The Annals of Statistics
1982, Vol. 10, No. 1. 167–173

SEQUENTIAL ESTIMATION THROUGH ESTIMATING EQUATIONS IN THE NUISANCE PARAMETER CASE

BY PEDRO E. FERREIRA[1]

CIENES, Organization of American States

Let $(X_1, X_2, \cdots)$ be a sequence of random variables and let the p.d.f. of $\mathbf{X}_n = (X_1, \cdots, X_n)$ be $p(\mathbf{x}_n, \theta)$, where $\theta = (\theta_1, \theta_2)$. An estimating equation rule for θ_1 is a sequence of functions $g(x_1, \theta_1)$, $g(x_1, x_2, \theta_1)$, $\cdots$. If the random sample size $N = n$, we estimate θ_1 through the estimating equation $g(\mathbf{X}_n, \theta_1) = 0$. In this paper, optimum estimation rules are obtained and, in particular, sufficient conditions for the optimality of the maximum conditional likelihood estimation rule are given. In addition, Bhapkar's concept of information in an estimating equation is used to discuss stopping criteria.

1. Introduction. Let $(X_1, X_2, \cdots)$ be a sequence of random variables with an associated sample space $\mathscr{X} = \mathscr{X}_1 \times \mathscr{X}_2 \times \cdots$, where each $\mathscr{X}_n$ is assumed to be a Euclidean space. Let B_n be the Borel sigma-field of subsets of $\mathscr{X}^n = \mathscr{X}_1 \times \cdots \times \mathscr{X}_n$ generated by $\mathbf{X}_n = (X_1, \cdots, X_n)$, and let P_θ^n, $\theta \in \Theta$, be a family of probability measures on $(\mathscr{X}^n, B_n)$. Further let $p_n(\mathbf{x}_n, \theta)$ be the probability density function of $\mathbf{X}_n$; i.e. the density of P_θ^n.

A stopping rule for a sequential procedure will be defined as a sequence of sets $R_n \in B_n$, $n = 1, 2, \cdots$, and the *sample size*, N, is the least integer n such that $\mathbf{X}_n \in R_n$. The subset of points in $\mathscr{X}^n$ corresponding to a stop at $N = n$ is denoted by $\tilde{R}_n$, where

$$\tilde{R}_n = \begin{cases} R_1 & \text{if} \quad n = 1 \\ \{\mathbf{x}_n : \mathbf{x}_1 \in \bar{R}_1, \cdots, \mathbf{x}_{n-1} \in \bar{R}_{n-1}, \mathbf{x}_n \in R_n\} & \text{if} \quad n \geq 2. \end{cases}$$

Let us assume that $\theta = (\theta_1, \theta_2)$, where $\theta_1 \in \Theta_1$ is the parameter of interest and $\theta_2 \in \Theta_2$ is the nuisance parameter. Further, we assume that Θ_1 is a real open interval and that $\Theta = \Theta_1 \times \Theta_2$. Now, an *estimating equation rule* g, for θ_1, is given by a sequence of estimating functions $g(\mathbf{X}_1, \theta_1)$, $g(\mathbf{X}_2, \theta_1)$, $\cdots$ such that for each $\theta_1 \in \Theta_1$, $g(\mathbf{X}_n, \theta_1)$ is B_n-measurable for all $n = 1, 2, \cdots$. If the (random) sample size $N = n$, we estimate θ_1 through the estimating equation $g(\mathbf{X}_n, \theta_1) = 0$.

Let us write $g_n = g(\mathbf{X}_n, \theta_1)$. Then we will say that the estimation rule g is *unbiased* if $E_\theta(g_N) = 0$ for all $\theta \in \Theta$. Of all such rules, g^* is said to be *optimum* if it minimizes $E_\theta\{g_N/E_\theta(\partial g_N/\partial\theta_1)\}^2$ for all $\theta \in \Theta$. The idea is to obtain equations with small variance $E_\theta(g_N^2)$, so that $g(\mathbf{X}_N, \theta_1) \cong 0$, and with high sensitivity, $|E_\theta(\partial g_N/\partial\theta_1)|$, both conditions roughly implying a small bias for the estimate. These definitions were given by Khan (1969) for the case where the nuisance parameter is absent, that is, in our notation, when $\theta_2 = \theta_{20}$, a specified value. In the same paper, extending results in Godambe (1960), the optimality of the maximum likelihood estimation rule, given by $g_n = \partial \log p_n/\partial\theta_1$, was shown and a generalization of Wolfowitz's (1947) inequality was obtained.

