-\infty}^x r(z)\, dz].$$

Further properties of IFR distributions have been discussed in [2].

Received 7 January 1964; revised 30 July 1964.

Let $\mathfrak{F}$ be the class of IFR distributions, and let $X_1 \leqq X_2 \leqq \cdots \leqq X_n$ be obtained by ordering a random sample from an unknown distribution F in $\mathfrak{F}$. It is not possible to obtain a maximum likelihood estimator for $F \varepsilon \mathfrak{F}$ directly by maximizing $\prod_{i=1}^n f(X_i)$, since for $F \varepsilon \mathfrak{F}$, $f(X_n)$ can be arbitrarily large. Consequently, we first consider the subclass $\mathfrak{F}^M$ of distributions F in $\mathfrak{F}$ with corresponding failure rates bounded by M, obtaining $\sup_{F\varepsilon\mathfrak{F}^M} \prod_i f(X_i) \leqq M^n$. We shall see that there is a unique distribution $\hat{F}_n^M$ in $\mathfrak{F}^M$ at which the supremum is attained. These conventional maximum likelihood estimators $\hat{F}_n^M$ for $\mathfrak{F}^M$ converge in distribution as $M \to \infty$ (i.e., as $\mathfrak{F}^M \to \mathfrak{F}$) to an estimator $\hat{F}_n \varepsilon \mathfrak{F}$ which we call *maximum likelihood for* $\mathfrak{F}$. Furthermore, the density $\hat{f}_n^M$ and failure rate $\hat{r}_n^M$ of $\hat{F}_n^M$ converge in a natural way to the density $\hat{f}_n$ and failure rate $\hat{r}_n$ (of the continuous part) of $\hat{F}_n$ as is shown below in Section 3.

It is not difficult to see that the maximum likelihood estimator $\hat{F}_n$ for F defined above coincides with the maximum likelihood estimator defined by Kiefer and Wolfowitz [11], p. 893.

3. Derivation of the estimators. From (2.1) we obtain that the log likelihood $L = L(F)$ is given, for $F \varepsilon \mathfrak{F}^M$, by

$$L = \sum_1^n \log r(X_i) - \sum_1^n \int_{-\infty}^{X_i} r(z)\, dz. \tag{3.1}$$

L is maximized over $\mathfrak{F}^M$ by a distribution with failure rate constant between observations, as shown by Grenander [10] as follows: Let $F \varepsilon \mathfrak{F}^M$ have failure rate r and let F^* be the distribution with failure rate

$$\begin{aligned} r^*(x) &= 0, && x < X_1 \\ &= r(X_i), && X_i \leqq x < X_{i+1}, \quad i = 1, 2, \cdots, n-1 \\ &= r(X_n), && x \geqq X_n . \end{aligned} \tag{3.2}$$

Then $F^* \varepsilon \mathfrak{F}^M$, and $r(x) \geqq r^*(x)$ so that $-\int_{-\infty}^{X_i} r(z)\, dz \leqq -\int_{-\infty}^{X_i} r^*(z)\, dz$ for all i; we conclude that $L(F) \leqq L(F^*)$. Thus, we may replace L by the function

$$\textstyle\sum_1^n \log r(X_i) - \sum_1^{n-1} (n - i)(X_{i+1} - X_i) r(X_i). \tag{3.3}$$

The maximization of (3.3) subject to $r(X_1) \leqq \cdots \leqq r(X_n) = M$ is performed in [10]; it can also be performed as a direct application of [7], Corollary 2.1 and the discussion following (see also [13] [14]). This yields for r (corresponding to $F \varepsilon \mathfrak{F}^M$) the estimator

$$\hat{r}_n^M(X_i) = \min\left(\min_{v \geqq i+1} \max_{u \leqq i} \{1/(v - u)[r_u^{-1} + \cdots + r_{v-1}^{-1}]\}^{-1}, M\right) \tag{3.4}$$

where $r_n = M$ and

$$r_j = [(n - j)(X_{j+1} - X_j)]^{-1} \quad \text{for } j = 1, 2, \cdots, n - 1. \tag{3.5}$$

The estimator $\hat{r}_n^M$ given in (3.4) differs in form from the one given in [10] but is equivalent to it. We will use the form given in (3.4) to establish consistency.

