The Annals of Statistics
1985, Vol. 13, No. 1, 413–417

INCONSISTENCY OF THE MAXIMUM LIKELIHOOD ESTIMATOR OF A DISTRIBUTION HAVING INCREASING FAILURE RATE AVERAGE

BY R. A. BOYLES, A. W. MARSHALL[1], AND F. PROSCHAN[2],

Lawrence Livermore National Laboratory, University of British Columbia, and Florida State University

Marshall and Proschan (1965) showed that the MLE for a life distribution with increasing failure rate is strongly consistent. In this note we show that the MLE for a life distribution with increasing failure rate average is *not* consistent; in fact, maximum likelihood estimation in the IFRA case yields estimators of the average failure rate and of the distribution function which, in general, converge a.s. to values other than the true values. In the decreasing failure rate average case, the MLE fails to exist.

1. Introduction and summary. Throughout this paper, "decreasing" means "nonincreasing" and "increasing" means "nondecreasing". A life distribution F (i.e., $F(0^-) = 0$) with survival function $\bar{F} \equiv 1 - F$ is said to have an *increasing failure rate average*, written "F is IFRA", if (a) $F(0) = 0$, and (b) $\bar{F}^{1/t}(t)$ is decreasing in $t > 0$. Condition (b) is equivalent to the condition (b)$'$ that the average failure rate $-t^{-1} \log \bar{F}(t)$ is increasing in $t > 0$.

The class of IFRA life distributions plays a fundamental role in reliability theory. It represents the smallest class containing the exponential distributions, closed under formation of coherent systems and taking limits in distribution (see Birnbaum, Esary, and Marshall, 1966). From a more practical point of view, a coherent system of independent increasing failure rate (IFR) components need not have an IFR life distribution, but must have an IFRA life distribution. The IFRA class also arises in a natural way in shock models and wear processes (Esary, Marshall, and Proschan, 1973; A-Hameed and Proschan, 1973, 1975; and Esary and Marshall, 1974). Moreover, it has been shown that the renewal quantity has a distribution on the integers which is IFRA (Esary, Marshall, and Proschan, 1973).

Thus it is of importance and of interest to estimate the distribution function and the average failure rate for an IFRA life distribution. We are motivated to find the maximum likelihood estimator (MLE) for IFRA distributions, since the MLE estimators of the distribution function and of the failure rate function are strongly consistent when the underlying distribution is known to be IFR (Marshall and Proschan, 1965).

Received June 1980; revised July 1984.

[1] Research sponsored in part by the National Science Foundation and in part by the National Sciences and Engineering Research Council of Canada.

[2] Research sponsored by the Air Force Office of Scientific Research Council, AFSC, USAF, under Grant AFOSR-82-K-0007.

AMS 1980 *subject classifications.* Primary 62G05; secondary 62N05.

Key words and phrases. Maximum likelihood estimation, increasing failure rate average, inconsistency.

In this note, we show that the MLE of the distribution function and of the average failure rate function of an IFRA life distribution converge a.s. to functions other than the true ones. Thus we have the curious situation that MLE yields strongly consistent estimators in the relatively small class of IFR distributions, in the very large class of all distributions, but *not* in the intermediate class of IFRA distributions. Another aspect of interest is that this example of the failure of the MLE to yield a good estimator arises from a realistic problem, rather than from a mathematically "pathological" class of distributions.

The inconsistency of the MLE for the similar case of starshaped distributions has been obtained by Barlow, Bartholomew, Bremner and Brunk (1972). They provide a consistent estimator based on isotonic regression that applies both to the case of starshaped distributions and to IFRA distributions.

We wish to thank Professor Ronald Pyke for his help.