In this paper optimal estimation rules are given in the nuisance parameter case. In particular, the following situation is analyzed.

ASSUMPTION 1.1. For each n and θ_1 fixed, a statistic $T_n = T_n(\mathbf{X}_n)$ which does not depend on θ and is sufficient for θ_2 exists; that is,

$$p_n(\mathbf{x}_n, \theta) = f_{T_n}(\mathbf{x}_n, \theta_1) h_n(T_n, \theta), \tag{1}$$

Received February 1980; revised April 1981.

[1] This research is based on the author's Ph.D. dissertation submitted to the University of Waterloo, June 1980.

AMS 1980 *subject classifications.* Primary 62F10; secondary 62L12.

Key words and phrases. Estimating equation, stopping rule, nuisance parameter.

where $f_{T_n}(\mathbf{x}_n, \theta_1)$ denotes the conditional density for $\mathbf{X}_n$ given T_n, and $h_n(T_n, \theta)$ denotes the marginal density for T_n.

Under this assumption, which is a sequential version of one discussed by Godambe (1976), the optimality of the maximum conditional likelihood rule, $g_n^* = \partial \log f_{T_n}/\partial\theta_1$, is investigated. In addition, the problem of choosing stopping procedures is discussed. Based on Bhapkar's (1972) concept of information in an estimating equation, criteria for stopping are proposed and illustrated through several applications.

2. Optimal rules. We make the following assumptions on the frequency functions:

(a) $\partial \log p_n/\partial\theta_1$ and $\partial^2 \log p_n/\partial\theta_1^2$ exist on $\mathscr{X}^n$ for each $n = 1, 2, \cdots$ $(\theta \in \Theta)$;

(b) $\int_{\tilde{R}_n} p_n \, d\mathbf{x}_n$ is differentiable under the integral sign w.r.t. θ_1 for each $n = 1, 2, \cdots$ and $\theta \in \Theta$;

(c) $\sum_{n=1}^{\infty} \int_{\tilde{R}_n} (\partial \log p_n/\partial\theta_1) p_n \, d\mathbf{x}_n$ converges uniformly on Θ;

(d) when Assumption 1.1 holds, conditions similar to (a), (b) and (c) are also valid for the frequency function f_{T_n} and h_n.

On the estimation rules we introduce the following regularity conditions:

(i) (unbiasedness) $E_\theta(g_N) = 0$ $(\theta \in \Theta)$;

(ii) $g_n = g(\mathbf{x}_n, \theta_1)$ is differentiable w.r.t. θ_1 for each $n = 1, 2, \cdots$ $(\mathbf{x}_n \in \mathscr{X}^n, \theta_1 \in \Theta)$;

(iii) $\int_{\tilde{R}_n} g_n p_n \, d\mathbf{x}_n$ is differentiable under the integral sign w.r.t. θ_1 for each $n = 1, 2, \cdots$ $(\theta \in \Theta)$;

(iv) $\sum_{n=1}^{\infty} \int_{\tilde{R}_n} g_n (\partial \log p_n/\partial\theta_1) p_n \, d\mathbf{x}_n$ and $\sum_{n=1}^{\infty} \int_{\tilde{R}_n} (\partial g_n/\partial\theta_1) p_n \, d\mathbf{x}_n$ both converge uniformly on Θ;

(v) when Assumption 1.1 holds, conditions similar to (iii) and (iv) hold with p_n replaced by f_{T_n} or h_n and $\partial\, log\, p_n/\partial\theta_1$ replaced by $\partial \log f_{T_n}/\partial\theta_1$ or $\partial \log h_n/\partial\theta_1$, respectively;

(vi) $0 < \{E_\theta(\partial g_N/\partial\theta_1)\}^2$, $\theta \in \Theta$. The class of all regular unbiased estimation rules will be denoted by $\mathscr{G}_1$.

We are now in position to formalize the definition of optimality.

