The Annals of Mathematical Statistics
1969, Vol. 40, No. 2, 568–574

THE CONSISTENCY OF CERTAIN SEQUENTIAL ESTIMATORS[1]

BY R. M. LOYNES[2]

Florida State University

1. Introduction and summary. The results described here have their roots in two areas, for in a certain sense we combine on the one hand the work of Girshick, Mosteller and Savage [5] and Wolfowitz [11] and [12] on sequential estimation of the binomial parameter, and on the other the result of Hoeffding [7] concerning the consistency of U-statistics. The link between the two is the Blackwell [2] procedure for obtaining another (better) estimator from a given one by taking expectations conditional on a sufficient statistic.

The main result is that if from a given estimator T of $\theta = ET$ we construct new estimators by the Blackwell procedure corresponding to a sequence of stopping-rules N_i, then this sequence of estimators is consistent provided N_i tends to infinity in probability; in fact it has also to be assumed that the N_i have a certain structural property.

2. Notation, terminology and universal assumptions. We suppose $X_1, X_2, \cdots$ to be a sequence of independent identically distributed random variables, taking their values in a space $\mathfrak{X}$.

A stopping-rule or *stopping-time* N is a random variable defined on the sequence $X_1, X_2, \cdots$ whose possible values are the positive integers, with the property that for each $n \geqq 1$ the event $\{N = n\}$ is determined by conditions on $X_1, X_2, \cdots, X_n$ only. We assume that all stopping-times are finite with probability one.

For brevity we occasionally denote the ordered n-tuple $(X_1, X_2, \cdots, X_n)$ by X^n. By Z_n we mean the *order-statistic* calculated from X^n. If the X_i are real-valued this is just the usual order-statistic; otherwise we can regard it as the function from $\mathfrak{X}$ to the integers which describes how many of $X_1, X_2, \cdots X_n$ are equal to a given x in $\mathfrak{X}$. (The description in these terms I owe to L. J. Savage). It is important to note that Z_{n+1} can be calculated from a knowledge of Z_n and X_{n+1}.

There are various assumptions to be made which we shall label A1, A2, etc.

A1: For each n a statistic V_n is given which is sufficient for X^n. It is not supposed that V_n is real-valued.

A2: For each n V_n is a function of Z_n.

Assumption A2 will certainly be true if $V_n = Z_n$ or if V_n is a minimal sufficient statistic.

When a stopping-rule N is given we shall write V_N for the random variable

Received 20 November 1967.

[1] This research was partially supported by the Army, Navy and Air Force under Office of Naval Research Contract No. NONR 988(08). Reproduction in whole or in part is permitted for any purpose of the United States Government.

[2] Presently at the Statistical Laboratory, Cambridge.

which equals V_n when $N = n$; when no confusion can be caused we shall merely write $V_N = V$. When, as in Section 4, we deal with a sequence of stopping-rules N_i, we shall write V^i rather than V_{N_i}.

We suppose that for some parameter θ we have constructed an unbiased estimator T (defined on the sequence $X_1, X_2, \cdots$) which will remain fixed throughout the discussion.

A3: $E_\theta T = \theta$; in particular $E_\theta T$ exists.

The estimator T will be assumed to depend on a finite (but not necessarily bounded) number of observations only: there is a stopping-rule M for T such that if $M = n$, T is a function of $X_1, X_2, \cdots, X_n$ only. Any such estimator we call a sequential estimator; a fixed sample-size estimator, depending on a fixed number of X_i, will be a special case of this obtained by putting $M = m$ with probability one for an appropriate integer m.

A4: Any stopping-rule N or N_i considered satisfies $N \geqq M$ or $N_i \geqq M$ with probability one.

The function of A4 is, in conjunction with A1, to ensure that the quantities $U(N)$ and $U(N_i)$ defined in (3.1) and (4.1) are functions of the observations only.

If A is an event, we shall often use the notation $I(A)$ for the indicator function of A, which is unity on A and zero elsewhere.

