To see that φ is the best possible, consider the case of the sequence $\{x_n\}$ independently distributed, each taking the values ± 1 with probabilities $\frac{1}{2}$, $\frac{1}{2}$. It follows from a result of Chernoff [3] that

$$\Pr \{x_1 + \cdots + x_n \geq n\epsilon\}^{1/n} \to \varphi \quad \text{as} \quad n \to \infty,$$

so that our φ is exact.

REFERENCES

[1] D. BLACKWELL, "On optimal systems," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 394–397.
[2] J. V. USPENSKY, *Introduction to Mathematical Probability*, McGraw-Hill, 1937.
[3] H. CHERNOFF, "A measure of asymptotic efficiency for test of a hypothesis based on a sum of observations," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 493–507.

A REMARK ON THE ROOTS OF THE MAXIMUM LIKELIHOOD EQUATION

BY C. KRAFT AND L. LECAM[1]

University of California, Berkeley

1. Introduction and summary. The statistical literature combines two types of investigations concerning the consistency of maximum likelihood (M.L.) estimates. A few of these, such as the most excellent paper of A. Wald [1], do prove directly the consistency of M.L. estimates. However, most investigators seem to have concentrated their efforts on proving the existence and consistency of suitably selected roots of the successive likelihood equations. Some authors, see [2], for example, add the supplementary remark that such consistent roots will eventually be unique in suitably small neighborhoods of the true value and will achieve a local maximum.

It is the purpose of the present note to point out by means of examples that this second mode of attack is not adequate. In the examples given below, the "usual regularity conditions" of Cramér [3] or Wald [4] are satisfied, but the M.L. estimates are not consistent. It should also be pointed out that the direct proofs of existence of roots, simple in the case of a unidimensional parameter, become unwieldy in more than one dimension. On the other hand, if one has proved the consistency of the M.L. estimates, the existence of roots follows trivially from the fact that when a differentiable function reaches its maximum in an open set, the derivatives vanish at that point.

2. Examples with independent identically distributed variables. The first example given below has the following characteristics:

Received September 21, 1955.

[1] This paper was prepared with the support of the Office of Ordnance Research, U. S. Army, under Contract DA-04-200-ORD-171.

(1) Cramér's conditions are satisfied and the condition of identifiability is satisfied.

(2) The likelihood equation has roots.

(3) The M.L. estimate does not exist, except maybe for sets of sample points of measure zero.

(4) There exist consistent estimates. (See Section 3, below.)

For every nonnegative integer k, let A_k be the open interval

$$A_k = (2k, 2k + 1).$$

Let $\Theta = \bigcup_{k=1}^{\infty} A_k$, and let $\{a_k\}$, $k = 0, 1, 2, \cdots$, be an arbitrary ordering of the rationals of the interval (1, 2). Define $\rho(\theta) = a_k$ if $\theta \varepsilon A_k$. For each $\theta \varepsilon \Theta$, let the vector (X, Y) have a normal distribution with $E(X \mid \theta) = \rho(\theta) \cos 2\pi\theta$ and $E(Y \mid \theta) = \rho(\theta) \sin 2\pi\theta$, and covariance matrix the identity. If $\{(X_i, Y_i)\}$, $i = 1, 2, \cdots, n \cdots$, is a sequence of independent random vectors with the distribution of (X, Y), the logarithm of the probability density of the first n observation is given by

$$-2 \log p_n = K_n + n[\bar{X}_n - \rho(\theta) \cos 2\pi\theta]^2 + n[\bar{Y}_n - \rho(\theta) \sin 2\pi\theta]^2,$$

where $(\bar{X}_n, \bar{Y}_n)$ is the sample mean.