DEFINITION 2.1. An estimation rule $g^* \in \mathscr{G}_1$ is said to be optimum if

$$E_\theta\{g_N/E_\theta(\partial g_N/\partial\theta_1)\}^2 \geq E_\theta\{g_N^*/E_\theta(\partial g_N^*/\partial\theta_1)\}^2 \tag{2}$$

for all $g \in \mathscr{G}_1$ and all $\theta \in \Theta$.

2.1. *Complete sequential procedures.* We will restrict the attention to stopping rules satisfying $E_\theta(N) > \infty$ and such that the decision on whether or not to continue sampling after n observations depends on $T_n(\mathbf{X}_n)$. In particular we need the following

ASSUMPTION 2.1. $\tilde{R}_n$ depends on $\mathbf{x}_n$ through T_n alone.

It follows (Blackwell, 1947) that (T_N, N), where N is the random sample size, is a sufficient statistic for θ_2, for each fixed θ_1. Further, we assume that for each fixed θ_1, the sequential procedure satisfies the Lehmann and Stein's (1950) definition of *completeness*; i.e. we assume that for each fixed θ_1, the family of distributions of (T_N, N) is complete. Illustrations of this kind of sequential procedures will be presented in the case where the dimension of the nuisance parameter in $p_n(\mathbf{x}_n, \theta)$ increases when the sample size increases; see Section 3 and Appendix. Now, notice that, in general, Assumption 2.1 will not hold, even for stopping rules based on the sequence $\{T_n\}$. However, it will hold for such a rule if the following condition is valid:

ASSUMPTION 2.2. T_m is a function of T_n for $m < n$. For complete sequential procedures satisfying Assumption 2.1 the following sequential version of Theorem 3.2 in Godambe (1976), proving the optimality of the maximum conditional likelihood rule, holds.

THEOREM 2.1. *Consider a complete sequential procedure such that the decision on whether or not to continue sampling after n observations depends on* $T_n(\mathbf{X}_n)$. *Let*

Assumptions 1.1 *and* 2.1 *hold and, in addition, assume that* g^* *defined by*

$$g^*(\mathbf{X}_n, \theta_1) = \partial \log f_{T_n}/\partial\theta_1 \tag{3}$$

belongs to $\mathscr{G}_1$. *Then this rule is optimum in the sense of Definition* 2.1 *and further,*

$$E_\theta\{g_N/E_\theta(\partial g_N/\partial\theta_1)\}^2 \geq 1/E_\theta(\partial \log f_{T_N}/\partial\theta_1)^2 \tag{4}$$

for all $g \in \mathscr{G}_1$ *and* $\theta \in \Theta$, *the equality being attained when* g_n *is given by* (3) *up to a non-null multiplicative constant* $Z(\theta_1)$ *and up to sets of measure zero.*

The proof proceeds as in the original paper by Godambe (1976) and is omitted. It is important to notice that the conditional distribution of $\mathbf{X}_n$ given $N = n$ and $T_N = t$ has density

$$f^\dagger_{n,t}(\mathbf{x}_n, \theta_1) = I_{\tilde{R}_n}(\mathbf{x}_n) f_{T_n=t}(\mathbf{x}_n, \theta_1)/k(\theta_1),$$

where $k(\theta_1) = P(\mathbf{X}_n \in \tilde{R}_n \mid T_n = t, \theta_1)$, and reduces to $f_{T_n=t}(\mathbf{x}_n, \theta_1)$ if Assumption 2.1 holds for $\mathbf{x}_n \in \tilde{R}_n$. This in particular implies that $E_{\theta_1}(\partial \log f_{T_N}/\partial\theta_1 \mid N, T_N) = 0$; hence g^* defined in (3) is unbiased.

2.2. *Bounded sampling plans.* Let us restrict our attention to bounded sampling plans, i.e. those satisfying

$$P_\theta(N \leq n_0) = 1, \qquad \theta \in \Theta.$$

In addition, let us consider a set of statistics T_n, $n = 1, 2, \cdots, n_0$, satisfying Assumption 1.1 and, for simplicity, let us write $T = T_{n_0}$. We will now replace the condition of completeness of the sequential procedure by the following

ASSUMPTION 2.3. For each θ_1 fixed, the family of probability distributions induced by T,

$$\mathscr{P}^T_{\theta_1} = \{P^T_\theta : \theta_2 \in \Theta_2, \theta_1 \text{ fixed}\}$$

is complete.

