DERIVATION OF A BROAD CLASS OF CONSISTENT ESTIMATES

By R. C. Davis

U. S. Naval Ordnance Test Station, Inyokern, California

1. Summary. Given a chance vector $\mathbf{X}$ with distribution function $F(\mathbf{X}, \boldsymbol{\theta}_T)$, where $\boldsymbol{\theta}_T$ denotes the true unknown parameter vector, a broad class of estimates of $\boldsymbol{\theta}_T$ is derived which is shown to be identical with the class of all consistent estimates of $\boldsymbol{\theta}_T$. A sub-class is obtained each member of which has the following properties: a.) Its construction depends upon the solution of an equation involving a single vector function of the parameter vector $\boldsymbol{\theta}$ and the members of a sequence $\{\mathbf{X}_n\}$ of independent and identically distributed chance vectors; b.) the estimate so obtained converges almost certainly to $\boldsymbol{\theta}_T$; c.) it is a symmetric function of the members of the sequence $\{\mathbf{X}_n\}$. In order to obtain this sub-class it is postulated that a function of $\mathbf{X}$ and $\boldsymbol{\theta}$ exists (continuous in $\boldsymbol{\theta}$ for a certain neighborhood of the true parameter $\boldsymbol{\theta}_T$ and existing for each $\mathbf{X}$ in a subset of the sample space) which satisfies a Lipschitz condition in $\boldsymbol{\theta}$. In particular if a density function $f(\mathbf{X}, \boldsymbol{\theta}_T)$ exists satisfying certain conditions, the consistency of the maximum likelihood estimate can be established under regularity conditions quite different from those usually assumed [1]. This is not to be interpreted as a weakening of the usual regularity conditions but rather as an extension of the class of consistent likelihood estimates obtained under the usual regularity conditions.

2. Introduction. The present work is the result of investigations into the following question posed by J. Neyman: What happens to the asymptotic properties of the maximum likelihood estimate of $\boldsymbol{\theta}_T$ when the usual regularity conditions on $F(\mathbf{X}, \boldsymbol{\theta})$ are relaxed? The consistency and efficiency of the estimate are the properties in question, and the present work arose from the observation that consistency at least can be obtained under conditions much different than those usually assumed [1]. The assumptions made below are existential in nature, and no general methods are given for the actual construction of consistent estimates. As stated above, however, the results of this work can be used to widen the class of consistent maximum likelihood estimates established heretofore. Although simple upper and lower bounds for the variance of a consistent estimate are obtained, no answer is given to the question of determining the efficiency of such an estimate. In regard to consistent estimates, J. Neyman and E. Scott have discussed recently [2] the need for a systematic method of obtaining consistent estimates. Wald has given necessary and sufficient conditions [3] for the existence of a uniformly consistent estimate of an unknown parameter $\boldsymbol{\theta}$ when there exists a density function continuous jointly in all of its arguments, and it is assumed that the domain of each of the unknown parameters is a closed and bounded set. It is hoped that the class of consistent estimates derived below will help shed some light on a general method for actually obtaining such estimates. In this connection it is important to point out that if necessary and sufficient conditions were known for the existence and uniqueness of a fixed point for a transformation on E_n to E_n, the weakest possible conditions could be expressed for the existence of consistent estimates obtained in the manner given below. It is surmised that the use of a Hölder condition of order one as presented below is stronger than required.

