ASYMPTOTIC PROPERTIES OF THE MAXIMUM LIKELIHOOD ESTIMATE OF AN UNKNOWN PARAMETER OF A DISCRETE STOCHASTIC PROCESS

By Abraham Wald

Columbia University

Summary. Asymptotic properties of maximum likelihood estimates have been studied so far mainly in the case of independent observations. In this paper the case of stochastically dependent observations is considered. It is shown that under certain restrictions on the joint probability distribution of the observations the maximum likelihood equation has at least one root which is a consistent estimate of the parameter θ to be estimated. Furthermore, any root of the maximum likelihood equation which is a consistent estimate of θ is shown to be asymptotically efficient. Since the maximum likelihood estimate is always a root of the maximum likelihood equation, consistency of the maximum likelihood estimate implies its asymptotic efficiency.

1. Introduction. Let $\{X_i\}$, $(i = 1, 2, \cdots, \text{ad. inf.})$, be a sequence of chance variables. It is assumed that for any positive integral value n the first n chance variables $X_1, \cdots, X_n$ admit a joint probability density function $p_n(x_1, \cdots, x_n, \theta)$ involving an unknown parameter θ. The consistency relations

$$\int_{-\infty}^{+\infty} p_{n+1}(x_1, \cdots, x_{n+1}, \theta)\, dx_{n+1} = p_n(x_1, \cdots, x_n, \theta) \tag{1.1}$$

are assumed to hold.

In what follows, for any chance variable u the symbol $E(u \mid \theta)$ will denote the expected value of u when θ is the true parameter value.

Let $t_n(x_1, \cdots, x_n)$ be an unbiassed estimate of θ. Cramér [1] and Rao [2] have shown that under some weak regularity conditions on the distribution function $p_n(x_1, \cdots, x_n, \theta)$, the variance of t_n cannot fall short of the value

$$\frac{1}{c_n(\theta)} = \frac{1}{E\left[\left(\frac{\partial \log p_n}{\partial \theta}\right)^2 \Big| \theta\right]}. \tag{1.2}$$

Thus, for any unbiassed estimate t_n the variate $\sqrt{c_n(\theta)}(t_n - \theta)$ has mean value zero and variance $\geqq 1$. An estimate t_n is called efficient if $\sqrt{c_n(\theta)}(t_n - \theta)$ has mean value zero and variance 1.

A sequence $\{t_n\}$, $(n = 1, 2, \cdots, \text{ad. inf.})$, of estimates is said to be asymptotically efficient if the mean of $\sqrt{c_n(\theta)}\,(t_n - \theta)$ is zero and the variance of $\sqrt{c_n(\theta)}\,(t_n - \theta)$ is 1 in the limit as $n \to \infty$. In the literature usually the additional requirement is made that the limiting distribution of $\sqrt{c_n(\theta)}\,(t_n - \theta)$ be normal.

To make a distinction between the two cases when the condition concerning the limiting distribution of $\sqrt{c_n(\theta)}\ (t_n - \theta)$ is fulfilled or not, we shall say that $\{t_n\}$ is asymptotically efficient in the wide sense if it satisfies the conditions concerning the mean and the variance of $\sqrt{c_n(\theta)}\ (t_n - \theta)$. If, in addition, the limiting distribution of $\sqrt{c_n(\theta)}\ (t_n - \theta)$ is normal, we shall say that $\{t_n\}$ is asymptotically efficient in the strict sense. Clearly, if $\{t_n\}$ is asymptotically efficient in the strict sense, it is also asymptotically efficient in the wide sense.

A word of clarification is needed as to the meaning of the conditions concerning the mean and variance of $\sqrt{c_n(\theta)}\ (t_n - \theta)$. One interpretation would be that the requirement is that

(1.3) $$\lim_{n=\infty} E[\sqrt{c_n(\theta)}\ (t_n - \theta) \mid \theta] = 0$$

and

(1.4) $$\lim_{n=\infty} E[c_n(\theta)\ (t_n - \theta)^2 \mid \theta] = 1.$$

Another interpretation would be that the requirement is that the limiting distribution of $\sqrt{c_n(\theta)}\ (t_n - \theta)$, provided that the limit distribution exists as $n \rightarrow \infty$, should have zero mean and unit variance. These two interpretations are certainly not equivalent. It seems to the author that the mean and variance of the limiting distribution is more relevant than the limits of the mean and the variance. We shall, therefore, adopt the following definition of asymptotic efficiency:

Definition: A sequence $\{t_n\}$ of estimates is said to be asymptotically efficient in the wide sense if a sequence $\{u_n\}$, $(n = 1, 2, \cdots$, ad. inf.), of chance variables exists such that

(1.5) $$\lim_{n=\infty} E(u_n \mid \theta) = 0, \qquad \lim_{n=\infty} E(u_n^2 \mid \theta) = 1$$

and

(1.6) $$\sqrt{c_n(\theta)}(t_n - \theta) - u_n$$

converges stochastically to zero as $n \rightarrow \infty$. If, in addition, the limiting distribution of $\sqrt{c_n(\theta)}\ (t_n - \theta)$ exists and is normal, $\{t_n\}$ is said to be asymptotically efficient in the strict sense.