ASSUMPTION 2.4. The conditional distributions of $\mathbf{X}_n$ given T and given T_n satisfy

$$f_T(\mathbf{x}_n, \theta_1) = f_{T_n}(\mathbf{x}_n, \theta_1), \qquad n \leq n_0. \tag{5}$$

Under these assumptions we now prove again the optimality of the maximum conditional likelihood rule. Notice that this theorem is a useful tool in the sense that the stopping procedure is only restricted by the condition of boundedness.

THEOREM 2.2. *Let Assumptions* 1.1, 2.3 *and* 2.4 *hold and assume that* g^* *in* (3) *belongs to* $\mathscr{G}_1$. *Then* g^* *is optimum according to Definition* 2.1 *and inequality* (4) *holds for all* $g \in \mathscr{G}_1$ *and* $\theta \in \Theta$, *the equality being attained when* g_n *is given by* (3) *up to a non-null multiplicative constant* $Z(\theta_1)$ *and up to sets of measure zero.*

PROOF. Let us define a class of $\mathscr{G}_{1T}$ of estimation rules, adding to the conditions in $\mathscr{G}_1$ the following requirement

(vi)$_T$: $$E_{\theta_1}(g_N \mid T) = 0 \quad \text{a.e.,} \quad \theta_1 \in \Theta_1.$$

Notice that the expectation in (vi)$_T$ depends only on θ_1 because of Assumption 1.1. Now, for every $g \in \mathscr{G}_{1T}$, condition (vi)$_T$ may be written as

$$E_{\theta_1}\{\textstyle\sum_{n=1}^{n_0} I_{\{\tilde{R}_n\}}(\mathbf{X}_n) g(\mathbf{X}_n, \theta_1) \mid T\} = 0. \tag{6}$$

Differentiating (6) with respect to θ_1 (condition (v)) and taking into account (5), we obtain

$$E_{\theta_1}\{\textstyle\sum_{n=1}^{n_0} I_{\{\tilde{R}_n\}}(\mathbf{X}_n)(\partial g_n/\partial\theta_1 + g_n \partial \log f_{T_n}/\partial\theta_1) \mid T\} = 0.$$

Taking expectation on T, we have that

$$E_\theta(\partial g_N/\partial\theta_1) + E_\theta(g_N g_N^*) = 0 \tag{7}$$

and then, applying the Cauchy-Schwartz inequality and taking into account (vi), (4) is obtained. The equality is attained only if $g_N = Z(\theta_1) g_N^*$ a.e., where $Z(\theta_1) \neq 0$ for all $\theta_1 \in \Theta_1$. Further, if $g^* \in \mathscr{G}_{1T}$, we can replace g_N by g_N^* in (7), obtaining

$$E_\theta(\partial g_N^*/\partial\theta_1) + E_\theta(g_N^{*2}) = 0. \tag{8}$$

The optimality of g^* follows from (8) and (4).

Now, for every $g \in \mathscr{G}_1$, condition (i) may be written as $E_\theta\{E_{\theta_1}(g_N \mid T)\} = 0$, $\theta \in \Theta$. Then, the completeness of T implies (vi)$_T$. Hence $\mathscr{G}_1 \subset \mathscr{G}_{1T}$, and since obviously $\mathscr{G}_{1T} \supset \mathscr{G}_1$, the proof is complete.

Notice that from (5) and (6) we can write generally,

$$E_{\theta_1}(g_N \mid T) = \textstyle\sum_{n=1}^{n_0} \displaystyle\int I_{\{\tilde{R}_n\}}(\mathbf{x}_n) g(x_n, \theta_1) f_{T_n}(\mathbf{x}_n, \theta_1)\, d\mathbf{x}_n. \tag{9}$$

Then it follows that g^* will be unbiased since $(\partial/\partial\theta_1)E_{\theta_1}(1 \mid T)$ coincides with $E_{\theta_1}(g_N^* \mid T)$ according to (9).

We now present an example that shows the non-completeness in general of a sequential procedure satisfying the assumptions in Theorem 2.2.

