EXAMPLES OF INCONSISTENCY OF MAXIMUM LIKELIHOOD ESTIMATES

By R. R. BAHADUR

Indian Statistical Institute, Calcutta

SUMMARY. This note contains two examples, concerning independent observations from a fixed population of real values, where maximum likelihood estimates of the population distribution function do not converge, as the sample size tends to infinity, to the actual distribution function. Such examples are of interest since in previously published examples of the failure of the method, the estimated distribution, if it exists at all, does converge to the actual one.

The maximum likelihood (m.l.) method estimates the entire population distribution from given data, even if the statistician is interested only in some particular parameter, i.e., some functional of the population distribution. It therefore seems interesting and appropriate, if not logically necessary, to enquire whether the method is consistent in the sense that, in increasing samples of independent observations from a given population, the estimated distribution converges, in a given intrinsic sense, to the actual one. According to this viewpoint, the consistency of the m.l. estimate of a particular parameter or set of parameters (e.g., a set specifying the population distribution) is a subsidiary problem, determined mainly by such non-stochastic questions as whether the parameters of interest are identifiable or whether they are continuous functionals of the population distribution.

Wald's famous proof of the consistency of m.l. estimates (Wald, 1949; c.f. also Kiefer and Wolfowitz, 1956) can easily be formulated so as to yield regularity conditions sufficient for consistency in the sense described above. However, the need for regularity conditions, except in so far as they guarantee the existence of m.l. estimates, is not clear from the examples in the literature (Basu, 1955; Kiefer and Wolfowitz, 1956; Lecam and Kraft, 1956). In some of these examples, the likelihood function is unbounded, so that m.l. or even approximate m.l. estimates do not exist. In the others, where m.l. or approximate m.l. estimates do exist, it can be shown that the estimated distribution converges to the actual distribution in the strong sense that, with probability one, the estimated probability of any event in the sample space of a single observation tends, uniformly in events, to the correct probability. It seems to the writer that in such examples the m.l. estimate can hardly be claimed to be inconsistent, since the claim must be based on convergence definitions remote from statistics, or on the failure of the estimated value of a discontinuous functional to converge to the population value.

The two examples that follow exhibit m.l. estimates that fail to converge to the population distribution even in the weakest of intrinsic senses. In both examples, a single observation from the population is denoted by X, and X takes real values. For definiteness, it is understood that a sequence $P_1, P_2, \ldots$ of probability distributions of X 'converges' to a probability distribution Q if $P_k(X \leqslant t) \to Q(X \leqslant t)$ at every continuity point of the latter distribution function as $k \to \infty$.

Example 1. (*The continuous case*). Suppose that it is known only that X is distributed according to some Q such that

$$dQ = f(x)dx, \int_0^1 f(x)dx = 1, \quad \ldots \quad (1)$$

where f is continuous and

$$0 \leqslant f(x) \leqslant 2 \quad \ldots \quad (2)$$

for $0 \leqslant x \leqslant 1$. Given n independent observations on X, say $X_1, X_2, \ldots$ and X_n, it is clear that P_n is an m.l. estimate of Q if and only if $dP_n = f_n(x)dx$, where f_n is continuous and satisfies (1), (2) and

$$f_n(X_i) = 2 \text{ for } i = 1, 2, \ldots, n. \quad \ldots \quad (3)$$

The inconsequential requirement (3) is satisfied by many inconsistent estimates and also by some consistent ones. The verification is omitted. It might appear at first sight that m.l. estimation fails here because the set of alternative distributions of X is infinite dimensional, but it is easy to construct versions of the same example in which the set in question is only countably infinite.

In the preceding example, the m.l. principle provides no indication of the actual distribution. In the following one, the principle provides a definite but ultimately misleading indication.

Example 2. (*The discrete case*). In this example, (i) X is confined to the non-negative integers $j = 0, 1, 2, \ldots$ ad inf.; (ii) the set of alternative distributions of X is a countably infinite set $\{P_1, P_2, \ldots, P_\infty\}$, where

$$\lim_{k \to \infty} P_k = P_\infty; \quad \ldots \quad (4)$$

(iii) for each $n = 1, 2, \ldots$ and any set $X_1, X_2, \ldots, X_n$ of n sample values, the (or an) m.l. estimate exists; and (iv) no matter what the actual distribution of X, *any* sequence of m.l. estimated distributions converges to P_∞; more precisely, if $h_n = h_n(X_1, X_2, \ldots, X_n)$ is the least index h such that P_h is an m.l. estimate based on the first n observations, then

$$P_k^{(\infty)} [\lim_{n \to \infty} h_n = \infty] = 1 \quad \ldots \quad (5)$$

for each $k = 1, 2, \ldots, \infty$, where $P_k^{(\infty)}$ denotes the probability measure on the space of infinite sequences $X_1, X_2, \ldots$ when X is distributed according to P_k.

