THE EFFICIENCY OF SEQUENTIAL ESTIMATES AND WALD'S EQUATION FOR SEQUENTIAL PROCESSES

BY J. WOLFOWITZ

Columbia University

1. Summary. Let n successive independent observations be made on the same chance variable whose distribution function $f(x, \theta)$ depends on a single parameter θ. The number n is a chance variable which depends upon the outcomes of successive observations; it is precisely defined in the text below. Let $\theta^*(x_1, \cdots, x_n)$ be an estimate of θ whose bias is $b(\theta)$. Subject to certain regularity conditions stated below, it is proved that

$$\sigma^2(\theta^*) \geq \left(1 + \frac{db}{d\theta}\right)^2 \left[EnE\left(\frac{\partial \log f}{\partial \theta}\right)^2\right]^{-1}.$$

When $f(x, \theta)$ is the binomial distribution and θ^* is unbiased the lower bound given here specializes to one first announced by Girshick [3], obtained under no doubt different conditions of regularity. When the chance variable n is a constant the lower bound given above is the same as that obtained in [2], page 480, under different conditions of regularity.[1]

Let the parameter θ consist of l components $\theta_1, \cdots, \theta_l$ for which there are given the respective unbiased estimates $\theta_1^*(x_1, \cdots, x_n), \cdots, \theta_l^*(x_1, \cdots, x_n)$. Let $\| \lambda_{ij} \|$ be the non-singular covariance matrix of the latter, and $\| \lambda^{ij} \|$ its inverse. The concentration ellipsoid in the space of $(k_1, \cdots, k_l)$ is defined as

$$\sum_{i,j} \lambda^{ij}(k_i - \theta_i)(k_j - \theta_j) = l + 2.$$

(This valuable concept is due to Cramér). If a unit mass be uniformly distributed over the concentration ellipsoid, the matrix of its products of inertia will coincide with the covariance matrix $\| \lambda_{ij} \|$. In [4] Cramér proves that no matter what the unbiased estimates $\theta_1^*, \cdots, \theta_l^*$, (provided that certain regularity conditions are fulfilled), when n is constant their concentration ellipsoid always contains within itself the ellipsoid

$$\sum_{i,j} \mu_{ij}(k_i - \theta_i)(k_j - \theta_j) = l + 2$$

where

$$\mu_{ij} = nE\left(\frac{\partial \log f}{\partial \theta_i} \frac{\partial \log f}{\partial \theta_j}\right).$$

[1] To whom this result is to be ascribed is not clear from the context in which Professor Cramér describes it (in [2]). After the present paper was completed the author learned of the papers by Rao [8] and Aitken and Silverstone [9], both of which deal with this question. The author is indebted to Prof. M. S. Bartlett for drawing his attention to these papers.

Consider now the sequential procedure of this paper. Let $\theta_1^*, \cdots, \theta_l^*$ be, as before, unbiased estimates of $\theta_1, \cdots, \theta_l$, respectively, recalling, however, that the number n of observations is a chance variable. It is proved that the concentration ellipsoid of $\theta_1^*, \cdots, \theta_l^*$ always contains within itself the ellipsoid

$$\sum_{i,j} \mu'_{ij}(k_i - \theta_i)(k_j - \theta_j) = l + 2$$

where

$$\mu'_{ij} = EnE\left(\frac{\partial \log f}{\partial \theta_i}\frac{\partial \log f}{\partial \theta_j}\right).$$

When n is a constant this becomes Cramér's result (under different conditions of regularity).

In section 7 is presented a number of results related to the equation $EZ_n = EnEX$, which is due to Wald [6] and is fundamental for sequential analysis.

2. Introduction. Let X be a chance variable whose distribution function $f(x, \theta)$ depends on the parameter θ. It is assumed that X either has a probability density function (which we then denote by $f(x, \theta)$) or that it can take only an at most denumerable number of discrete values (in the latter case $f(x, \theta) = P\{X = x\}$, where the latter symbol denotes the probability of the relation in braces). Let $\omega = x_1, x_2, \cdots$ be an infinite sequence of observations on X, and let Ω be the space of "points" ω. Let there be given an infinite sequence of Borel measurable functions $\varphi_1(x_1), \varphi_2(x_1, x_2), \cdots, \varphi_j(x_1, \cdots, x_j), \cdots$ defined for all ω in Ω, such that each takes only the values zero and one. It is well known that the function $f(x, \theta)$ defines a measure (probability) on a Borel field in Ω. We assume that everywhere in Ω, except possibly on a set whose probability is zero for all θ under consideration, at least one of the functions $\varphi_1, \varphi_2, \cdots$ takes the value one. Let $n(\omega)$ be the smallest integer at which this occurs. Thus $n(\omega)$ is a chance variable.