Let $\{\mathbf{X}_i\}$, $i = 1, 2, \cdots, n, \cdots$, be a sequence of chance vectors in which $\mathbf{X}_i$ possesses the probability distribution function $F_i(\mathbf{X}, \boldsymbol{\theta})$ depending upon an unknown parameter vector $\boldsymbol{\theta}$. The vector $\mathbf{X}$ has components X_i, $i = 1, 2, \cdots, s$, where X_i is a chance variable, and $\boldsymbol{\theta}$ has components θ_j, $j = 1, 2, \cdots, m$. The problem is to obtain a function of the $\mathbf{X}_i$ which is a consistent estimate of $\boldsymbol{\theta}$. We denote by E_s the real Euclidean space of s dimensions and by E_s' a subset of E_s excluding at most a set of probability measure zero. For convenience we use the symbol $\| \theta \|$ to denote the norm of $\boldsymbol{\theta}$, where

$$\| \boldsymbol{\theta} \| = (\theta_1^2 + \theta_2^2 + \cdots + \theta_m^2)^{1/2}.$$

We define in a similar manner the norm of any function which assumes values in E_m. The following assumption is made:

Assumption 1. *There exists a point* $\boldsymbol{\theta}_0$ *and a neighborhood* $W(\boldsymbol{\theta}_0, a)$ *of* $\boldsymbol{\theta}_0$ *having radius* a $(a > 0)$ *which contains the true parameter vector* $\boldsymbol{\theta}_T$ *as an interior point and there exists an infinite sequence of functions* $G_n(\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n; \boldsymbol{\theta})$, $n = 1, 2, \cdots$, *ad inf. on* $E_s \times E_m$ *to* E_m *such that*

(a) *for each* n *the equation*

$$G_n(\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n; \boldsymbol{\theta}) = 0$$

has a unique solution $\boldsymbol{\theta} = \boldsymbol{\theta}_n^*(\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n)$ *in* $W(\boldsymbol{\theta}_0, a)$. (*For the sake of brevity we usually write* $G_n(\mathbf{X}; \boldsymbol{\theta}) = G_n(\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n; \boldsymbol{\theta})$.)

(b) *For every pair of values of* $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2$ *in* $W(\boldsymbol{\theta}_0, a)$ *and for some* K *with* $0 < K < 1$

$$\lim_{n\to\infty} P\{\|G_n(\mathbf{X}, \boldsymbol{\theta}_1) - G_n(\mathbf{X}, \boldsymbol{\theta}_2) - (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2)\| \leqq K \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|\} = 1.$$

(c) *For every* $\epsilon > 0$,

$$\lim_{n\to\infty} P\{\| G_n(\mathbf{X}, \boldsymbol{\theta}_T) \| < \epsilon\} = 1.$$

3. A consistent estimate of $\boldsymbol{\theta}_T$.

Theorem 3.1. *The solution* $\boldsymbol{\theta} = \boldsymbol{\theta}_n^*(\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n)$ *of the equation*

$$G_n(\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n; \boldsymbol{\theta}) = 0$$

is a consistent estimate of $\boldsymbol{\theta}_T$, *providing* $G_n(\mathbf{X}; \boldsymbol{\theta})$ *satisfies Assumption* 1.

Proof: From Assumption 1b it follows that given $\delta > 0$, we have for all $n > N'(\delta)$,

$$P\{\| G_n(\mathbf{X}, \boldsymbol{\theta}_T) - (\boldsymbol{\theta}_T - \boldsymbol{\theta}_n^*) \| \leqq K \| \boldsymbol{\theta}_T - \boldsymbol{\theta}_n^* \|\} > 1 - \frac{\delta}{2}, \tag{3.1}$$

since $G_n(\mathbf{X}, \boldsymbol{\theta}_n^*) = 0$. It follows from (3.1) that for all $n > N'(\delta)$,

$$P\left\{\frac{|| G_n(\mathbf{X}, \boldsymbol{\theta}_T) ||}{1 + K} \leqq || \boldsymbol{\theta}_T - \boldsymbol{\theta}_n^* || \leqq \frac{|| G_n(\mathbf{X}, \boldsymbol{\theta}_T) ||}{1 - K}\right\} > 1 - \frac{\delta}{2}. \tag{3.2}$$

From Assumption 1c it follows that there exists $N''(\epsilon, \delta)$ such that $n > N''(\epsilon, \delta)$ implies

$$P\{|| G_n(\mathbf{X}, \boldsymbol{\theta}_T) || < \epsilon(1 - K)\} > 1 - \frac{\delta}{2}. \tag{3.3}$$