2. Derivation of the MLE for an IFRA distribution. First assume $X_1 < X_2 < \cdots < X_n$ constitute the order statistics based on a random sample from an IFRA life distribution F. Since there is no σ-finite measure relative to which every IFRA distribution is absolutely continuous, we employ the concept of nonparametric maximum likelihood proposed by Kiefer and Wolfowitz (1956). As noted by Barlow (1968), this concept leads to the likelihood

$$L = L(F) = n! \prod_{i=1}^{n} [F(X_i) - F(X_i^-)],$$

which must be maximized subject to the IFRA constraint. With $X_0 = 0$, it follows that the MLE $\hat{F}_n$ assumes the form

$$-\log \hat{\bar{F}}_n(x) = \begin{cases} \lambda_j x & \text{for} \quad X_j \le x < X_{j+1}, \quad j = 0, \cdots, n-1, \\ \infty & \text{for} \quad X_n \le x, \end{cases} \tag{2.1}$$

where $\lambda_0 = 0 \le \lambda_1 \le \lambda_2 \le \cdots \le \lambda_{n-1}$. Note that $-x^{-1}\log \hat{\bar{F}}_n(x)$ is increasing on $[0, \infty)$ so that $\hat{F}_n$ is IFRA. Note further that by choosing $-\log \hat{\bar{F}}_n$ linear between ordered observations, no probability is "wasted", but instead, is assigned to the fullest extent at the order statistics, where it makes the maximum contribution to the likelihood function.

Thus the likelihood function L can be expressed as:

$$L = n![\prod_{j=1}^{n-1} (\exp(-\lambda_{j-1}X_{j-1}) - \exp(-\lambda_j X_j))]\exp(-\lambda_{n-1}X_n). \tag{2.2}$$

With $\Delta\lambda_i = \lambda_i - \lambda_{i-1}$, $i = 1, \cdots, n-1$ and $\lambda_0 \equiv 0$, rewrite L as

$$L = n! \prod_{j=1}^{n-1} U_j$$

where $U_j = \exp(-\Delta\lambda_j(\sum_{j+1}^{n} X_i))(1 - \exp(-\Delta\lambda_j X_j))$. To maximize L by appropriate choice of $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_{n-1}$, it suffices to maximize separately the U_j by appropriate choice of $\Delta\lambda_j \ge 0$, $j = i, \cdots, n-1$. Set $d \log U_j/d\Delta\lambda_j = 0$ to obtain the maximizing

$$\Delta\hat{\lambda}_j = X_j^{-1}(\log \textstyle\sum_j^n X_i - \log \sum_{j+1}^n X_i) > 0, \quad j = 1, \cdots, n-1.$$

It follows that the corresponding $\hat{\lambda}_j$ is given by

$$\hat{\lambda}_j = \textstyle\sum_1^j \Delta\hat{\lambda}_k = \sum_{k=1}^{j} X_k^{-1}(\log \sum_k^n X_i - \log \sum_{k+1}^n X_i). \tag{2.3}$$

Thus

$$\hat{\bar{F}}_n(X_j) = e^{-\hat{\lambda}_j X_j} = \prod_{\ell=2}^{j+1} \left(\sum_\ell^n X_i / \sum_{\ell-1}^n X_i\right)^{X_j/X_{\ell-1}} \tag{2.4}$$

constitutes the MLE of the survival probability $\bar{F}(X_j)$, $j = 1, \cdots, n - 1$. For values of the MLE of $\bar{F}(x)$ for x between the ordered observations, simply replace λ_j in (2.1) by $\hat{\lambda}_j$ given in (2.3).

If tied observations occur, results (2.3) and (2.4) still hold. This is a consequence of the fact that $\hat{\lambda}_j$ and $\hat{\bar{F}}_n(X_j)$ are continuous functions of $X_1, \cdots, X_n$.

3. Inconsistency of MLE. To demonstrate the inconsistency of the MLE derived in Section 2, assume F is continuous and let $Y_1, Y_2, \cdots, Y_n$ denote the unordered sample. Then for $X_j \le x < X_{j+1}$ we have from (2.3)

$$\hat{\lambda}_n(x) = \hat{\lambda}_j = \sum_{k=1}^j X_k^{-1} \log[1 + X_k(\sum_{k+1}^n X_i)^{-1}]. \tag{3.1}$$