The reason that a sequence $\{u_n\}$ of chance variables is considered in the above definition, instead of the limiting distribution of $\sqrt{c_n(\theta)}\ (t_n - \theta)$, is that the existence of a limiting distribution of $\sqrt{c_n(\theta)}\ (t_n - \theta)$ is not postulated. If a limiting distribution of $\sqrt{c_n(\theta)}\ (t_n - \theta)$ exists and if this limiting distribution has zero mean and unit variance, a sequence $\{u_n\}$ of chance variables satisfying the conditions (1.5) and (1.6) always exists. This can be seen as follows: Let T_n denote the chance variable $\sqrt{c_n(\theta)}\ (t_n - \theta)$ and let $F_n(t) =$ prob. $\{T_n < t\}$. If a limiting distribution of T_n exists and if this limiting distribution has zero mean and unit variance, then

$$(1.7) \qquad \lim_{a=\infty}\left[\lim_{n=\infty}\int_{-a}^{a} t\,dF_n(t)\right] = 0 \qquad \text{and} \qquad \lim_{a=\infty}\left[\lim_{n=\infty}\int_{-a}^{a} t^2\,dF_n(t)\right] = 1.$$

From (1.7) it follows that there exists a sequence $\{a_n\}$, $(n = 1, 2, \cdots$, ad. inf.), of positive values such that the following conditions are fulfilled:

$$(1.8) \quad \lim_{n=\infty}\int_{-a_n}^{a_n} t\,dF_n(t) = 0; \; \lim_{n=\infty}\int_{-a_n}^{a_n} t^2\,dF_n(t) = 1; \; \lim_{n=\infty} \text{Prob}\,\{|\,T_n\,| > a_n\} = 0.$$

Let u_n be a chance variable which is equal to T_n whenever $|\,T_n\,| \leqq a_n$, and equal to zero otherwise. Clearly, the sequence $\{u_n\}$ will satisfy conditions (1.5) and (1.6).

In the following section we shall formulate some assumptions concerning the probability density function $p_n(x_1, \cdots, x_n, \theta)$. It will then be shown in section 3 that there exists a root of the maximum likelihood equation

$$(1.9) \qquad \frac{\partial \log p_n}{\partial \theta} = 0$$

which is asymptotically efficient at least in the wide sense.

2. Assumptions concerning the probability density $p_n(x_1, \cdots, x_n, \theta)$. We shall assume that there exists a finite non-degenerate interval A on the θ-axis such that the following conditions hold:

Condition 1. The derivatives $\dfrac{\partial^i p_n}{\partial \theta^i}$, $(i = 1, 2, 3)$, exist for all θ in A and for all samples $(x_1, \cdots, x_n)$ except perhaps for a set of measure zero. We have furthermore,

$$(2.1) \qquad \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \underset{\theta \in A}{\text{l.u.b.}} \left|\frac{\partial^i p_n}{\partial \theta^i}\right| dx_1 \cdots dx_n < \infty, \qquad (i = 1, 2).$$

Condition 2. For any θ in A we have $\lim_{n=\infty} c_n(\theta) = \infty$.

Condition 3. For any θ in A the standard deviation of $\dfrac{\partial^2 \log p_n}{\partial \theta^2}$ divided by the expected value of $\dfrac{\partial^2 \log p_n}{\partial \theta^2}$ (both computed under the assumption that θ is true) converges to zero as $n \to \infty$.

Condition 4. There exists a positive δ such that for any θ in A the expression

$$(2.2) \qquad \frac{1}{c_n(\theta)} E\left[\underset{\theta'}{\text{l.u.b.}} \left|\frac{\partial^3 \log p_n(x_1, \cdots, x_n, \theta')}{\partial \theta'^3}\right| \,\Big|\, \theta\right]$$

is a bounded function of n where θ' is restricted to the interval $|\,\theta' - \theta\,| \leqq \delta$.

In what follows in this section, as well as in section 3, the domain of θ will be restricted to interior points of the interval A unless a statement to the contrary is explicitly made.