(3.2), (3.3), and a familiar formula in probability imply for all

$$n > \max [N'(\delta), N''(\epsilon, \delta)],$$
$$P\{|| \boldsymbol{\theta}_T - \boldsymbol{\theta}_n^* || < \epsilon\} > 1 - \delta.$$

It is noted that (3.2) characterizes the speed of convergence of the estimate $\boldsymbol{\theta}_n^*$. The following uniqueness property is noted: *If a given sequence of functions $G_n(\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n; \boldsymbol{\theta})$ satisfies Assumption 1, then $\boldsymbol{\theta}_T$ is the unique parameter vector in $W(\boldsymbol{\theta}_0, a)$ which satisfies item* c *of Assumption* 1. The proof of this remark is left to the reader.

The following remark demonstrates the extreme generality of the class of consistent estimates obtained in the above manner: *The set of estimates of the parameter vector $\boldsymbol{\theta}_T$ obtained from the class of all sequences of functions*

$$G_n(\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n; \boldsymbol{\theta})$$

satisfying Assumption 1 *is identical with the set of all consistent estimates of the parameter vector $\boldsymbol{\theta}_T$.* The proof of this remark is quite obvious and is left to the reader.

4. Properties of a sub-class of consistent estimates. The question arises naturally concerning a general method for the construction of a sequence of functions $G_n(\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n; \boldsymbol{\theta})$ satisfying Assumption 1. The author knows of no general method. It is possible to describe a sub-class of the class of consistent estimates, the construction of which depends upon the existence of one function rather than a sequence of functions. This is possible by application of the strong law of large numbers, and in this way consistent estimates of the parameter vector are obtained which converge almost certainly to the true value $\boldsymbol{\theta}_T$. Moreover it is clear that under certain conditions the function

$$G_n(\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n; \boldsymbol{\theta}_T)$$

defined as in equation 4.1 below is an asymptotically m-variate normal variable

ASSUMPTION 2. *Let $\{\mathbf{X}_i\}, i = 1, 2, \cdots, n, \cdots$, be a sequence of independently and identically distributed chance vectors with common distribution function $F(\mathbf{X}; \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is again the unknown parameter vector.*

ASSUMPTION 3. *There exists a function $g(\mathbf{X}, \boldsymbol{\theta})$ on $E_s \times E_m$ to E_m such that* (a) *for every $\mathbf{X} \in E_s'$ and every distinct pair $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ in $W(\boldsymbol{\theta}_0, a)$,*

$$|| g(\mathbf{X}, \boldsymbol{\theta}_1) - g(\mathbf{X}, \boldsymbol{\theta}_2) - (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2) || \leqq K || \boldsymbol{\theta}_1 - \boldsymbol{\theta}_2 ||,$$

where $0 < K < 1$ *and* $\| g(\mathbf{X}, \boldsymbol{\theta}_0) \| < (1 - K)a$.

(b) $Eg(\mathbf{X}, \boldsymbol{\theta}_T) = \int_{-\infty}^{\infty} g(\mathbf{X}, \boldsymbol{\theta}_T)\, dF(\mathbf{X}, \boldsymbol{\theta}_T) = 0$.

We define the function $G_n(\mathbf{X}, \boldsymbol{\theta})$ as follows:

$$G_n(\mathbf{X}, \boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^{n} g(\mathbf{X}_i, \boldsymbol{\theta}). \tag{4.1}$$

The following lemmas are required:

LEMMA 4.1. *$G_n(\mathbf{X}, \boldsymbol{\theta})$ as defined in* (4.1) *satisfies the conditions in Assumption* 3 *with $G_n(\mathbf{X}, \boldsymbol{\theta})$ replacing $g(\mathbf{X}, \boldsymbol{\theta})$.*

The proof is sufficiently obvious to be omitted.