The maximization procedure which yields (3.4) may be described as follows. First, find the maximum of (3.3) obtaining (3.5). If there is a reversal, say $r_i > r_{i+1}$, then set $r(X_i) = r(X_{i+1})$ in (3.3) and repeat the procedure. After, at most, n steps of this kind, a monotone estimator is obtained. The maximum derived with $r(X_i) = r(X_{i+1})$ can be directly obtained by replacing r_i and r_{i+1} by their harmonic mean, $(r_i^{-1} + r_{i+1}^{-1})^{-1}$. Succeeding steps amount to further such averaging which is extended just to the point necessary to eliminate all reversals. It can be seen that this is exactly what is called for in (3.4) taking into account $r(x) \leqq M$. (In this connection, see also [1], [6].) The resulting estimator $\hat{r}_n^M$ is of the form

$$\begin{aligned} \hat{r}_n^M(x) &= 0, & x &< X_1 \\ &= \min(r_{n_i+1,n_{i+1}}, M), & X_{n_i+1} \leqq x &< X_{n_{i+1}+1} \\ &= M, & x &\geqq X_n, \end{aligned}$$

where $r_{1,n_1} \leqq r_{n_1+1,n_2} \leqq \cdots \leqq r_{n_k+1,n-1}$, $0 = n_0 < n_1 < \cdots < n_k < n - 1$, and $r_{n_i+1,n_{i+1}}$ is the harmonic mean of $r_{n_i+1}, r_{n_i+2}, \cdots, r_{n_{i+1}}$. The n_i are determined by the rule which determines the extent of the averaging.

The estimator for r corresponding to $F \varepsilon \mathfrak{F}$ is obtained by letting $M \to \infty$ in (3.4), and is given by

$$\begin{aligned} \hat{r}_n(X_i) = \min_{v \geqq i+1} \max_{u \leqq i} [v - u][(n - u)(X_{u+1} - X_u) \\ + \cdots + (n - v + 1)(X_v - X_{v-1})]^{-1}. \end{aligned} \tag{3.6}$$

$i = 1, 2, \cdots, n - 1$ and $\hat{r}_n(X_n) = \infty$. For the remaining values of x, $\hat{r}_n(x)$ is determined by (3.2) with $\hat{r}_n$ replacing r and r^*. The corresponding estimators $\hat{F}_n$ and $\hat{f}_n$ for F and f are obtained from $\hat{r}_n$ using (2.1) and the relation $\hat{f}_n(x) = \hat{r}_n(x)[1 - \hat{F}_n(x)]$.

It is of interest to note that the estimator $\hat{r}_n$ can also be written in the form

$$\hat{r}_n(x) = \inf_{F_n(v) \geqq F_n(x)} \sup_{F_n(u) < F_n(x)} [F_n(v) - F_n(u)]/\textstyle\int_u^v [1 - F_n(y)]\, dy, \tag{3.7}$$

where F_n is the empirical distribution. Similarly, since r is increasing, $r(x)$ is given by (3.7) with F replacing F_n. Note that the infimum of (3.7) may be taken over the set $v \geqq x$, but the supremum may be taken over $u < x$ only if $x = X_i$ for some i (in the corresponding formula for r with F replacing F_n, $F(u) < F(x)$ can without restriction be replaced by $u < x$). It may be possible to demonstrate consistency of the maximum likelihood estimate of F using (3.7) and the Glivenko-Cantelli theorem.