Consider now the inequality

$$0 < 1/s - (1/t)\log(1 + t/s) < t/s(s + t), \tag{3.2}$$

valid for all s, $t > 0$. We apply (3.2) to each term in the sum in (3.1), taking $t = X_k$ and $s = \sum_{k+1}^n X_i$, to obtain

$$0 < \sum_{k=1}^j (\sum_{k+1}^n X_i)^{-1} - \hat{\lambda}_n(x) < \sum_{k=1}^j X_k(\sum_{k+1}^n X_i)^{-1}(\sum_k^n X_i)^{-1}. \tag{3.3}$$

The latter term is bound above by

$$(\sum_1^j X_k)(\sum_{j+1}^n X_i)^{-2} = (\sum_1^n Y_k[1 - I_k(x)])(\sum_1^n Y_i I_i(x))^{-2} \tag{3.4}$$

where $I_k(x)$ is the indicator function of $\{Y_k > x\}$. Since $E(Y) < +\infty$ if Y has an IFRA distribution, the strong law of large numbers (SLLN) implies that the right-hand side of (3.4) is a.s. $O(n^{-1})$. It therefore follows from (3.3) and (3.4) that

$$\hat{\lambda}_n(x) = \sum_{k=1}^j (\sum_{k+1}^n X_i)^{-1} + O(n^{-1}).$$

In terms of the empirical distribution function F_n this expression becomes

$$\hat{\lambda}_n(x) = \int_{(0,x]} G_n(y)\, dF_n(y) + O(n^{-1}) \tag{3.5}$$

where

$$G_n(y) = [\sum_1^n Y_i I_i(y)/n]^{-1}.$$

The representation (3.5) is valid when there are no tied observations, which is why we restrict attention to continuous F.

By SLLN, with probability one

$$\lim_{n\to\infty} G_n(y) = G(y) \equiv \left[\int_y^\infty z\, dF(z)\right]^{-1}.$$

We actually have a stronger result, from which the inconsistency of $\hat{F}_n$ will follow.

LEMMA. *Assume* $0 < F(x) < 1$. *Then as* $n \to \infty$

$$\sup_{0 \le y \le x} |G_n(y) - G(y)| \to_{\text{a.s.}} 0. \tag{3.6}$$

PROOF. Evidently, G is continuous and nondecreasing, and G_n is right continuous and nondecreasing. Now, with x fixed, we define right continuous distribution functions H_n, H by

$$H_n(y) = \begin{cases} 0 & y < 0 \\ G_n(y)/G_n(x) & 0 \le y \le x \\ 1 & x < y \end{cases}$$

$$H(y) = \begin{cases} 0 & y < 0 \\ G(y)/G(x) & 0 \le y \le x \\ 1 & x < y. \end{cases}$$

Since $G_n(y) \to_{\text{a.s.}} G(y)$ for each y, the arguments used to prove the Glivenko-Cantelli Theorem yield

$$\sup_{-\infty < y < \infty} |H_n(y) - H(y)| \to_{\text{a.s.}} 0 \tag{3.7}$$

(cf. Chung, 1974, page 133). Since $G_n(x) \to_{\text{a.s.}} G(x) < +\infty$ and

$$\begin{aligned} |G_n(y) - G(y)| &\le |G_n(y) - G_n(x)H(y)| + |G_n(x)H(y) - G(y)| \\ &= G_n(x)|H_n(y) - H(y)| + H(y)|G_n(x) - G(x)| \end{aligned}$$

for $0 \le y \le x$, (3.6) follows from (3.7).

Using the lemma, we now have with probability one

$$\begin{aligned} \lim_{n\to\infty} \hat{\lambda}_n(x) &= \lim_{n\to\infty} \int_{(0,x]} G_n(y)\, dF_n(y) = \lim_{n\to\infty} \int_{(0,x]} G(y)\, dF_n(y) \\ &= \int_0^x G(y)\, dF(y), \end{aligned}$$

the last step being another application of SLLN. Consequently, $\hat{\lambda}_n(x)$ is a consistent estimator of

$$\gamma_F(x) \equiv \int_0^x \left[\int_y^\infty z\, dF(z)\right]^{-1} dF(y) \tag{3.8}$$

rather than a consistent estimator of the true failure rate average

$$\lambda(x) = x^{-1} \int_0^x \left[\int_y^\infty dF(z)\right]^{-1} dF(y). \tag{3.9}$$