3. Construction of the Blackwell estimators. Given the estimator T and a stopping-rule N we define the new estimator $U = U(N)$ as the conditional expectation of T given N and V.

$$U(N) = E[T \mid N, V_N]. \tag{3.1}$$

Equivalently $U = U_n(v)$ when $N = n$ and $V_n = v$ where

$$U_n(v) = E[T \mid N = n, V_n = v] = E[TI(N = n) \mid V_n = v] \cdot \{E[I(N = n) \mid V_n = v]\}^{-1}. \tag{3.2}$$

Assumptions A1 and A4 clearly imply that $U_n(v)$ is a function of $X_1, X_2, \cdots, X_n$ only, so that U is indeed a statistic with stopping-rule N. Then U is also unbiased for θ, and has at least as small a loss as T for any convex loss function (indeed unbiasedness is not necessary for this property.)

As a particular case we may take $V_n = Z_n$, which would lead in the case of fixed M and N to U-statistics (see e.g. Fraser [4]), and we may therefore regard such estimators as generalised U-statistics. We observe that in any case, because of A2, U is a symmetric function of $X_1, X_2, \cdots, X_N$.

The obvious question, though not particularly relevant to the present investigation, is when is U the unique minimum variance unbiased estimator for the given N. It will of course be unique when the pair (N, V) is complete, but this is no more than a restatement of the question in different terms. Lehmann and Stein [9] have some results, but the problem seems to be appreciably more difficult to treat than in the fixed sample-size case.

Blackwell [2], in a paper apparently directly inspired by the sequential binomial estimation procedures developed by Girshick, Mosteller and Savage [5], gave an almost identical discussion, except that he imposed an unnecessary further condition on N.

4. The consistency of certain sequences of estimators. Wolfowitz [12] showed that if a sequence of estimators were constructed in the binomial case by the method of Section 3 corresponding to a sequence of stopping-rules N_i satisfying a certain condition, it would be a consistent sequence. His condition was that $n_{0,i} \to \infty$ as $i \to \infty$, where $n_{0,i}$ is the smallest value of n for which $P[N_i = n] > 0$. It will be shown here that even in the general case consistency follows from weaker and more appealing conditions on the N_i.

We shall write

$$U^i \equiv U(N_i) = E[T \mid N_i, V^i]. \tag{4.1}$$

Theorem 1. *Suppose that assumptions* A1 *to* A4, *and in addition the following conditions* C1 *and* C2 *are satisfied.* C1: *For any fixed k, $P[N_i \leqq k] \to 0$ as $i \to \infty$.* C2: *There exists an integer $\lambda(i, k)$, which is monotone in k and which tends to ∞ if i and k both tend to ∞, such that for each k, $N_i = k$ if and only if $N_i \geqq k$ and the set of random variables $Z_{\lambda(i,k)}, X_{\lambda(i,k)+1}, \cdots, X_k$ satisfies some condition (depending of course on i and k).*

Then U^i converges to θ in probability and in the mean of order 1 *as $i \to \infty$.*

Remarks. (i) C1 states that $N_i \to \infty$ in probability. Wolfowitz' condition implies that $N_1 \to \infty$ with probability 1.

(ii) C2 is not necessary, since for example it obviously need only be required to hold for large i, but some condition obviously is. Its function is (effectively) to ensure that T and N_i are nearly independent when i is large. (We do not need to make this precise, but the intention is to exclude stopping-rules such as $N_i = [iX_1]$.) The form chosen here is a convenient one which is reasonable for most (though not all) applications. A somewhat simpler condition which implies C2 is

C2$'$: *For each k, $N_i = k$ if and only if $N_i \geqq k$ and Z_k satisfies some condition.*

This is very similar to the extra condition imposed by Blackwell [2]. As examples the stronger condition C2$'$ is satisfied by the stopping-rules (a) stop when $\sum^n X_j$ first crosses a barrier, (b) stop when the estimated variance of the sample mean s^2/n first becomes smaller than some prescribed value, and (in the binomial case), (c) stop when the point whose co-ordinates are number of successes and number of failures first enters a region R. An example in which C2 is satisfied but not C2$'$, is given by defining $N_i = n$ if X_n is the first X_j for which $X_j > \max(X_1, X_2, \cdots, X_i)$.