Clearly

$$E\left(\frac{\partial \log p_n}{\partial \theta} \,\middle|\, \theta\right) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \frac{\partial p_n}{\partial \theta} \, dx_1 \cdots dx_n. \tag{2.3}$$

It follows from Condition 1 that

$$\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \frac{\partial p_n}{\partial \theta} \, dx_1 \cdots dx_n = \frac{\partial}{\partial \theta} \int_{-\infty}^{+\infty} p_n \, dx_1 \cdots dx_n = 0. \tag{2.4}$$

Hence,

$$E\left(\frac{\partial \log p_n}{\partial \theta} \,\middle|\, \theta\right) = 0. \tag{2.5}$$

We have

$$\frac{\partial^2 \log p_n}{\partial \theta^2} = \frac{1}{p_n} \frac{\partial^2 p_n}{\partial \theta^2} - \left(\frac{\partial \log p_n}{\partial \theta}\right)^2. \tag{2.6}$$

Hence

$$E\left(\frac{\partial^2 \log p_n}{\partial \theta^2} \,\middle|\, \theta\right) = E\left(\frac{1}{p_n} \frac{\partial^2 p_n}{\partial \theta^2} \,\middle|\, \theta\right) - c_n(\theta). \tag{2.7}$$

But

$$E\left(\frac{1}{p_n} \frac{\partial^2 p_n}{\partial \theta^2} \,\middle|\, \theta\right) = 0, \tag{2.8}$$

because of Condition 1. From (2.7) and (2.8) we obtain

$$E\left(\frac{\partial^2 \log p_n}{\partial \theta^2} \,\middle|\, \theta\right) = -c_n(\theta). \tag{2.9}$$

Conditions 3 and 4 will generally be fulfilled when the stochastic dependence of x_j on x_i decreases sufficiently fast with increasing value of $| i - j |$. For, in such cases, the following order relations will generally hold: The standard deviation of $\frac{\partial^2 \log p_n}{\partial \theta^2}$ will, in general, be of the order $\sqrt{n}$, the expected value of

$$\underset{|\theta' - \theta| \leq \delta}{\text{l.u.b.}} \left| \frac{\partial^3 \log p_n}{\partial \theta'^3} \right|$$

will usually be of the order n, and $\frac{c_n(\theta)}{n}$ will generally have a positive lower bound and a finite upper bound.

3. Proof that the maximum likelihood equation has a root which is an asymptotically efficient estimate of θ (at least in the wide sense). Let θ_0 denote the true parameter value and let θ be any other value. We put

$$\frac{\partial \log p_n}{\partial \theta} = \Phi_n, \quad \frac{\partial^2 \log p_n}{\partial \theta^2} = \Phi_n' \quad \text{and} \quad \frac{\partial^3 \log p_n}{\partial \theta^3} = \Phi_n''. \tag{3.1}$$

Expanding $\Phi_n(x_1, \cdots, x_n, \theta)$ in a Taylor expansion around $\theta = \theta_0$ we obtain

$$\begin{aligned} \Phi_n(x_1, \cdots, x_n, \theta) &= \Phi_n(x_1, \cdots, x_n, \theta_0) + (\theta - \theta_0)\Phi_n'(x_1, \cdots, x_n, \theta_0) \\ &\quad + \tfrac{1}{2}(\theta - \theta_0)^2\Phi_n''(x_1, \cdots, x_n, \theta_n^*) \end{aligned} \tag{3.2}$$

where θ_n^* is some value between θ_0 and θ. Dividing both sides of (3.2) by $c_n(\theta_0)$ we obtain

$$\begin{aligned} \frac{\Phi_n(x_1, \cdots, x_n, \theta)}{c_n(\theta_0)} &= \frac{\Phi_n(x_1, \cdots, x_n, \theta_0)}{c_n(\theta_0)} \\ &\quad + (\theta - \theta_0)\frac{\Phi_n'(x_1, \cdots, x_n, \theta_0)}{c_n(\theta_0)} + \tfrac{1}{2}(\theta - \theta_0)^2\frac{\Phi''(x_1, \cdots, x_n, \theta_n^*)}{c_n(\theta_0)}. \end{aligned} \tag{3.3}$$

From Condition 3 and equation (2.9) it follows that

$$\operatorname*{plim}_{n=\infty} \frac{\Phi_n'(x_1, \cdots, x_n, \theta_0)}{c_n(\theta_0)} = -1 \tag{3.4}$$

where the operator plim stands for convergence in probability (stochastic convergence).