Note that, in general, the limit of $\hat{\lambda}_n(x)$ given in (3.8) differs from the true value given in (3.9), and similarly the limit of $\hat{F}_n(x) = 1 - \exp(-x\hat{\lambda}_n(x)$ differs from the true value $F(x) = 1 - \exp(-x\lambda(x))$. That is, the ML estimators of the failure rate average and distribution function are *not* strongly consistent.

EXAMPLE 1. (Exponential distribution). Let $\bar{F}(x) = \exp(-\lambda x)$ for $x \geq 0$. Then

$$\gamma_F(x) = \int_0^x \left[\int_y^\infty z\lambda \exp(-\lambda z)\, dz\right]^{-1} \lambda \exp(-\lambda y)\, dy = \lambda \log(1 + \lambda x),$$

after simplification, whereas the true failure rate average is $\equiv \lambda$. Thus $\hat{\bar{F}}_n(x)$ converges almost surely to $(1 + \lambda x)^{-\lambda x}$ rather than to $\bar{F}(x) = \exp(-\lambda x)$. Note that the limiting distribution is IFR, hence IFRA.

EXAMPLE 2. (Uniform distribution). Let $F(x) = x$ for $0 \leq x \leq 1$. Then

$$\gamma_F(x) = \int_0^x \left[\int_y^1 z\, dz\right]^{-1} dy = \log\left[\frac{1 + x}{1 - x}\right],$$

after simplification. Thus, $\hat{\bar{F}}_n(x)$ converges almost surely to $[(1 - x)/(1 + x)]^x$ rather than to $\bar{F}(x) = 1 - x$.

REFERENCES

BARLOW, R. E. (1968). Likelihood ratio tests for restricted families of probability distributions. *Ann. Math. Statist.* **39** 547–560.

BARLOW, R. E., BARTHOLOMEW, D. J., BREMNER, J. M. and BRUNK, H. D. (1972). *Statistical Inference under Order Restrictions.* Wiley, New York.

BIRNBAUM, Z. W., ESARY, J. D. and MARSHALL, A. W. (1966). Stochastic characterization of wearout for components and systems. *Ann. Math. Statist.* **37** 816–825.

CHUNG, K. L. (1974). *A Course in Probability Theory,* 2nd ed. Academic, New York.

ESARY, J. D. and MARSHALL, A. W. (1974). Families of components, and systems, exposed to a compound Poisson damage process, in *Reliability and Biometry: Statistical Analysis of Lifelength,* 31–46, ed. by F. Proschan and R. J. Serfling, Soc. Ind. Appl. Math.

ESARY, J. D., MARSHALL, A. W. and PROSCHAN, F. (1973). Shock models and wear processes. *Ann. Probab.* **1** 627–629.

A-HAMEED, M. S. and PROSCHAN, F. (1973). Nonstationary shock models. *Stochastic Process. Appl.* **1** 383–404.

A-HAMEED, M. S. and PROSCHAN, F. (1975). Shock models with underlying birth process. *J. Appl. Probab.* **12** 18–28.

KIEFER, J. and WOLFOWITZ, J. (1950). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Ann. Math. Statist.* **27** 887–900.

MARSHALL, A. W. and PROSCHAN, F. (1965). Maximum likelihood estimation for distributions with monotone failure rate. *Ann. Math. Statist.* **36** 69–77.

R. A. BOYLES
NERCO MINERALS CO.
111 S.W. COLUMBIA
10TH FLOOR
PORTLAND, OREGON 97201

A. W. MARSHALL
DEPARTMENT OF STATISTICS
UNIVERSITY OF BRITISH COLUMBIA
1984 MATHEMATICS ROAD
VANCOUVER BC
CANADA V6T 1W5

F. PROSCHAN
FLORIDA STATE UNIVERSITY
TALLAHASSEE, FLORIDA