Defining $r_n > 0$ and φ_n, $0 \leqq \varphi_n < 1$, by $\bar{X}_n = r_n \cos 2\pi\varphi_n$ and $\bar{Y}_n = r_n \sin 2\pi\varphi_n$, the above equation can also be written as

$$-2 \log p_n = n \log 2\pi + [r_n - \rho(\theta)]^2 + 2r_n\rho(\theta)[1 - \cos 2\pi(\theta - \varphi_n)].$$

Accordingly, the likelihood equation is $r_n \sin 2\pi(\theta - \varphi_n) = 0$. Therefore, all values of the form $\theta = \varphi_n + k/2$ which belong to Θ are solutions of the likelihood equation. However, if r_n is not rational, the M.L. estimate does not exist, since $\rho(\theta)$ can be chosen close to r_n but not equal to it.

One could define approximate maximum likelihood estimates as follows. Let $\{\epsilon_n\}$ be a sequence of positive numbers tending to zero. For each n, let S_n be the set of values of θ such that

$$\sup_{t \varepsilon \Theta} p_n(x_1, x_2, \cdots, x_n \mid t) \leqq (1 + \epsilon_n) p_n(x_1, \cdots, x_n \mid \theta).$$

Since every interval, however small, contains an infinity of rationals, for every $\epsilon_n > 0$, the set S_n will, in our example, have elements in common with an infinite number of the intervals A_k, and therefore the sequence $\{S_n\}$ cannot converge to a point.

One might object to the preceding example for two reasons. In the usual proofs of "consistency" of roots of the M.L. equation, it is assumed that the random variables are real-valued. However, this assumption is irrelevant to the proofs given, so that the bivariate character of the example is no detraction. It is, of course, possible to build analogous univariate examples.

Another feature to which objections can be raised is the nonexistence of the M.L. estimate. This is also irrelevant, as shown by the next example, which possesses the following characteristics:

(1) Cramér's conditions and the condition of identifiability are satisfied.

(2) The likelihood equation has roots.

(3) With probability tending to unity, the maximum likelihood estimate exists, is unique, and is a root of the likelihood equation.

(4) The M.L. estimate is not consistent.

(5) There exist consistent estimates.

Let Θ be $\bigcup_k A_k$ as in the first example, and let $\{\alpha_k\}$ be an ordering of the rationals of the interval (0, 1). Let $\rho(\theta) = \alpha_k$ if $\theta \varepsilon A_k$ and let (X_i, Y_i, Z_i) be multinomially distributed with probabilities $p_1 = \rho(\theta) \cos^2 2\pi\theta$, $p_2 = \rho(\theta) \sin^2 2\pi\theta$, and $p_3 = 1 - \rho(\theta)$. For n independent observations, the likelihood function is

$$\log p_n = n_1 \log p_1 + n_2 \log p_2 + n_3 \log p_3 + f(n_1, n_2, n_3),$$

where $n_1 = \sum_{i=1}^n X_i$, $n_2 = \sum_{i=1}^n Y_i$, $n_3 = \sum_{i=1}^n Z_i$, and f is a function which does not depend on θ. Again, the likelihood equation has solutions of the form $2\pi\theta = \tan^{-1}\sqrt{n_2/n_1}$. Since the density is maximized by taking $p_i = n_i/n$, if this is possible, only one of these solutions is the M.L. estimate. With probability tending to unity, the M.L. estimate $\hat{\theta}_n$ is such that $1 - \rho(\hat{\theta}_n) = n_3/n$. For $\hat{\theta}_n$ to be consistent, it must eventually stay in a fixed interval A_k so that $n_3/n = 1 - \alpha_k$, but the probability of this equality tends to zero as n tends to infinity.

3. Existence of consistent estimates. In the discussion of the first and second examples given above, it is stated that there exist consistent estimates. This follows from the lemma stated in the present section.

Consider a situation where the following assumptions are satisfied:

(1) Observations are made on a sequence of independent identically distributed variables $\{X_n\}$, $n = 1, 2, \cdots$, taking their values in a Euclidean space $\mathfrak{X}$.

(2) The parameter space Θ is a measurable subset of a Euclidean space.