It is easily seen from (3.6) or (3.7) that

$$[\hat{r}_n(X_i)]^{-1} = \sup_{v \geqq i+1} \inf_{u \leqq i} [\varphi(v) - \varphi(u)]/(v - u), \tag{3.8}$$

where $\varphi(j) = n\int_0^{X_j} (1 - F_n(x)]\, dx$. Let φ^* be the convex minorant of φ (i.e., φ^* is the supremum of convex functions which at each j do not exceed $\varphi(j)$). Then as shown in [10], $[\hat{r}(x)]^{-1}$ is the right-hand derivative of φ^* at x. This representation may be of some importance for computation, since φ^* is easily obtained graphically from φ.

4. Consistency. It can be verified that the regularity conditions of [11] are satisfied for the family $\mathfrak{F}^M$ (though they are violated for $\mathfrak{F}$). Thus from the results of [11], it follows that $\hat{F}_n^M(t)$ is a consistent estimator of $F(t)$ when F is in $\mathfrak{F}^M$. For fixed $t < \beta$, choose $M > r(t)$; then it follows that with probability one for sufficiently large n, $\hat{F}_n^M(t) = \hat{F}_n(t)$. We conclude that $\hat{F}_n(t)$ is a consistent estimator of $F(t)$ when F is in $\mathfrak{F}$. Similar conclusions can also be obtained from the results of [9]. However, rather than verify the regularity conditions, we choose to give a direct proof of consistency. This direct proof can also be adapted to the DFR case, where it is not clear that consistency can be demonstrated using the results of [9] or [11].

THEOREM 4.1. *If r is increasing, then for every t_0,*

$$r(t_0 -) \leqq \liminf \hat{r}_n(t_0) \leqq \limsup \hat{r}_n(t_0) \leqq r(t_0 +) \tag{4.1}$$

with probability one.

PROOF. The right-hand inequality is trivial if $r(t_0) = 0$ or $r(t_0 +) = \infty$; otherwise, let $t_1 > t_0$ satisfy $r(t_1) < \infty$, and let $a_j(n) + 1$ be the index of the largest observation $\leqq t_j$, $j = 0, 1$. Let $N_1(n)$ and $N_2(n)$ be defined by

$$\hat{r}_n(t_0) = [N_2(n) - N_1(n)][\textstyle\sum_{i=N_1(n)+1}^{N_2(n)} (n - i)(X_{i+1} - X_i)]^{-1}.$$

Let $Y = -[r(t_1)]^{-1} \log [1 - F(X)]$, so that $P\{Y > y\} = P\{1 - F(X) < e^{-r(t_1)y}\} = e^{-r(t_1)y}$, i.e., Y has an exponential distribution. Since the X_i are order statistics from the distribution F, $Y_i = -[r(t_1)]^{-1} \log [1 - F(X_i)]$ are order statistics from the exponential distribution, and $(n - i)(Y_{i+1} - Y_i)$ are independent, identically distributed exponential random variables, with mean $1/r(t_i)$. Finally,

$$\begin{aligned} Y_{i+1} - Y_i &= [r(t_1)]^{-1}[\textstyle\int_{-\infty}^{X_{i+1}} r(z)\,dz - \int_{-\infty}^{X_i} r(z)\,dz] \\ &= \textstyle\int_{X_i}^{X_{i+1}} [r(z)/r(t_1)]\,dz \leqq X_{i+1} - X_i, \qquad i \leqq a_1(n). \end{aligned} \tag{4.2}$$

From (3.6) and (4.2), it follows that

$$\begin{aligned} \hat{r}_n(t_0) &\leqq [a_1(n) - N_1(n)][\textstyle\sum_{i=N_1(n)+1}^{a_1(n)} (n - i)(X_{i+1} - X_i)]^{-1} \\ &\leqq [a_1(n) - N_1(n)][\textstyle\sum_{i=N_1(n)+1}^{a_1(n)} (n - i)(Y_{i+1} - Y_i)]^{-1}. \end{aligned}$$

This implies that

$$[\hat{r}_n(t_0)]^{-1} \geqq \min_{j \leqq a_0(n)} [a_1(n) - j]^{-1} \textstyle\sum_{i=j+1}^{a_1(n)} (n - i)(Y_{i+1} - Y_i).$$