It would be straightforward to prove the theorem if Wolfowitz' condition were satisfied, and the main difficulty is in fact in taking advantage of the fact that it is almost satisfied. By C1 there exists an increasing sequence of integers m_i, tending to ∞ with i, with the property that

$$P[N_i < m_i] \to 0. \tag{4.2}$$

For convenience we shall write

$$E_i \equiv \{N_i \geqq m_i\}, \tag{4.3}$$

and

$$k_i = \lambda(i, m_i) \quad \text{so that} \quad k_i \to \infty \tag{4.4}$$

and

$$P(E_i) \to 1. \tag{4.5}$$

Define

$$W_i = E[E[T \mid I(E_i), Z_{k_i}, X_{k_i+1}, X_{k_i+2}, \cdots] \mid N_i, V^i], \tag{4.6}$$

then when $N_i = j$ and $V_j = v$

$$W_i = E[E[T \mid I(E_i), Z_{k_i}, X_{k_i+1}, \cdots] I(N_i = j) \mid V_j = v] \cdot \{E[I(N_i = j) \mid V_j = v]\}^{-1}. \tag{4.7}$$

Now if $j \geqq m_i$ it follows from C2 that $N_i = j$ if and only if $N_i \geqq m_i$ and the set of random variables $Z_{k_i}, X_{k_i+1}, \cdots, X_j$ satisfy some condition; consequently if $j \geqq m_i$

$$\begin{aligned} W_i &= E[E[TI(N_i = j) \mid I(E_i), Z_{k_i}, X_{k_i+1}, \cdots] \mid V_j = v] \\ &\quad \cdot \{E[I(N_i = j) \mid V_j = v]\}^{-1} \\ &= E[TI(N_i = j) \mid V_j = v]\{E[I(N_i = j) \mid V_j = v]\}^{-1} \end{aligned} \tag{4.8}$$

because of A2. Hence $W_i = U^i$ on $\{N_i \geqq m_i\}$ and by (4.5)

$$W_i - U^i \to 0 \quad \text{in probability}. \tag{4.9}$$

Now write

$$Y_i = E[T \mid I(E_i), Z_{k_i}, X_{k_i+1}, \cdots] \tag{4.10}$$

so that

$$W_i = E[Y_i \mid N_i, V^i]. \tag{4.11}$$

Then when $N_i \geqq m_i$, or $I(E_i) = 1$,

$$Y_i = E[TI(E_i) \mid Z_{k_i}, X_{k\ +1}, \cdots]\{E[I(E_i) \mid Z_{k_i}, X_{k_i+1}, \cdots]\}^{-1}. \tag{4.12}$$

We have

$$\begin{aligned} &E[|1 - E[I(E_i) \mid Z_{k_i}, X_{k_i+1}, \cdots]|] \\ &\qquad = E[|E[1 - I(E_i) \mid Z_{k_i}, X_{k_i+1}, \cdots]|] \leqq E[1 - I(E_i)] \to 0 \end{aligned} \tag{4.13}$$

so that the denominator of (4.12) tends in the mean of order 1 and *a fortiori* in probability, to 1. If now we write

$$S_i = E[T \mid Z_{k_i}, X_{k_i+1}, \cdots] \tag{4.14}$$

it follows similarly that the difference between the numerator of (4.12) and S_i tends in probability to 0, and hence finally, recalling C1,

$$Y_i - S_i \to 0 \quad \text{in probability.} \tag{4.15}$$

Now the S_i form a backwards martingale, so that by Theorem 4.2 of Chapter VII of Doob [3] S_i converges with probability 1 and in the mean of order 1 to a limiting random variable with expectation $ET = \theta$. From the zero-one law of Hewitt and Savage [6] it follows that the limit is in fact constant, and therefore equal to θ with probability 1.

$$S_i \to \theta. \tag{4.16}$$

Hence by (4.15) and (4.16)

$$Y_i \to \theta \quad \text{in probability.} \tag{4.17}$$

Now, if $\epsilon > 0$, writing $Y_i' = Y_i - \theta$ we have

$$\begin{aligned} E[|Y_i - \theta|] &= E[|Y_i - \theta| I(|Y_i'| > \epsilon)] + E[|Y_i - \theta| I(|Y_i'| \leqq \epsilon)] \\ &\leqq E[E[|T - \theta| \mid I(E_i) Z_{ki}, X_{k_i+1}, \cdots] I(|Y_i'| > \epsilon)] + \epsilon \\ &= E[E[|T - \theta| I(|Y_i'| > \epsilon) \mid I(E_i), Z_{k_i}, X_{k_i+1}, \cdots]] + \epsilon \\ &= E[|T - \theta| I(|Y_i'| > \epsilon)] + \epsilon \\ &\to 0 \end{aligned} \tag{4.18}$$