EXAMPLE 2.1. Let $X_i = (X_{i1}, X_{i2})$, $i = 1, 2$, the random variables X_{ij}, $i, j = 1, 2$, being independent and such that $X_{ij} \sim N(\theta_{2i}, \theta_1)$. Let $R_1 = \{x_1 : x_{11} < x_{12}\}$ and $R_2 = \{\mathbf{x}_2 : x_{11} \geq x_{12}\}$, so that $n_0 = 2$. Consider the statistic T_n defined by $T_1 = X_{11} + X_{12}$ and $T_2 = (X_{11} + X_{12}, X_{21} + X_{22})$. Let us define a function of (T_N, N) as follows: $u(T_1, 1) = 1$ and $u(T_2, 2) = -1$. Then $E_\theta\{u(T_N, N)\} = 0$ for all $\theta \in \Theta$ and therefore (T_N, N) is not complete.

2.3. *Further results.* We will now extend to the sequential setting three additional results in Godambe (1976). The proofs and the regularity conditions required are obvious sequential versions of those given by Godambe and are omitted.

THEOREM 2.3. *If the estimating equation rules $g_1^* \in \mathscr{G}_1$ and $g_2^* \in \mathscr{G}_1$ are optimum according to Definition* 2.1, *then up to sets of measure zero,*

$$g_{1N}^*/E_\theta(\partial g_{1N}^*/\partial\theta_1) = g_{2N}^*/E_\theta(\partial g_{2N}^*/\partial\theta_1), \qquad \theta \in \Theta.$$

REMARK 2.1. The conclusion of Theorem 2.3 may be equivalently expressed as follows: for some function $S(\theta_1)$,

$$g_{1n}^* = S(\theta_1) g_{2n}^* \quad \text{for all} \quad \mathbf{x}_n \in \tilde{R}_n, \qquad n = 1, 2, \cdots, \quad \theta \in \Theta.$$

THEOREM 2.4. *If a rule g^* such that*

$$g_n^* = A(\theta)\partial \log p_n(\mathbf{X}_n, \theta)/\partial\theta_1 + r_n(\mathbf{X}_n, \theta)$$

satisfies (a) $g^* \in \mathscr{G}_1$ *and* (b) $E_\theta(r_N g_N) = 0$ *for all* $\theta \in \Theta$ *and* $g \in \mathscr{G}_1$, *then it is optimum according to Definition* 2.1.

REMARK 2.2. Notice that the choice

$$r_n = C(\theta)\{(\partial \log p_n/\partial\theta_2)^2 + (\partial^2 \log p_n/\partial\theta_2^2)\} \qquad \text{for} \quad \mathbf{x}_n \in \tilde{R}_n$$

ensures that (b) holds and also that g_N^* is unbiased. This result is a generalization of a theorem in Godambe and Thompson (1974).

ASSUMPTION 2.5. There exists a real function $\alpha(\theta_1, \theta_2)$ such that h_n in (1) depends on θ only through α and such that $\partial \log h_n/\partial\alpha$, $\partial\alpha/\partial\theta_1$ and $\partial\alpha/\partial\theta_2$ exist and $\partial\alpha/\partial\theta_2 \neq 0$, $\theta \in \Theta$.

THEOREM 2.5. *Let Assumptions* 1.1 *and* 2.5 *hold. Then, if* g^* *defined by* $g_n^* = \partial \log f_{T_n}(\mathbf{X}_n, \theta_1)/\partial\theta_1$ *belongs to* $\mathscr{G}_1$, *it gives the optimum estimating rule according to Definition* 2.1, *unique up to a constant multiple* $Z(\theta_1)$ *and up to sets of measure zero.*

Generalizations of this theorem are easily obtained. For example, suppose that, on taking new parameters (θ_1, α), the density factorizes as $p_n = a(\mathbf{x}_n, \theta_1)b(\mathbf{x}_n, \alpha)$, where, locally, θ_1 and α are not functionally related. Then $(\partial/\partial\theta_1)\log a(\mathbf{X}_n, \theta_1)$ gives the optimum rule. Another generalization to the case where h_n depends on θ only through real functions $\alpha_i(\theta_1, \theta_{2i})$, $i = 1, \cdots, n$, was given by Ferreira (1980).