In statistical applications the chance variable $n(\omega)$ may be interpreted as a rule for terminating a sequence of observations on the chance variable X, the probability of termination being one, and the decision to terminate depending only upon the observations obtained. A sequential test is an example of this procedure. The converse is, however, not true, because the process described above does not require that any statistical decision should be reached when the process of drawing observations is terminated.

An "estimate" of θ is a function $\theta^*(x_1, \cdots, x_n)$ of the observations $x_1, \cdots, x_n$ (those obtained prior to the "termination" of the process of drawing observations). In the sequel we shall limit ourselves to estimates whose second moments are finite. The estimate is "unbiased" if $E\theta^*$, the expected value of θ^*, is θ. When this is not so $E\theta^* - \theta$ is called the bias, $b(\theta)$, of θ^*. In general the bias is a function of θ. It is obvious that the function θ^* may be undefined on a set of points $(x_1, \cdots, x_n)$ whose probability is zero for all θ under consideration.

In the present paper we shall be concerned with an upper bound on the efficiency of a sequential estimate, or, more precisely, with a lower bound on its variance. This lower bound is intimately related to certain results on the efficiency of the maximum likelihood estimate from a sample of fixed size. This is not surprising since fixed-size sampling is a special instance of sequential sampling. The results obtained in this paper are also obviously and intimately related to those due to Cramér [4] and those described by him in [2], pp. 477–488. Naturally the conditions of regularity (restrictions on $f(x, \theta)$, θ^*, etc.) under which the results are proved are different. For example, no restrictions on the sequential sampling procedure need appear in the statement of a theorem which deals only with samples of fixed size.

The argument below proceeds as if $f(x, \theta)$ were a probability density function. The results apply equally well to the case where $f(x, \theta)$ is the probability function of a discrete chance variable provided:

1). Integration is replaced by summation wherever this is obviously required.

2). The phrase "almost all points" in a Euclidean space of any finite dimensionality is understood

a). as all points in the space with the possible exception of a set of Lebesgue measure zero, when $f(x, \theta)$ is a probability density function

b). as all points in the space with the possible exception of points one of whose coordinates is a member of the set Z, when $f(x, \theta)$ is the probability function of a discrete chance variable. The set Z consists of all points z such that $f(z, \theta) = 0$ identically for all θ under consideration.

3. Conditions of regularity. In this section we shall formulate the restrictions which we impose on f, the estimates, and the sequential process. They are intended to be such as will be satisfied in most cases of statistical interest. No doubt they can be weakened, but the author has decided against attempting to do so here. The list may seem long for two reasons. Seldom in the literature are the assumptions which, for example, lead to validation of differentiation under the integral sign etc., formulated explicitly. The presence of a sequential procedure means that additional restrictions must be imposed.

In this section we assume that θ is a single parameter. The case where θ has more than one component is treated later.

(3.1). *The parameter θ lies in an open interval D of the real line. D may consist of the entire line or of an entire half-line.*

(3.2). *The derivative $\frac{\partial f}{\partial \theta}$ exists for all θ in D and almost all x. We define $\frac{\partial \log f(x, \theta)}{\partial \theta}$ as zero whenever $f(x, \theta) = 0$; thus $\frac{\partial \log f}{\partial \theta}$ is defined for all θ in D and almost all x. We postulate that $E \frac{\partial \log f(x, \theta)}{\partial \theta} = 0$ and that $E \left(\frac{\partial \log f(x, \theta)}{\partial \theta}\right)^2$ be not zero for all θ in D.*

(3.3).
$$E\left(\sum_{i=1}^{n}\left|\frac{\partial \log f(x_i, \theta)}{\partial\theta}\right|\right)^2$$

exists for all θ in D.