as $i \to \infty$ and then $\epsilon \to 0$, by (4.17). Consequently from (4.11)

$$\begin{aligned} E[|W_i - \theta|) &= E[|E[Y_i - \theta \mid N_i, V^i]|] \leqq E[E[|Y_i - \theta| \mid N_i, V^i]] \\ &= E[|Y_i - \theta|] \to 0, \end{aligned} \tag{4.19}$$

and *a fortiori* W_i converges to θ in probability, so that by (4.9)

$$U^i \to \theta \quad \text{in probability.} \tag{4.20}$$

That U^i also converges to θ in the mean of order 1 follows, by applying an argument exactly parallel to that in (4.18).

Intuition suggests that almost sure convergence of U^i to θ ought to hold, provided the requirements on N_i are strengthened: possible conditions which suggest themselves are $N_i \to \infty$ with probability one, or $N_{i+1} \geqq N_i$ for all i. Attempts to prove this have, however, been unsuccessful, except in the following non-sequential but otherwise rather general case. (It is sometimes possible to deal with the sequential case by special arguments—see Examples 3 and 4 in Section 5.) Suppose T is a function of X^m for some fixed m and that for each n V_n is a function of V_{n-1} and X_n (which again will be true if $V_n = Z_n$ or if V_n is minimal); we shall without loss of generality suppose $N_i = i$ for each i. Then $E[T \mid V_n] \to \theta$ with probability one. For if $n \geqq m$

$$E[T \mid V_n] = E[T \mid V_n, X_{n+1}, X_{n+2}, \cdots], \tag{4.21}$$

and these variables therefore form a backwards martingale which converges with probability one. If we set $V_n = Z_n$ we have a proof of the almost sure consistency of U-statistics (see Hoeffding [7]), modelled closely on Doob's proof of the strong law of large numbers ([3], p. 341]. The same proof was given by Berk [1].

5. Examples. We give five examples. The first two are entirely concerned with fixed sample sizes. All except the fifth have T depending only on a fixed sample size. All except the fourth satisfy C2′.

(1) Suppose that X_j are normally and independently distributed with mean μ and variance 1, that $T = X_1^2 - 1$ so that $\theta = \mu^2$, that $N_i = i$, and that we use the minimal sufficient statistic $V_n = \sum^n X_j$. Then we find

$$U^i = (\textstyle\sum^i X_j)^2 i^{-2} - i^{-1}, \tag{5.1}$$

which converges to θ in probability and in the mean of order 1 according to the theorem, and with probability 1 according to the remarks at the end of Section 4 (and in any case of course according to the strong law.) As a matter of interest the variance of U^i is

$$2i^{-2} + 4\mu^2 i^{-1}, \tag{5.2}$$

whereas that of the U-statistic associated with T is

$$2i^{-1} + 4\mu^2 i^{-1}. \tag{5.3}$$

(2) Suppose that the X_j are independently distributed, uniformly on $(0, 2\theta)$, that $T = X_1$, that $N_i = i$, and that $V_n = M_n$, where

$$M_n = \max (X_1, X_2, \cdots, X_n) \tag{5.4}$$

the minimal sufficient statistic. Then

$$U^i = (i + 1)(2i)^{-1} M_i, \tag{5.5}$$

whose behaviour is in general terms the same as that in example (1).

(3) Let the (real) X_j be independent and identically distributed with continuous distribution extending to $+\infty$, let $V_n = Z_n$, and let N_i be the first n for which $M_n > i$; again suppose $T = X_1$. Then if $N_i = 1$, $U^i = X_1$; otherwise, if $N_i = n \geqq 2$ and $Z_n = (z_1, z_2, \cdots, z_n)$ where $X_n = z_n > i$ and $z_1 \leqq z_2 \leqq \cdots \leqq z_{n-1}$, we have $X_1 = z_j$ with probability $(n - 1)^{-1}$ for each j $(1 \leqq j \leqq n - 1)$. Thus

$$\begin{aligned} U^i &= z_1 = X_1 && \text{if} \quad N_i = 1 \\ &= \textstyle\sum^{n-1} z_j (n-1)^{-1} = \sum^{n-1} X_j (n-1)^{-1} && \text{if} \quad N_i = n \geqq 2. \end{aligned} \tag{5.6}$$

By the theorem U^i converges to $\theta = EX$ in probability and in the mean of order 1. In both this example and the next we can also show that convergence with probability 1 occurs, for Theorem 1 of Richter [10] applies.