Writing $r(t_1) = \mu^{-1}$, we have that

$$\begin{aligned} &P\{|\min_{j \leqq a_0(n)} [a_1(n) - j]^{-1} \textstyle\sum_{i=j+1}^{a_1(n)} [(n - i)(Y_{i+1} - Y_i) - \mu]| \geqq \epsilon\} \\ &\qquad = P\{|\max_{a_1(n)-a_0(n) \leqq k \leqq a_1(n)} [-k^{-1} \textstyle\sum_{i=1}^{k} (Z_i - \mu)]| \geqq \epsilon\} \leqq P(B_n), \end{aligned}$$

where $B_n = \{\max_{a_1(n)-a_0(n)\leqq k\leqq a_1(n)}|k^{-1}\sum_{i=1}^k (Z_i - \mu)| \geqq \epsilon\}$, $Z_i \equiv Z_{i,n}$ are independent exponentially distributed random variables with mean μ and $a_0(n)$, $a_1(n)$ are random variables depending on the Z_i.

In order to conclude that $\limsup \hat{r}_n(t_0) \leqq r(t_0+)$ with probability one, we wish to show that $P\{\limsup B_n\} = 0$. For arbitrary δ satisfying $0 < 2\delta < F(t_1) - F(t_0)$, let $A_n = \{|a_i(n) - nF(t_i)| < n\delta,\ i = 1, 2\}$, so that $P\{\limsup A_n^c\} = 0$ by the strong law of large numbers (A_n^c is the complement of A_n). Thus

$$P\{\limsup B_n\} = P\{\limsup A_n \cap B_n\} + P\{\limsup A_n^c \cap B_n\}$$
$$= P\{\limsup A_n \cap B_n\}.$$

Therefore, by the Borel-Cantelli lemma, we can conclude $P\{\limsup B_n\} = 0$, by showing $\sum P\{A_n \cap B_n\} < \infty$. To do this, we use the following generalized form of the Hájek-Rényi inequality, obtained with an obvious change of variables from (2.6) of [5]: If $W_1, \cdots, W_m$ are random variables such that $EW_1 = 0$, $E(W_i \mid W_1, \cdots, W_{i-1}) = 0$ a.e., $i = 2, 3, \cdots, m$, $r \geqq 1$ and $\epsilon_l \geqq \cdots \geqq \epsilon_m$, then

$$P\{\max_{l\leqq k\leqq m} \epsilon_k|W_1 + \cdots + W_k| \geqq 1\}$$
$$\leqq \epsilon_l^r E|W_1 + \cdots + W_l|^r + \textstyle\sum_{k=l+1}^m \epsilon_k^r EW_k^r.$$

From this inequality (in our case, the W_i are independent exponential random variables) and with $\theta_1 = F(t_1) - F(t_0) - 2\delta$, $\theta_2 = \delta + F(t_1)$, it follows that

$$P\{A_n \cap B_n\} \leqq P\{\max_{n\theta_1<k<n\theta_2} |(k\epsilon)^{-1}\textstyle\sum_1^k (Z_i - \mu)| \geqq 1\}$$
$$\leqq (n\theta_1\epsilon)^{-4}E|\textstyle\sum_1^{[n\theta_1+1]} (Z_i - \mu)|^4 + \sum_{k=[n\theta_1+1]+1}^{[n\theta_2]-1} (k\epsilon)^4 EZ_i^4$$
$$\leqq (n\theta_1\epsilon)^{-4}\cdot 3\mu^4 n\theta_1(n\theta_1 + 2) + 9\mu^4 n(\theta_2 - \theta_1)(n\theta_1 + 1)^{-4}\epsilon^{-4}.$$

Thus, it is clear that $\sum P\{A_n \cap B_n\} < \infty$. A similar proof yields the left-hand inequality.||

COROLLARY 4.2. *If r is increasing, then for all t, $\lim_{n\to\infty} \hat{F}_n(t) = F(t)$ with probability one.*