Let

$$\alpha_j = 1 + 2\left[\frac{\log \log \log (j+a)}{\log \log (j+a)}\right] \quad \ldots \quad (6)$$

for $j = 1, 2, \ldots$ where a is a constant such that α_j is well-defined and positive for $j \geqslant 1$, e.g. $a = 27$. Next, let

$$m_j = b/(j+a)[\log (j+a)]^{\alpha_j} \quad \ldots \quad (7)$$

for $j = 1, 2, \ldots$ where b is chosen so that

$$m_j > 0, \sum_{j=1}^{\infty} m_j = \frac{1}{4}. \qquad \ldots \quad (8)$$

Define

$$P_{\infty}(X = j) = \begin{cases} 3/4 & \text{for } j = 0 \\ m_j & \text{for } j > 0 \end{cases} \qquad \ldots \quad (9)$$

and for each $k = 1, 2, \ldots$ define

$$P_k(X = j) = \begin{cases} 3/4-d_k & \text{for } j = 0 \\ m_j & \text{for } j > 0, j \neq k \\ \sqrt{m_k} & \text{for } j = k \end{cases} \qquad \ldots \quad (10)$$

where

$$d_k = \sqrt{m_k}-m_k. \qquad \ldots \quad (11)$$

Since $0 < d_k < \frac{1}{2}$ and $d_k \to 0$ as $k \to \infty$ by (8), it is clear from (8), (9), (10) and (11) that $P_1, P_2 \ldots, P_{\infty}$ is a set of distinct probability distributions, and that (4) holds.

Now let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed observations on X. For any $n = 1, 2, \ldots$ let $f_{jn} = f_{jn}(X_1, \ldots, X_n)$ denote the frequency of the event $\{X = j\}$ in the first n observations and let $L_n(k) = L_n(X_1, X_2 \ldots, X_n | k)$ be the logarithm of the likelihood function. Then, for $k < \infty$, we have

$$L_n(k) - L_n(\infty) = f_{0n} \log [1-(4d_k/3)]+(\tfrac{1}{2})f_{kn} \log (1/m_k), \qquad \ldots \quad (12)$$

by (9) and (10). Let

$$Y_n = \max\{X_1, X_2, \ldots, X_n\}. \qquad \ldots \quad (13)$$

Since $f_{jn} = 0$ for all $j > Y_n$, it follows from (12) that $L_n(k) \leqslant L_n(\infty)$ for all $k > Y_n$. Consequently there exists an h such that $L_n(h) = \sup_k\{L_n(k)\}$ (i.e. an m.l. estimate exists), and there must be such an h in the set $\{1, 2, \ldots, Y_n, \infty\}$. Let h_n be the smallest maximizing h. We proceed to show that (5) holds. It will be shown, incidentally, that

$$P_k^{(\infty)} [h_n \leqslant Y_n \text{ for all sufficiently large } n] = 1 \qquad \ldots \quad (14)$$

for each $k = 1, 2, \ldots, \infty$. It is interesting to note that (14) holds even if h_n is the *largest* maximizing index, so that the distribution P_{∞}, to which all m.l. estimates converge, is comparatively unattractive to the m.l. principle, even when P_{∞} itself obtains.

Choose and fix a $k = 1, 2, \ldots$, or ∞ and assume henceforth that X is distributed according to P_k. Since $Y_n \to \infty$ with probability 1, since $f_{jn} \geqslant 1$ for $j = Y_n$, and since

$$\log \left(\frac{1}{m_j} \right) \geqslant \log (j+a)-\log (b) \qquad \ldots \quad (15)$$

for $j \geqslant 1$ by (6) and (7), it follows from (12) that

$$\liminf_{n\to\infty}\left\{\frac{L_n(Y_n)-L_n(\infty)}{n}\right\} \geqslant \liminf_{n\to\infty}\left\{\frac{\log Y_n}{2n}\right\} \qquad \dots (16)$$

with probability one. Now let M be an arbitrary positive integer. It follows from (12) that

$$\max_{1\leqslant j\leqslant M}\left\{\frac{L_n(j)-L_n(\infty)}{n}\right\} \leqslant N \qquad \dots (17)$$

where N is a constant depending only on M. Suppose for the moment that

$$P_k^{(\infty)}\left[\lim_{n\to\infty}\left\{\frac{\log Y_n}{n}\right\}=\infty\right]=1. \qquad \dots (18)$$

It will then follow from (16), (17) and the definition of $\hat{h}_n$ that, with probability one, $M < \hat{h}_n \leqslant Y_n$ for all sufficiently large n. Since M is arbitrary, this will establish (5) and (14).

It remains therefore to verify (18). For any $z \geqslant 1$, let m_z be defined by writing z instead of j in (6) and (7). Then m is a decreasing function of z. Hence we have

$$P_k(X \geqslant z) = \sum_{j\geqslant z} m_j \geqslant \int_{z+1}^{\infty} m_z\, dz = \frac{b}{\log\, \log(a+z+1)} \qquad \dots (19)$$

for all $z \geqslant 1$ if $k=\infty$ and for all $z > k$ if $k < \infty$, by (9) and (10). Let c be a constant, $0 < c < \infty$. A straightforward computation using (19) shows that

$$\sum_{n=1}^{\infty} P_k^{(\infty)}\left[\log\, Y_n < cn\right] = \sum_{n=1}^{\infty}\left[P_k(X < e^{cn})\right]^n < \infty. \qquad \dots (20)$$

Hence, by the Borel-Cantelli lemma, $\log\, Y_n \geqslant cn$ for all sufficiently large n, with probability one. Since c is arbitrary, (18) is established, and this completes the verification of the example.

In conclusion, the writer wishes to thank his associates D. Basu, C. R. Rao, and V. S. Varadarajan for several discussions and helpful suggestions.

REFERENCES

BASU, D. (1955): An inconsistency of the method of maximum likelihood. *Ann. Math. Stat.*, 26, 144.

KIEFER, J. AND WOLFOWITZ, J. (1956): Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Ann. Math. Stat.*, 27, 887.

KRAFT, C. AND LECAM, L. (1956): A remark on the roots of the maximum likelihood equation. *Ann. Math. Stat.*, 27, 1174.

WALD, A. (1949): Note on the consistency of the maximum likelihood estimate. *Ann. Math. Stat.*, 20, 595.

Paper received : February, 1958.