(3.4). *Let R_j, ($j = 1, 2, \cdots$), be the set of points $(x_1, \cdots, x_j)$ in the j-dimensional Euclidean space such that*

$$\varphi_i(x_1, \cdots, x_i) = 0 \qquad i = 1, 2, \cdots, j-1$$
$$\varphi_j(x_1, \cdots, x_j) = 1.$$

For any integral j there exists a non-negative L-measurable function $T_j(x_1, \cdots, x_j)$ such that

a).
$$\left|\theta^*(x_1, \cdots, x_j)\frac{\partial}{\partial\theta}\prod_{i=1}^{j} f(x_i, \theta)\right| < T_j(x_1, \cdots, x_j)$$

for all θ in D and almost all $(x_1, \cdots, x_j)$ in R_j

b).
$$\int_{R_j} T_j(x_1, \cdots, x_j)\, dx_1 \cdots dx_j$$

is finite.

(3.5). *Let*

$$t_j(\theta) = \int_{R_j} \theta^*(x_1, \cdots, x_j)\prod_{i=1}^{j} f(x_i, \theta)\, dx_i, \qquad (j = 1, 2, \cdots).$$

We postulate the uniform convergence of the series

$$\sum_j \frac{dt_j(\theta)}{d\theta}$$

(*the existence of* $\frac{dt_j(\theta)}{d\theta}$ *is a consequence of Assumption* (3.4)) *for all θ in D.*

4. The case of one parameter. In this section we assume that $f(x, \theta)$ depends on a single parameter θ. In sections 5 and 6 we shall discuss the case when θ is a vector with more than one component.

We have $E\frac{\partial \log f(x, \theta)}{\partial\theta} = 0$

by (3.2). Define the chance variable

$$Y_n = \sum_{i=1}^{n}\frac{\partial \log f(x_i, \theta)}{\partial\theta}.$$

By an argument almost identical with that of [1], Theorem 1, or of Theorem 7.1 below, we have

(4.1)
$$EY_n = 0.$$

From Theorem 7.2 below we obtain

$$\sigma^2(Y_n) = EnE\left(\frac{\partial \log f(x, \theta)}{\partial \theta}\right)^2. \tag{4.2}$$

Let θ^* $(x_1, \cdots, x_n)$ be an estimate of θ such that

$$E\theta^* = \theta + b(\theta).$$

Then

$$\sum_{j=1}^{\infty} \int_{R_j} \theta^*(x_1, \cdots, x_j) \prod_{i=1}^{j} f(x_i, \theta)\, dx_i = \theta + b(\theta). \tag{4.3}$$

Differentiation of both members of (4.3) with respect to θ (Assumptions (3.4) and (3.5)) gives

$$E\theta^* Y_n = 1 + \frac{db}{d\theta}. \tag{4.4}$$

From (4.1) it follows that (4.4) gives the covariance between θ^* and Y_n. Hence from (4.2)

$$\sigma^2(\theta^*) \geq \left(1 + \frac{db}{d\theta}\right)^2 \left[EnE\left(\frac{\partial \log f(x, \theta)}{\partial \theta}\right)^2\right]^{-1}. \tag{4.5}$$

When the bias $b(\theta)$ is constant, for example when $b(\theta) \equiv 0$ in case θ^* is an unbiased estimate, we have from (4.5)

$$\sigma^2(\theta^*) \geq \left[EnE\left(\frac{\partial \log f(x, \theta)}{\partial \theta}\right)^2\right]^{-1}. \tag{4.6}$$

The equality sign in (4.6) will hold if θ^* may be written as $Z'(\theta)Y_n + Z''(\theta)$, where Z' and Z'' are functions of θ. However, θ^* itself should not be a function of θ if our argument is to remain valid. The subject is connected with the question of the existence of a sufficient estimate.

Let $f(x, \theta)$ be defined as follows:

$$f(x, \theta) = \theta^x(1 - \theta)^{1-x}, \qquad (x = 0 \text{ or } 1; 0 < \theta < 1).$$

Then

$$\frac{\partial \log f(x, \theta)}{\partial \theta} = \frac{x}{\theta} - \frac{(1 - x)}{(1 - \theta)}, \qquad E\left(\frac{\partial \log f}{\partial \theta}\right)^2 = \frac{1}{\theta(1 - \theta)}.$$

Suppose θ^* is unbiased. Then $\sigma^2(\theta^*) \geq \theta(1 - \theta)(En)^{-1}$, a result first given by Girshick [3] under unspecified regularity conditions.