3. Information, stopping rules and examples. Up to now we have considered the problem of finding an optimum estimation rule for a given stopping procedure. We will now briefly consider the problem of choosing an optimum stopping rule. Firstly, following Bhapkar (1972), we define the amount of *information* $J_g(\theta \mid R)$ in the estimation rule g, given the stopping procedure R_n, $n = 1, 2, \cdots$, as the inverse of the left hand side of (2). The quotient $J_g(\theta \mid R)/J_{g^*}(\theta \mid R)$ will be called the *efficiency* of the estimation rule g.

Based on the above concept of information, we now introduce criteria of optimality for stopping procedures:

(i) Subject to the condition $J_{g^*}(\theta \mid R) \geq J_0 > 0$, $\theta \in \Theta$, minimize the expected sample size $E_\theta(N)$;

(ii) subject to the condition $E_\theta(N) \leq m$, maximize the information in the optimum estimation rule;

(iii) minimize $F(R, \theta) = AJ_{g^*}^{-1}(\theta \mid R) + E_\theta(N)$ where $A > 0$ is fixed.

In the particular case where $\theta_2 = \theta_{20}$, a specific value, i.e. θ is a single parameter, Khan (1969, Theorem 5.1) proved that $g_n^* = \partial \log p_n/\partial\theta_1$ is optimum. Further, if $(X_1, X_2, \cdots)$ are independently and identically distributed, we have that $J_{g^*}(\theta \mid R) = E_\theta(N)E_\theta(\partial \log p/\partial\theta_1)^2$. Then applying criterion (i), we must minimize $E_\theta(N)$ subject to

$$E_\theta(N) \geq J_0/E_\theta(\partial \log p/\partial\theta_1)^2. \tag{10}$$

If the right hand side of (10) does not depend on θ, then a fixed sample size procedure is optimum. However, if it does depend on θ, we will not be able to determine the size of the sample. In this situation, an approximate criterion for stopping may be suggested through the *fixed sample size* information $J_{g^*}^{(n)}(\theta) = E_\theta(-\partial^2 \log p_n/\partial\theta_1^2)$. For example, we may stop when an estimate of $J_{g^*}^{(n)}(\theta)$ is greater than a given constant J_0. Similar considerations are valid in the nuisance parameter case, but no general formula like (10) is available.

EXAMPLE 3.1. Let $(X_1, X_2, \cdots)$ be i.i.d. $N(\theta_1, \theta_2)$. If $\theta_2 = \theta_{20}$, known, then (Khan, 1969, Theorem 5.1) the optimum estimation rule is

$$g_n^* = \partial \log p_n/\partial\theta_1 = n\theta_2^{-1}(\bar{X}_n - \theta_1).$$

Further, $J_{g^*}(\theta \mid R) = E_\theta(N)\theta_2^{-1}$ and hence, applying criterion (i), a fixed sample size procedure with $n = J_0\theta_2$ is obtained. This procedure is also obtained when sampling till the variance of $\hat{\theta}_1 = \bar{X}$ reaches a given level J_0^{-1}. On the other hand, if θ_2 is unknown we may apply Theorem 2.4 with $A(\theta) = \theta_2$ and $r_n = 0$, obtaining the optimum rule $g_n^* = n(\bar{X} - \theta_1)$. Further, $J_{g^*}^{(n)}(\theta) = n\theta_2^{-1}$ may be estimated by ns^{-2}, where $s^2 = \sum_1^n (X_i - \bar{X})^2/(n - 1)$, and hence we obtain the well known rule $n \geq J_0 s^2$ (Kendall and Stuart, 1979, page 650).

EXAMPLE 3.2. Let $X_i = (X_{1i}, X_{2i})$, $i = 1, 2, \cdots$, be independently distributed, such that X_{1i} given $X_{2i} = x_{2i}$ is $N(\theta_1 x_{2i}, 1)$ and X_{2i} is $N(\theta_{2i}, 1)$. That is, we have a regression through the origin where the regression variable is subject to error, and where inference about the slope is of prime interest. Let us take $T_n = (X_{21}, \cdots, X_{2n})$. Then Assumption 1.1 holds and, further, if we restrict the attention to complete sequential procedures, then g^* in (3) is optimum and

$$J_{g^*}^{(n)}(\theta) = -E_\theta(\partial^2 \log f_{T_n}/\partial\theta_1^2) = E_\theta(\textstyle\sum_1^n X_{2i}^2).$$

Hence, we may stop sampling when $\sum_1^n X_{2i}^2 > J_0$. In the Appendix we show that this sequential procedure is complete and therefore, applying Theorem 2.1, we obtain the optimum rule $g_n^* = \sum_1^n (X_{1i} - \theta_1 X_{2i})X_{2i}$. Notice that g_n^* gives the usual slope estimate and that the stopping rule is determined by the size of its (conditional) variance.