(3) To each $\theta \epsilon \Theta$ there corresponds a measure P_θ on $\mathfrak{X}$, and the distribution of the sequence $\{X_n\}$ is the product measure corresponding to a P_θ of the family $\{P_\theta, \theta \epsilon \Theta\}$.

(4) $P_{\theta_1} = P_{\theta_2}$ implies $\theta_1 = \theta_2$.

(5) Θ is a locally compact subset of a Euclidean space and the map $\theta \to P_\theta$ is continuous in the sense that if $\theta_n \to \theta_0$, then $P_{\theta_n} \to P_{\theta_2}$ for Paul Lévy's distance.

One can easily obtain the following proposition:

LEMMA 1. *Let assumptions* (1) *to* (5) *be satisfied. Then there exists a sequence* $\{T_n\}$ *of estimates such that for every positive* ϵ *and every compact set* $K \subset \Theta$ *the quantity*

$$\sup_{\theta \epsilon K} P[\,|\,T_n - \theta\,| > \epsilon \,|\, \theta]$$

tends to zero as n tends to infinity.

The proof of this lemma has been given elsewhere, see [5]. It entails that Cramér's conditions in [3], when supplemented by (4), above, imply the existence of consistent estimates. Even in the simple case described by assumptions (1) to (4), the problem of finding necessary and sufficient conditions for the existence of consistent estimates has not been solved. Partial results have been obtained by C. Stein [6] and Doob [7].

4. Independent, not identically distributed, variables. In the case of independent identically distributed variables, the lemma given in Section 3 ensures the existence of consistent estimates in a wide variety of circumstances. If the variables are not identically distributed, much more freedom is available, as indicated by the next example which possesses the following characteristics:

(1) Wald's conditions [4] and the condition of identifiability are satisfied.

(2) There does not exist any consistent estimate.

Let Θ be the open interval $(0, 3\pi)$. For each θ, let X_{2i} be normal with mean $\cos a_i\theta$ and variance 1 and let X_{2i+1} be normal with mean $\sin \theta$ and variance 1. It can be verified that if a_i tends to unity, Wald's conditions for the existence of consistent roots are satisfied. However, a necessary condition for the existence of consistent estimates given in [8] is not always satisfied. In the present case, this condition would require that for any two values θ_1, θ_2, the quantity

$$n(\sin \theta_1 - \sin \theta_2) + \sum_{i=1}^{n} [\cos a_i\theta_1 - \cos a_i\theta_2]^2$$

increases to infinity. If θ_2 is taken equal to $\theta_1 + 2\pi$ and the a_i's tend to unity sufficiently fast, this condition is not satisfied, although the condition of identifiability can readily be satisfied.

It should be noted that the above is not contradictory to Wald's assertion that there is a sequence of roots which converge to the true parameter value. However there can be, as in this example, more than one limit point to the set of all roots. There is no consistent estimate because it cannot be determined from the sample values alone which convergent subsequences of the roots are the appropriate ones.

REFERENCES

[1] A. WALD, "Note on the consistency of the maximum likelihood estimate," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 595–601.

[2] V. S. HUZURBAZAR, "The likelihood equation, consistency, and the maximum of the likelihood function," *Ann. Eugenics*, Vol. 14 (1948), pp. 185–200.

[3] H. CRAMÉR, *Mathematical Methods of Statistics*, Princeton University Press, 1946.

[4] A. WALD, "Asymptotic properties of the maximum likelihood estimate of an unknown parameter of a discrete stochastic process," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 40–46.

[5] L. LECAM, "On the asymptotic theory of estimation and testing hypotheses," *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, to be published.

[6] C. STEIN, unpublished.

[7] J. DOOB, "Applications of the theory of martingales," *Colloques Internationaux du Centre National de la Recherche Scientifique*, 1949, pp. 23–27.

[8] C. KRAFT, "Some conditions for consistency and uniform consistency of statistical procedures," *Univ. of California Publ. Stat.*, Vol. 2, No. 6 (1955), pp. 125–