Let the functions $\varphi_1, \varphi_2, \cdots$ be such that $n(\omega)$ is a constant. We are then dealing with samples of fixed size. The result (4.5) is then given in [2], p. 480, under different conditions of regularity.

5. Regularity conditions for the case when θ has more than one component. We suppose that $\theta = (\theta_1, \cdots, \theta_l)$ and that simultaneous estimates

$\theta_1^*(x_1, \cdots, x_n), \cdots, \theta_l^*(x_1, \cdots, x_n)$ of the components of θ are under discussion. In the sequel we shall limit ourselves to the case when these estimates are all unbiased.

We postulate the following regularity conditions which are sufficient to validate section 6:

(5.1). *The covariance matrix of the estimates* $\theta_1^*, \cdots, \theta_l^*$ *is non-singular for all* θ *in* D (*this time* D *is an open interval of the l-dimensional parameter space*).

(5.2). *The conditions of section 3 are satisfied for each* θ_i *and* θ_i^* $(i = 1, \cdots, l)$.

6. The ellipsoid of concentration when θ has more than one component.

Let

$$\theta = (\theta_1, \cdots, \theta_l).$$

We shall first describe briefly the result of Cramér [4] which refers to samples of fixed size $n > l$. Let $\theta_i^*(x_1, \cdots, x_n)$ be an unbiased estimate of θ_i, $(i = 1, \cdots, l)$. Let $\| \lambda_{ij} \|$ be the non-singular covariance matrix of the θ_i^*, and let $\| \lambda^{ij} \|$ be its inverse. The "ellipsoid of concentration" in the space of points $(k_1, \cdots, k_l)$ is defined as

$$\sum_{i,j=1}^{l} \lambda^{ij}(k_i - \theta_i)(k_j - \theta_j) = l + 2. \tag{6.1}$$

If a unit mass be distributed uniformly over this ellipsoid it will have the point $(\theta_1, \cdots, \theta_l)$ as its center of gravity and λ_{ij} as its product of inertia about the corresponding axes. Cramér proves that, subject to certain regularity conditions, there is a fixed ellipsoid

$$\sum_{i,j=1}^{l} \mu_{ij}(k_i - \theta_i)(k_j - \theta_j) = l + 2 \tag{6.2}$$

where

$$\mu_{ij} = nE\left(\frac{\partial \log f}{\partial \theta_i} \frac{\partial \log f}{\partial \theta_j}\right)$$

which is always contained entirely within the concentration ellipsoid of any set of unbiased estimates. The two ellipsoids coincide only under certain conditions, among which is that the θ_i^* be jointly sufficient estimates of the θ_i.

Let us now consider the sequential procedure of this paper and postulate the regularity conditions of section 5. Let

$$K = \| k_{ij} \|$$

be a matrix with real elements such that $| K | = 1$ and let

$$K^{-1} = \| k^{ij} \|$$

be its inverse. Let

$$\| \theta \| = \begin{Vmatrix} \theta_1 \\ \cdot \\ \cdot \\ \cdot \\ \theta_l \end{Vmatrix}, \qquad \| \theta^* \| = \begin{Vmatrix} \theta_1^* \\ \cdot \\ \cdot \\ \cdot \\ \theta_l^* \end{Vmatrix}, \qquad \| \psi \| = \begin{Vmatrix} \psi_1 \\ \cdot \\ \cdot \\ \cdot \\ \psi_l \end{Vmatrix}$$

be column matrices. Suppose

$$\| \psi \| = K \| \theta \|. \tag{6.3}$$

Then

$$\| \theta \| = K^{-1} \| \psi \|. \tag{6.4}$$

Define

$$\| \psi^* \| = \begin{Vmatrix} \psi_1^* \\ \cdot \\ \cdot \\ \cdot \\ \psi_l^* \end{Vmatrix} = K \| \theta^* \|.$$