LEMMA 4.2. *$G_n(\mathbf{X}, \boldsymbol{\theta}_T) \to 0$ almost certainly as $n \to \infty$, if Assumptions* 2 *and* 3b *hold.*

PROOF: Since $Eg(\mathbf{X}_i, \boldsymbol{\theta}_T) = 0$, $i = 1, 2, \cdots, n$, and the chance variables $g(\mathbf{X}_i, \boldsymbol{\theta}_T)$ are independently and identically distributed, this follows immediately from a theorem due to Kolmogorov [5].

THEOREM 4.1. *If Assumptions* 2 *and* 3 *hold, then the equation $G_n(\mathbf{X}, \boldsymbol{\theta}) = 0$ has a unique solution $\boldsymbol{\theta} = \boldsymbol{\theta}_n^*(\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n)$ in $W(\boldsymbol{\theta}_0, a)$, where $\boldsymbol{\theta}_n^*$ is a consistent estimate of $\boldsymbol{\theta}_T$ and is moreover a symmetric function of the observation vectors $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n$.*

PROOF: We obtain the solution $\boldsymbol{\theta}_n^*$ by the method of successive substitutions. Define

$$\boldsymbol{\theta}_1 = \boldsymbol{\theta}_0 - G_n(\mathbf{X}, \boldsymbol{\theta}_0), \cdots, \boldsymbol{\theta}_{q+1} = \boldsymbol{\theta}_q - G_n(\mathbf{X}, \boldsymbol{\theta}_q).$$

In view of Lemma 4.1 we can apply a well known existence theorem [4] in the theory of functions to prove that the sequence $\{\boldsymbol{\theta}_q\}$ converges to a limit $\boldsymbol{\theta}_n^*$ which is also in $W(\boldsymbol{\theta}_0, a)$. The same theorem establishes the uniqueness of the solution in $W(\boldsymbol{\theta}_0, a)$. This uniqueness property together with lemmas 4.1 and 4.2 establish the fact that the sequence $\{G_n(\mathbf{X}, \boldsymbol{\theta})\}$ as defined in equation (4.1) satisfies Assumption 1. It follows immediately from Theorem 3.1 that $\boldsymbol{\theta}_n^*$ is a consistent estimate of $\boldsymbol{\theta}_T$. We can, however, prove a stronger relationship.

THEOREM 4.2. *The estimate $\boldsymbol{\theta}_n^*$ defined in Theorem* 4.1 *converges almost certainly to $\boldsymbol{\theta}_T$.*

PROOF: From Lemma 4.2 we know that given any number $\epsilon > 0$, there exists an integer $N(\epsilon)$ such that for all $n > N(\epsilon)$

$$P\{\| G_n(\mathbf{X}, \boldsymbol{\theta}_T) \| < \epsilon(1 - K)\} = 1.$$

From Assumption 3a and Lemma 4.1 we see that

$$\| G_n(\mathbf{X}, \boldsymbol{\theta}_T) - (\boldsymbol{\theta}_T - \boldsymbol{\theta}_n^*) \| \leqq K \| \boldsymbol{\theta}_T - \boldsymbol{\theta}_n^* \|,$$

since $G_n(\mathbf{X}, \boldsymbol{\theta}_n^*) = 0$. Then

$$\| G_n(\mathbf{X}, \boldsymbol{\theta}_T) \| \geqq (1 - K) \| \boldsymbol{\theta}_T - \boldsymbol{\theta}_n^* \|.$$

Clearly the set of $\mathbf{X} \epsilon E_s'$ for which $\| \boldsymbol{\theta}_T - \boldsymbol{\theta}_n^* \| < \epsilon$ includes the set of $\mathbf{X}$ for which $\| G_n(\mathbf{X}, \boldsymbol{\theta}_T) \| < \epsilon(1 - K)$.