The use of criterion (ii) being $J_g^*(\theta \mid R) = c(\theta)E_\theta(N)$, where $c(\theta)$ does not depend on R, entails the maximization of $E_\theta(N)$ and hence a fixed sample size procedure constitutes an optimum rule.

EXAMPLE 3.3. Let $X_1 = (X_{11}, X_{21})$, $X_2 = (X_{12}, X_{22})$, $\cdots$ be independent random variables such that X_{1i} and X_{2i} have independent Poisson distributions with means θ_1 and $\theta_1 + \theta_2$, respectively, being $\theta_1 > 0$ and $\theta_2 > 0$. Taking $T_n = (X_{21}, X_{22}, \cdots, X_{2n})$ and $\alpha = \theta_1 + \theta_2$, Theorem 2.5 applies and

$$g^*(\mathbf{X}_n, \theta_1) = -n + \textstyle\sum_{j=1}^n X_{1j}/\theta_1. \tag{11}$$

Further, $J_{g^*}(\theta \mid R) = E_\theta(N)/\theta_1$ for all possible R and hence a fixed sample size procedure is optimum according to (ii). On the other hand, if we estimate θ_1 through (11) and proceed to sampling till $J_{g^*}^{(n)}(\hat{\theta}) = n/\hat{\theta}_1 \geq J_0$, the rule $R_n = \{\sum_{j=1}^n X_{1j} < cn^2\}$ is reached. This is the same rule obtained when estimating the mean of a Poisson distribution with a given precision (Kendall and Stuart, 1979, page 651).

Notice that, according to criterion (iii), when $J_{g^*}(\theta \mid R)$ is of the form $cE_\theta(N)$, c being a positive constant, then a fixed sample size procedure will be optimum. Otherwise we may proceed as in Robbins (1959), first finding the value of n minimizing

$$F_{(\theta)}^{(n)} = AJ_{g^*}^{(n)-1}(\theta) + n, \tag{12}$$

say $n_0 = \phi(\theta)$, and then continuing sampling till $n > \phi(\hat{\theta}_n)$.

EXAMPLE 3.4. Let $(X_1, X_2, \cdots)$ be independent random variables such that $X_i = (X_{1i}, X_{2i})$ has two independent components each $N(\theta_{2i}, \theta_1)$. Take $T_n = (X_{11} + X_{21}, \cdots, X_{1n} + X_{2n})$. Then Assumption 2.1 follows from Theorem 7.3 of Lehmann and Scheffé (1955). Hence, if the sequential procedure is bounded, Theorem 2.2 applies and

$$g_n^* = -\frac{n}{2\theta_1} + \frac{1}{4\theta_1^2}\textstyle\sum_1^n (X_{1i} - X_{2i})^2. \tag{13}$$

Minimizing $F_{(\theta)}^{(n)} = A2\theta_1^2 n^{-1} + n$ (see (12)) as a function of n, we obtain $n = \sqrt{2A}\theta_1$. Hence, if θ_1 is estimated through (13) and the condition $\{N \leq n_0\}$ is imposed, we arrive at the following procedure: stop when either $n^2 \geq \sqrt{A/2}\ \sum_1^n (X_{1i} - X_{2i})^2$ or $n \geq n_0$.

Acknowledgments. The author wishes to acknowledge with gratitude the encouragement and general guidance received from Prof. V. P. Godambe. He is also indebted to Prof. D. R. Cox, M. Thompson and S. Zacks and two referees for valuable comments.