According to equation (2.5) the expected value of $\Phi_n(x_1, \cdots, x_n, \theta_0)$ is zero. Since the variance of $\Phi_n(x_1, \cdots, x_n, \theta_0)$ is equal to $c_n(\theta_0)$, and since $\lim_{n=\infty} c_n(\theta) = \infty$, we have

$$\operatorname*{plim}_{n=\infty} \frac{\Phi_n(x_1, \cdots, x_n, \theta_0)}{c_n(\theta_0)} = 0. \tag{3.5}$$

It follows from Condition 4 that for any θ with $|\theta - \theta_0| \leqq \delta$ we have

$$\frac{1}{c_n(\theta_0)} E(|\Phi_n''(x_1, \cdots, x_n, \theta_n^*|) = 0(1). \tag{3.6}$$

According to Markoff's inequality the probability that a positive random variable will exceed λ-times its expected value is not greater than $\frac{1}{\lambda}$. Hence, it follows from (3.6) that for any $\epsilon > 0$ we can find a positive value k_ϵ such that

$$\limsup_{n=\infty} \text{Prob}\left\{\frac{1}{c_n(\theta_0)} |\Phi_n''(x_1, \cdots, x_n, \theta_n^*)| \geqq k_\epsilon\right\} \leqq \epsilon. \tag{3.7}$$

Let ρ be any given positive number. The probability that the maximum likelihood equation

$$\Phi_n(x_1, \cdots, x_n, \theta) = 0 \tag{3.8}$$

will have a root in the interval $(\theta_0 - \rho, \theta_0 + \rho)$ converges to one as $n \to \infty$. This follows easily from (3.3), (3.4), (3.5) and (3.7). Thus, we have shown that the maximum likelihood equation has a root $\bar{\theta}_n$ which is a consistent estimate, i.e. it satisfies the relation

$$\text{plim}\,(\bar{\theta}_n - \theta_0) = 0. \tag{3.9}$$

We shall now show that if $\bar{\theta}_n$ is a root of the maximum likelihood equation (3.8) and if $\bar{\theta}_n$ is a consistent estimate, then $\bar{\theta}_n$ is also asymptotically efficient, at least in the wide sense. For this purpose we substitute $\bar{\theta}_n$ for θ in (3.3) and multiply both sides of the equation by $\sqrt{c_n(\theta_0)}$. We then obtain

$$0 = \frac{\Phi_n(x_1, \cdots, x_n, \theta_0)}{\sqrt{c_n(\theta_0)}} + \sqrt{c_n(\theta_0)}\,(\bar{\theta}_n - \theta_0)\,\frac{\Phi'_n(x_1, \cdots, x_n, \theta_0)}{c_n(\theta_0)} + \sqrt{c_n(\theta_0)}\,(\bar{\theta}_n - \theta_0)^2\, v_n \tag{3.10}$$

where

$$v_n = \frac{1}{2}\,\frac{\Phi''_n(x_1, \cdots, x_n, \theta_n^*)}{c_n(\theta_0)}. \tag{3.11}$$

Let

$$y_n = \frac{\Phi_n(x_1, \cdots, x_n, \theta_0)}{\sqrt{c_n(\theta_0)}} \quad \text{and} \quad z_n = \sqrt{c_n(\theta_0)}\,(\bar{\theta}_n - \theta_0). \tag{3.12}$$

Then (3.10) given

$$-y_n = z_n\,\frac{\Phi'_n(x_1, \cdots, x_n, \theta_0)}{c_n(\theta_0)} + z_n(\bar{\theta}_n - \theta_0)\, v_n. \tag{3.13}$$

It follows from (3.7) and (3.9) that

$$\operatorname*{plim}_{n=\infty}\,(\bar{\theta}_n - \theta_0)\, v_n = 0. \tag{3.14}$$

From (3.4), (3.13) and (3.14) we obtain

$$-\,y_n = z_n(-\,1 + \xi_n) \tag{3.15}$$

where

$$\operatorname*{plim}_{n=\infty}\,\xi_n = 0. \tag{3.16}$$

Since $Ey_n = 0$ and $Ey_n^2 = 1$, it follows from (3.15) and (3.16) that

$$\operatorname*{plim}_{n=\infty}\,(z_n - y_n) = 0. \tag{3.17}$$

The asymptotic efficiency (in the wide sense) of $\bar{\theta}_n$ is an immediate consequence of (3.17). Our main result may be summarized in the following theorem:

THEOREM. *If the true value of the parameter θ is an interior point of an inter-*

val A satisfying the conditions 1 — 4, then the maximum likelihood equation (1.9) *has a root*[1] *which is a consistent estimate of* θ. *Furthermore, any root of* (1.9) *which is a consistent estimate of* θ *is also asymptotically efficient at least in the wide sense.*

Since the maximum likelihood estimate is a root of (1.9), it follows from the above theorem that whenever the maximum likelihood estimate is consistent, it is also asymptotically efficient at least in the wide sense.

REFERENCES

[1] H. Cramér, *Mathematical Methods of Statistics*, Princeton Univ. Press, 1946.
[2] C. R. Rao, "Information and the accuracy attainable in the estimation of statistical parameters", *Bull. Calcutta Math. Soc.*, Vol. 37 (1945).

[1] The probability that (1.9) has at least one root converges to unity as $n \to \infty$.