Therefore, for $n > N(\epsilon)$,

$$P\{\| \boldsymbol{\theta}_T - \boldsymbol{\theta}_n^* \| < \epsilon\} \geqq P\{\| G_n(\mathbf{X}, \boldsymbol{\theta}_T) \| < \epsilon(1 - K)\} = 1,$$

and the proof is completed.

The uniqueness of the parameter value $\boldsymbol{\theta}_T$ in the neighborhood $W(\boldsymbol{\theta}_0, a)$ follows immediately from the remark succeeding Theorem 3.1 since Assumption 1 is valid in Theorems 4.1 and 4.2.

It is interesting to note that the application of a theorem in the theory of functions of a real variable gives the result that if the function $g(\mathbf{X}, \boldsymbol{\theta})$ is continuous on a bounded and closed set in $E_s \times E_m$ and if we take for E_s' a bounded and closed set, then $\boldsymbol{\theta}_n^*(\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n)$ is a continuous function of $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n$ for $\mathbf{X}_i \epsilon E_s'$ $(i = 1, 2, \cdots, n)$. If we assume the continuity of $g(\mathbf{X}, \boldsymbol{\theta})$ in $\mathbf{X}$ for each $\boldsymbol{\theta}$ in $W(\boldsymbol{\theta}_0, a)$ the following remark demonstrates an interesting relationship concerning the uniqueness of the solution for $\boldsymbol{\theta}$ in the equation $Eg(\mathbf{X}, \boldsymbol{\theta}) = 0$: *If in addition to Assumption 3 we assume that $g(\mathbf{X}, \boldsymbol{\theta})$ is continuous in $\mathbf{X}$ for every $\mathbf{X}$ in E_s and every $\boldsymbol{\theta}$ in $W(\boldsymbol{\theta}_0, a)$ and if at least one of the components $g_i(\mathbf{X}, \boldsymbol{\theta})$, $1 \leqq i \leqq m$ of the m-dimensional vector function $g(\mathbf{X}, \boldsymbol{\theta})$ satisfies also a Lipschitz condition:*

$$\| g_i(\mathbf{X}, \boldsymbol{\theta}_1) - g_i(\mathbf{X}, \boldsymbol{\theta}_2) - (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2) \| \leqq K \| \boldsymbol{\theta}_1 - \boldsymbol{\theta}_2 \|$$

for every distinct pair $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2$ in $W(\boldsymbol{\theta}_0, a)$, then for all $\boldsymbol{\theta}$ in $W(\boldsymbol{\theta}_0, a)$, $\boldsymbol{\theta}_T$ is the unique solution for $\boldsymbol{\theta}$ of the equation $Eg(\mathbf{X}, \boldsymbol{\theta}) = 0$.
The proof of this remark is left to the reader.

5. Upper and lower bounds for the expected squared error of $\boldsymbol{\theta}_n^*(\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n)$. Denote by $g_i(\mathbf{X}, \boldsymbol{\theta}), i = 1, 2, \cdots, m$, the m components of the chance vector $g(\mathbf{X}, \boldsymbol{\theta})$. We now make an additional assumption.

ASSUMPTION 4.

$$E[g_i(\mathbf{X}, \boldsymbol{\theta}_T)g_j(\mathbf{X}, \boldsymbol{\theta}_T)] = \lambda_{ij}$$

exists for $i = 1, 2, \cdots, m$ and $j = 1, 2, \cdots, m$.

It follows from Assumptions 2, 3b, and 4 and the Lindeberg-Lévy form of the Central Limit Theorem that the vector $\sqrt{n}G_n(\mathbf{X}, \boldsymbol{\theta}_T)$ tends in probability to an m-variate normal distribution with means zero and moment matrix (λ_{ij}).