PROOF. It is sufficient to prove the theorem for t satisfying $F(t) < 1$, in which case $\hat{F}_n(t) < 1$ for sufficiently large n. By Theorem 4.1, $\lim_{n\to\infty} \hat{r}_n(z) = r(z)$ except possibly for z in a set of Lebesgue measure zero. For $z \in [x, t]$, $x > -\infty$, $\hat{r}_n(z) < \infty$, and by the Lebesgue dominated convergence theorem,

$$\lim_{n\to\infty} \int_x^t \hat{r}_n(z)\, dz = \int_x^t r(z)\, dz$$

with probability one. Then, by (2.1),

$$\lim_{n\to\infty} \frac{1 - \hat{F}_n(t)}{1 - \hat{F}_n(x)} = \frac{1 - F(t)}{1 - F(x)} \quad \text{with probability one.} \tag{4.3}$$

If we knew that $F(x) = 0$ for some $x > -\infty$, this would complete the proof.

In order to obtain an upper bound for $\int_{-\infty}^{x} \hat{r}_n(z)\,dz$, we first note that

$$\begin{aligned} r_{n_i+1,n_{i+1}} &= [(n_{i+1}+1)-(n_i+1)][(n-(n_i+1))(X_{n_i+2}) \\ &\quad + \cdots + (n-n_{i+1})(X_{n_{i+1}+1}-X_{n_{i+1}})]^{-1} \\ &\leqq [(n_{i+1}+1)-(n_i+1)][(n-n_{i+1})(X_{n_i+2}-X_{n_i+1}) \\ &\quad + \cdots + (n-n_{i+1})(X_{n_{i+1}+1}-X_{n_{i+1}})]^{-1} \\ &= (n_{i+1}-n_i)/(n-n_{i+1})(X_{n_{i+1}}-X_{n_i+1}). \end{aligned}$$

Let $k \equiv k(n)$ be the index of the largest observation not greater than x; if X_k is in $[X_{n_j+1}, X_{n_{j+1}+1})$, we obtain by (3.6),

$$\begin{aligned} \hat{r}_{n_j+1,n_{j+1}} &\leqq [(k+1)-(n_j+1)] \\ &\quad \cdot[(n-(n_j+1))(X_{n_j+2}-X_{n_j+1})+\cdots+(n-k)(X_{k+1}-X_k)]^{-1} \\ &\leqq [k-n_j][(n-k)(X_{k+1}-X_{n_j+1})]^{-1}. \end{aligned}$$

From these estimates, it follows that

$$\begin{aligned} \int_{-\infty}^{x} \hat{r}_n(z)\,dz \leqq \sum_{i=0}^{j-1} \hat{r}_{n_i+1,n_{i+1}}(X_{n_{i+1}+1}-X_{n_i+1}) \\ + \hat{r}_{n_j+1,n_{j+1}}(X_{k+1}-X_{n_j+1}) \leqq k/(n-k). \end{aligned}$$

If $0 < \epsilon < \frac{1}{2}$ and x satisfies $F(x) \leqq \epsilon$, then $\lim k/(n-k) = F(x)/[1-F(x)] < \epsilon/(1-\epsilon) < 2\epsilon$ with probability one, so that $\lim\sup \int_{-\infty}^{x} \hat{r}_n(z)\,dz < 2\epsilon$, and by (2.1), $\lim\inf[1-\hat{F}_n(x)] \geqq e^{-2\epsilon} \geqq 1-2\epsilon$. This together with (4.3) completes the proof.||

COROLLARY 4.3. *If r is increasing and continuous on $[a, b]$, then*
(i) $\lim_{n\to\infty} \sup_{t\in[a,b]} |\hat{r}_n(t) - r(t)| = 0$,
(ii) $\lim_{n\to\infty} \sup_{-\infty<t<\infty} |\hat{F}_n(t) - F(t)| = 0$,
(iii) $\lim_{n\to\infty} \sup_{t\in[a,b]} |\hat{f}_n(t) - f(t)| = 0$,
each with probability one.