(4) Suppose T, X_j, and V_n are as in example (3), and let N_i be the first $n > i$

for which $M_n > M_i$. Then when $N_i = n$ and $Z_n = (z_1, z_2, \cdots, z_n)$ we have $X_n = z_n$, and z_{n-1} occurs among $X_1, X_2, \cdots, X_i$; thus $T = z_{n-1}$ with probability i^{-1}, and $T = z_j$ with probability $(i-1)/(n-2)i$ for $1 \leqq j \leqq n-2$. It follows that

$$U^i = z_{n-1}i^{-1} + (i-1)i^{-1}(z_1 + z_2 + \cdots + z_{n-2})(n-2)^{-1}. \tag{5.7}$$

The convergence behavior is as in example (3).

(5) Suppose the situation is as in example (3), except that $T = X_{N_1}$, and consider as before the case when $N_i = n$ and $Z_n = (z_1, z_2, \cdots, z_n)$. Then clearly, if $n = 1, N_1 = 1$; if $n > 1$ and $z_n \leqq 1, N_1 = n$; and if $n > 1$ and $z_1 < z_2 < \cdots z_r \leqq 1 < z_{r+1} < \cdots < z_n$, $X_{N_1} = z_j$ with probability $(n-r-1)^{-1}$ for $r+1 \leqq j \leqq n-1$. Thus

$$\begin{aligned} U^i &= z_1 = X_1 && \text{if } n = 1 \\ &= z_n = X_n && \text{if } n > 1 \text{ and } z_{n-1} \leqq 1 \\ &= (z_{r+1} + \cdots + z_{n-1})(n-r-1)^{-1} && \text{if } n > 1 \text{ and } z_r \leqq 1 < z_{r+1} \end{aligned} \tag{5.8}$$

and convergence occurs in probability and in the mean of order 1 to $\theta = ET$. Presumably convergence with probability 1 also occurs, but in this example it does not seem obvious.

REFERENCES

[1] Berk, Robert H. (1966). Limiting behaviour of posterior distributions when the model is incorrect. *Ann. Math. Statist.* **37** 51–58.

[2] Blackwell, David (1947). Conditional expectation and unbiased sequential estimation. *Ann. Math. Statist.* **18** 105–110.

[3] Doob, J. L. (1953). *Stochastic Processes.* Wiley, New York.

[4] Fraser, D. A. S. (1957). *Nonparametric Methods in Statistics.* Wiley, New York.

[5] Girshick, M. A., Mosteller, Frederick and Savage, L. J. (1946). Unbiased estimates for certain binomial sampling problems with applications. *Ann. Math. Statist.* **17** 13–23.

[6] Hewitt, Edwin and Savage, J. L. (1955). Symmetric measures on cartesian products. *Trans. Amer. Math. Soc.* **80** 470–501.

[7] Hoeffding, W. (1963). The strong law of large numbers for U-statistics. Univ. of N. Carolina. Mimeo series no. 312.

[8] Knight, William (1965). A method of sequential estimation applicable to the hypergeometric, binomial, Poisson and exponential distributions. *Ann. Math. Statist.* **36** 1494–1503.

[9] Lehmann, E. L. and Stein, Charles (1950). Completeness in the sequential case. *Ann. Math. Statist.* **21** 376–385.

[10] Richter, Wolfgang (1965). Limit theorems for sequences of random variables with sequences of random indices. *Theor. Prob. Appl.* **10** 74–84.

[11] Wolfowitz, J. (1946). On sequential binomial estimation. *Ann. Math. Statist.* **17** 489–493.

[12] Wolfowitz, J. (1947). Consistency of sequential binomial estimates. *Ann. Math. Statist.* **18** 131–135.