Now from Assumption 3a and Lemma 4.1

$$\frac{\| G_n(\mathbf{X}, \boldsymbol{\theta}_T) \|}{1 + K} \leqq \| \boldsymbol{\theta}_n^* - \boldsymbol{\theta}_T \| \leqq \frac{\| G_n(\mathbf{X}, \boldsymbol{\theta}_T) \|}{1 - K}. \tag{5.1}$$

For convenience define

$$\lambda = \sum_{i=1}^{m} \lambda_{ii}.$$

We obtain then

$$E \| G_n(\mathbf{X}, \boldsymbol{\theta}_T) \|^2 = \frac{\lambda}{n}.$$

It follows then from equation (5.1) that

$$\frac{\lambda}{n(1 + K)^2} \leqq E \| \boldsymbol{\theta}_n^* - \boldsymbol{\theta}_T \|^2 \leqq \frac{\lambda}{n(1 - K)^2}.$$

6. The consistency of maximum likelihood estimates. The results of this paper can be used to extend the class of consistent maximum likelihood estimates established heretofore [1].[1] Assume that $F(\mathbf{X}, \boldsymbol{\theta})$ admits a density function $f(\mathbf{X}, \boldsymbol{\theta})$ with the property

$$\frac{\partial}{\partial \boldsymbol{\theta}} \int_{-\infty}^{\infty} f(\mathbf{X}, \boldsymbol{\theta})\, d\mathbf{X} = \int_{-\infty}^{\infty} \frac{\partial f}{\partial \boldsymbol{\theta}} (\mathbf{X}, \boldsymbol{\theta})\, d\mathbf{X}.$$

Then

$$E\left[\frac{\partial}{\partial \boldsymbol{\theta}} \ln f(\mathbf{X}, \boldsymbol{\theta})\right] = 0.$$

The maximum likelihood estimate of $\boldsymbol{\theta}_T$ is obtained by solving the equation

$$\frac{\partial}{\partial \theta} \ln L(\mathbf{X}, \boldsymbol{\theta}) = 0,$$

where

$$L(\mathbf{X}, \boldsymbol{\theta}) = \prod_{i=1}^{n} f(\mathbf{X}_i, \boldsymbol{\theta}).$$

If a sample $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n$ is obtained as the result of n random independent drawings from the distribution having the c.d.f. $F(\mathbf{X}, \boldsymbol{\theta})$, the sample values will satisfy Assumption 2. Assumption 3b holds as assumed above. If we assume also that the function $\partial/\partial\boldsymbol{\theta} \ln f(\mathbf{X}, \boldsymbol{\theta})$ satisfies Assumption 3a, it follows directly from Theorem 4.2 that the maximum likelihood estimate converges almost certainly to the true parameter vector as the sample size approaches infinity.

The author wishes to acknowledge his indebtedness and gratitude to Professor Jerzy Neyman for the many helpful suggestions made during the preparation of the paper.

REFERENCES

[1] J. L. Doob, "Probability and statistics," *Trans. Am. Math. Soc.*, Vol. 36 (1934), p. 759.

[2] J. Neyman and Elizabeth L. Scott, "Consistent estimates based on partially consistent observations," *Econometrica*, Vol. 16 (1948), pp. 1–32.

[3] A. Wald, "Estimation of a parameter when the number of unknown parameters increases indefinitely with the number of observations," *Annals of Math. Stat.*, Vol. 19 (1948), pp. 220–227.

[1] Recently Wald [6] and Wolfowitz [7] have discussed the consistency of the maximum likelihood estimate from another approach than the one employed by Doob.

[4] L. M. GRAVES, *The Theory of Functions of Real Variables*, McGraw-Hill Book Co., 1946.
[5] A. KOLMOGOROFF, *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Chelsea Publishing Co., 1946.
[6] A. WALD, "Note on the consistency of the maximum likelihood estimate," *Annals of Math. Stat.*, Vol. 20 (1949), pp. 595–600.
[7] J. WOLFOWITZ, "On Wald's proof of the consistency of the maximum likelihood estimate," *Annals of Math. Stat.*, Vol. 20 (1949), pp. 601–602.