PROOF. (i) and (ii) follow from the same methods as in the usual proof of the Glivenko-Cantelli theorem. (iii) follows from (i), (ii), and the fact that $f(t) = r(t)[1-F(t)]$.||

5. Comparison between $r_n(t)$ and $\hat{r}_n(t)$. We shall show that with respect to a certain metric $\hat{r}_n(t)$ is closer to $r(t)$ than is $r_n(t)$, where

$$\begin{aligned} r_n(t) &= 0 && \text{for } 0 \leqq t < X_1 \\ &= \{(n-j)(X_{j+1}-X_j)\}^{-1} && \text{for } X_j \leqq t < X_{j+1}, \\ & && \qquad j = 1, 2, \cdots, n-1 \\ &= \infty && \text{for } X_n \leqq t < \infty. \end{aligned} \tag{5.1}$$

Note that $r_n(t)$ represents the "unaveraged" estimate of the failure rate, i.e., the estimate that does not take into account the requirement that $r(t)$ be increasing. The result is similar to an inequality of [1], p. 644 and is really a special case of the results of Brunk [8]. We give a simple proof for convenience and completeness, based on the following general result.

THEOREM 5.1. *Let h be nondecreasing, g be integrable with respect to the measure μ, the discontinuities of h distinct from the points at which μ places positive mass, and* $\tilde{g}(x) = \sup_{s \leqq x} \inf_{t \geqq x} \int_s^t g(\theta)\, d\mu(\theta)/[\mu(t) - \mu(s)]$. *Then*

$$\int (g - h)^2\, d\mu \geqq \int (\tilde{g} - h)^2\, d\mu + \int (g - \tilde{g})^2\, d\mu. \tag{5.2}$$

PROOF. It suffices to show $\int (\tilde{g} - h)(g - \tilde{g})\, d\mu \geqq 0$. The x-axis can be broken up into single points and maximal intervals on each of which $\tilde{g}(x)$ is constant. At a single point x, $\tilde{g}(x) = g(x)$. Let $[a, b]$ be an interval, with $\tilde{g}(x) = \tilde{g}$ on $[a, b]$. Define $G(x) = \int_{-\infty}^x g(\theta)\, d\mu(\theta)$. Then $\int_a^b (\tilde{g} - h(x))(g(x) - \tilde{g})\, d\mu(x) = \int_a^b \{G(x) - G(a) - \tilde{g}[\mu(x) - \mu(a)]\}\, d\{h(x) - \tilde{g}\} \geqq 0$, since $[G(x) - G(a)]/[\mu(x) - \mu(a)] \geqq \inf_{t \geqq x} [G(t) - G(a)]/[\mu(t) - \mu(a)] = \sup_{s \leqq x} \inf_{t \geqq x} [G(t) - G(s)]/[\mu(t) - \mu(s)] = \tilde{g}$.||

Identifying $h(t)$ as $r(t)$, $g(t)$ as $r_n(t)$, $\tilde{g}(t)$ as $\hat{r}_n(t)$, and $\mu(-\infty, t]$ as $F_n(t)$, the usual empirical distribution, we obtain

THEOREM 5.2. *With probability one,*

$$\begin{aligned} &\int_{-\infty}^{\beta-} \{r_n(t) - r(t)\}^2\, dF_n(t) \\ &\qquad \geqq \int_{-\infty}^{\beta-} \{\hat{r}_n(t) - r(t)\}^2\, dF_n(t) + \int_{-\infty}^{\beta-} \{r_n(t) - \hat{r}_n(t)\}^2\, dF_n(t). \end{aligned} \tag{5.3}$$

Thus, in the sense made precise by (5.3), $\hat{r}_n(t)$ is closer to $r(t)$ than is $r_n(t)$.

6. Decreasing failure rate. A distribution F is said to have decreasing failure rate (DFR) if the support of F is of the form $[\alpha, \infty)$, $\alpha > -\infty$, and if $\log [1 - F(x)]$ is convex on $[\alpha, \infty)$. Such distributions arise, e.g., as mixtures of exponentials (see [12]).