APPENDIX

Let $(X_1, X_2, \cdots)$ be independent random variables, X_i having a distribution in a family parameterized by (θ_1, θ_{2i}). Further, let the sequential procedure be such that Assumptions 2.1 and 2.2 hold and, in addition, the following conditions are satisfied:

(Ai) $\mathscr{P}_{\theta 1}^{T_1} = \{h(T_1, \theta_1, \theta_{21}) : \theta_1 \text{ fixed}\}$ is complete;

(Aii) Given $m < n$, $h_{T_m = t}(T_n, \theta)$ does not depend on θ_{2m} and constitutes a complete family when the parameters θ_1 and θ_{2i}, $i < m$, are held fixed, i.e. the family of conditional distributions of T_n given T_m is strongly complete (see Lehmann and Stein, 1950).

We will show that this sequential procedure is complete, i.e. the family of distributions of (T_n, N), θ_1 fixed, is complete. In fact, using Assumption 2.1 we write

$$
\begin{aligned}
E_\theta\{u(T_N, N)\} &= \textstyle\sum_{n=1}^{\infty} \displaystyle\int_{\tilde{R}_n} u(T_n, n) h(T_n, \theta)\, dT_n \\
&= \int v_1(T_1, \theta_1, \theta_{22}, \theta_{23}, \cdots) h(T_1, \theta_1, \theta_{21})\, dT_1 = 0,
\end{aligned} \tag{14}
$$

where

$$
v_1(t, \theta_1, \theta_{22}, \theta_{23}, \cdots) = I_{\{\tilde{R}_1\}}(t) u(t, 1) + \textstyle\sum_{n=2}^{\infty} \displaystyle\int I_{\{\tilde{R}_n\}}(T_n) u(T_n, n) h_{T_1=t}(T_n, \theta)\, dT_n. \tag{15}
$$

Assumption (Aii) implies that $h_{T_1=t}(T_n, \theta)$ in (15) does not depend on θ_{21} and the same holds for v_1. Hence, from (Ai) and (14) we conclude that $v_1 = 0$ for all possible t. Further, if $t \in \tilde{R}_1$, then the summation in the r.h.s. of (15) vanishes and hence $u(t, 1) = 0$. The proof follows, repeating the same reasoning, arguing by induction and using Assumption 2.2 to show that the conditional distribution of T_n given $T_1, \cdots, T_m$ coincides with that of T_n given T_m for $m < n$.

REFERENCES

BHAPKAR, V. P. (1972). On a measure of efficiency of an estimating equation. *Sankhyā Ser. A.* **34** 467–472.

BLACKWELL, D. (1947). Conditional expectation and unbiased sequential estimation. *Ann. Math Statist.* **18** 105–110.

FERREIRA, P. E. (1980). Contributions to the Theory of Estimating Equations. Ph.D. Thesis presented at the University of Waterloo.

GODAMBE, V. P. (1960). An optimum property of regular maximum likelihood equation. *Ann. Math. Statist.* **31** 1208–1211.

GODAMBE, V. P. (1976). Conditional likelihood and unconditional optimum estimating equations. *Biometrika* **63** 277–284.

GODAMBE, V. P. and THOMPSON, M. E. (1974). Estimating equations in the presence of a nuisance parameter. *Ann. Statist.* **2** 568–571.

KENDALL, M. G. and STUART, A. (1979). *The Advanced Theory of Statistics,* 2 4th ed. Griffin, London.

KHAN, R. (1969). Maximum likelihood estimation in sequential experiments. *Sankhyā Ser. A* **31** 49–56.

LEHMANN, E. L. and SCHEFFÉ, H. (1955). Completeness, similar regions and unbiased estimation. II. *Sankhyā* **15** 219–236.

LEHMANN, E. L. and STEIN, C. (1950). Completeness in the sequential case. *Ann. Math. Statist.* **21** 376–385.

ROBBINS, H. (1959). Sequential estimation of the mean of a normal population. *Probability and Statistics.* U. Grenander, ed. Wiley, New York. 235–245.

WOLFOWITZ, J. (1947). The efficiency of sequential estimates and Wald's equation for sequential processes. *Ann. Math. Statist.* **18** 215–230.

DEPARTMENTO DE ESTATISTICA
UNIVERSIDADE DE BRASILIA
ASA NORTE,
BRASILIA, D.F.,
BRASIL, CEP 70910