If F is DFR then by an argument similar to that used in the IFR case, it is absolutely continuous except possibly for a discontinuity at the point α. Thus, the measure determined by F is absolutely continuous with respect to $\mu_\alpha = \delta_\alpha + \lambda$, where δ_α places unit mass on $\{\alpha\}$ and λ is Lebesgue measure; we denote the density of F with respect to μ_α by f, and again define the failure rate of F by $r(x) = f(x)/[1 - F(x-)]$. If F is DFR, we always take a version of f for which r is decreasing in (α, ∞).

Allowing for the fact that f is a density with respect to μ_α, we see that (2.1) is replaced by

$$1 - F(x) = [1 - r(\alpha)] \exp [-\textstyle\int_\alpha^x r(z)\, dz]. \tag{6.1}$$

Estimation in the DFR case parallels that in the IFR case, but with some interesting differences. The first of these is that there are really two problems in the DFR case, depending on whether or not the point α is known.

First consider the case that α is known and suppose $\alpha = X_1 = \cdots = X_k < X_{k+1} < \cdots < X_n$ (in case $k = 0$, we define $X_0 = \alpha$). Using (6.1), $f(x) =$

$r(x)[1 - F(x-)]$ and the relations $r(\alpha) = f(\alpha) = F(\alpha+)$, we write the log likelihood in the form

$$k \log r(\alpha) + (n - k) \log (1 - r(\alpha)) + \sum_{i=k+1}^{n} \log r(X_i) - \sum_{i=k+1}^{n} \int_{\alpha}^{X_i} r(z)\, dz.$$

Maximization of the first two terms yields $\hat{r}_n(\alpha) = k/n = \hat{F}(\alpha+)$. Maximization of the last two terms is quite analogous to that in the IFR case, and yields for r the estimator $\hat{r}_n(x) = \hat{r}_n(X_i)$, $X_{i-1} < x \leqq X_i$, $i = k + 1, \cdots, n$, where

$$\hat{r}_n(X_i) = \max_{v \geqq i} \min_{u \leqq i-1} \{(v - u)^{-1}[(n - u)(X_{u+1} - X_u) + \cdots + (n - v + 1)(X_v - X_{v-1})]\}^{-1},$$

and $X_0 = \alpha$ in case $k = 0$.

Contrary to the IFR case, this DFR estimator is not unique; it is determined by the likelihood equation only for $x \leqq X_n$, and may be extended beyond X_n in any manner that preserves the DFR property.

Consider now the case that α is unknown, and assume for the moment that F is absolutely continuous with respect to Lebesgue measure. If F is DFR on $[\alpha, \infty)$ for $\alpha > X_1$, then the likelihood $\Lambda(F) = \prod f(X_i) = 0$. If F is DFR on $[\alpha, \infty)$ for $\alpha < X_1$, then $\Lambda(F) < \Lambda(\tilde{F})$ where $\tilde{F}$ is defined by

$$\begin{aligned} \tilde{F}(x) &= \{F(x) - F(X_1)\}/\{1 - F(X_1)\}, \qquad & x \geqq X_1 \\ &= 0, & x < X_1. \end{aligned}$$

Thus the maximum likelihood estimator for α unknown is found among those DFR distributions with support $[X_1, \infty)$, and the problem reduces to the case of known α.

The proof of consistency in the DFR case is similar to the proof in the IFR case.

7. The discrete case. A related problem of interest occurs in the case that F is discrete IFR. If F is a discrete distribution with mass p_i at x_i, $i = \cdots, -1, 0, 1, 2, \cdots$ and the x_i are ordered increasingly, the ratio $\rho_i = p_i/\sum_{j=i}^{\infty} p_j$, $i = \cdots, -1, 0, 1, 2, \cdots$, is called the (discrete) failure rate of F. If ρ_i is increasing, then F is said to be "discrete IFR." It is easily verified that $p_i = \rho_i \prod_{j=-\infty}^{i-1} (1 - \rho_j)$, $i = \cdots, -1, 0, 1, \cdots$.

If a sample of n independent observations from F consists of m_i occurrences at x_i, where for notational convenience, $i = 1, 2, \cdots, k$, then the log likelihood function is

$$L = \sum_{i=1}^{k} m_i \log p_i = \sum_{i=1}^{k} \{m_i \log \rho_i + (m_{i+1} + \cdots + m_k) \log (1 - \rho_i)\}.$$

We wish to maximize L subject to $\rho_1 \leqq \rho_2 \leqq \cdots \leqq \rho_k$.

With proper identification, this problem is exactly the one solved in [1]. The solution is obtained by averaging (through adding numerators and denominators) the quantities

$$\begin{aligned} \rho_i^* &= 0, & i < 1 \\ &= m_i/(m_i + \cdots + m_k), \qquad & i = 1, 2, \cdots, k \\ &= 1, & i > k, \end{aligned} \tag{7.1}$$

to eliminate any reversals $\rho_j^* > \rho_{j+1}^*$. After sufficient averaging a set of increasing estimates $\hat{\rho}_1, \cdots, \hat{\rho}_k$ are obtained which may be written as

$$(7.2) \quad \hat{\rho}_i = \min_{k \geq r \geq i} \max_{s \leq i} [m_s + m_{s+1} + \cdots + m_r] / \textstyle\sum_{j=r}^{s} (m_j + \cdots + m_k).$$

The estimator given in Section 3 for the continuous case may be derived from this as a limiting case. Consistency of the estimator (7.2) follows as in [1].

If ρ_i is decreasing, then F is said to be "discrete DFR." In this case, maximum likelihood estimators may be obtained and consistency proved using the same method as in the discrete IFR case with obvious modifications.

Acknowledgment. We would like to thank Professor Ronald Pyke for his helpful advice and suggestions.

REFERENCES

[1] Ayer, Miriam, Brunk, H. D., Ewing, G. M., Reid, W. T., and Silverman, Edward (1955). An empirical distribution function for sampling with incomplete information. *Ann. Math. Statist.* **26** 641–647.

[2] Barlow, R. E., Marshall, A. W., and Proschan, F. (1963). Properties of probability distributions with monotone hazard rate. *Ann. Math. Statist.* **34** 375–389.

[3] Barlow, R. E. and Proschan, F. (1962). Planned replacement, appearing in *Studies in Applied Probability and Management Science*, (Arrow, Karlin, and Scarf, Eds.) Stanford Univ. Press.

[4] Barlow, R. E. and Proschan, F. (1964). Comparison of replacement policies, and renewal theory implications. *Ann. Math. Statist.* **35** 577–589.

[5] Birnbaum, Z. W. and Marshall, A. W. (1961). Some multivariate Chebyshev inequalities with extensions to continuous parameter processes. *Ann. Math. Statist.* **32** 687–703.

[6] Brunk, H. D. (1955). Maximum likelihood estimates of monotone parameters. *Ann. Math. Statist.* **26** 607–616.

[7] Brunk, H. D. (1958). On the estimation of parameters restricted by inequalities. *Ann. Math. Statist.* **29** 437–454.

[8] Brunk, H. D. (1961). Best fit to a random variable by a random variable measurable with respect to a σ-lattice. *Pacific J. Math.* **11** 785–802.

[9] Crawford, Gordon and Saunders, S. C. (1964). Nonparametric maximum likelihood estimation. Boeing document D1-82-0308.

[10] Grenander, Ulf (1956). On the theory of mortality measurement, Part II. *Skan. Aktuarietidskr.* **39** 125–153.

[11] Kiefer, J. and Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Ann. Math. Statist.* **27** 887–906.

[12] Proschan, F. (1963). Theoretical explanation of observed decreasing failure rate. *Technometrics* **5** 375–383.

[13] van Eeden, C. (1956). Maximum likelihood estimation of ordered probabilities, *Indag. Math.* **18** 444–455.

[14] van Eeden, C. (1957). Maximum likelihood estimation of partially or completely ordered parameters. *Indag. Math.* **19** 128–136 and 201–211.