From section 4 we have

$$EnE \left(\frac{\partial \log f(x, \theta)}{\partial \psi_1} \right)^2 \geq [\sigma^2(\psi_1^*)]^{-1} \tag{6.5}$$

where the differentiation by which $\frac{\partial \log f}{\partial \psi_1}$ is obtained is performed with $\psi_2, \cdots, \psi_l$ held constant. Consider the last $(l - 1)$ rows of K as fixed and $(k_{11}, k_{12}, \cdots, k_{1l})$ as free to vary subject only to the restriction that $|K| = 1$. The left member of (6.5) is then a fixed quantity, while the right member is a function of the first row of K. The inequality (6.5) must remain valid for all admissible $(k_{11}, \cdots, k_{1l})$. Hence (6.5) will remain valid if the right member of (6.5) is replaced by its maximum with respect to $(k_{11}, \cdots, k_{1l})$. We shall obtain this maximum and find that (6.5) then implies a result about the minimal ellipsoid of concentration.

The problem is therefore to minimize $\sigma^2(\psi_1^*)$. Now

$$\sigma^2(\psi_1^*) = \sum_{i,j} \lambda_{ij} k_{1i} k_{1j}. \tag{6.6}$$

The family of ellipsoids in the space of $(k_{11}, \cdots, k_{1l})$

$$\sum_{i,j} \lambda_{ij} k_{1i} k_{1j} = c, \tag{6.7}$$

where c is a running parameter, has all centers located at the origin. Let

$$(k_{11}^0, \cdots, k_{1l}^0)$$

be the sought-for maximizing values of $(k_{11}, \cdots, k_{1l})$. From the definitions of K and K^{-1} we have

$$\sum_i k^{i1} k_{1i} = 1 \tag{6.8}$$

where $(k^{11}, k^{21}, \cdots, k^{l1})$ are constants. It follows that the minimum value c_0 of $\sigma^2(\psi_1^*)$ is such that the ellipsoid

$$\sum_{i,j} \lambda_{ij} k_{1i} k_{1j} = c_0 \tag{6.9}$$

is tangent to the hyperplane (6.8) at the point $(k_{11}^0, \cdots, k_{1l}^0)$. Now the tangent plane to (6.9) at this point is given by

$$\sum_{i,j} \lambda_{ij} k_{1i}^0 k_{1j} = c_0. \tag{6.10}$$

From (6.8) and (6.10) we obtain

$$c_0 k^{j1} = \sum_i k^0_{1i} \lambda_{ij}, \qquad (j = 1, \cdots, l). \tag{6.11}$$

Hence

$$c_0 \sum_i \lambda^{ij} k^{i1} = k^0_{1j}, \qquad (j = 1, \cdots, l) \tag{6.12}$$

from which

$$c_0 \sum_{i,j} \lambda^{ij} k^{i1} k^{j1} = 1. \tag{6.13}$$

We have

$$\frac{\partial \log f}{\partial \psi_1} = \sum_i k^{i1} \frac{\partial \log f}{\partial \theta_i}$$
$$\left(\frac{\partial \log f}{\partial \psi_1}\right)^2 = \sum_{i,j} k^{i1} k^{j1} \frac{\partial \log f}{\partial \theta_i} \frac{\partial \log f}{\partial \theta_j}. \tag{6.14}$$

From (6.5), (6.13), (6.14), and the definition of c_0 we conclude that

$$\sum_{i,j} \mu'_{ij} k^{i1} k^{j1} \geq \sum_{i,j} \lambda^{ij} k^{i1} k^{j1} \tag{6.15}$$

where

$$\mu'_{ij} = EnE\left(\frac{\partial \log f}{\partial \theta_i} \frac{\partial \log f}{\partial \theta_j}\right). \tag{6.16}$$

We may restate (6.15) as follows: The concentration ellipsoid

$$\sum_{i,j} \lambda^{ij}(k_i - \theta_i)(k_j - \theta_j) = l + 2 \tag{6.17}$$

of the unbiased estimates $\theta_1^*, \cdots, \theta_l^*$ always contains within itself the ellipsoid

$$\sum_{i,j} \mu'_{ij}(k_i - \theta_i)(k_j - \theta_j) = l + 2 \tag{6.18}$$

where the μ'_{ij} are defined by (6.16).

The question of the coincidence of the two ellipsoids is connected with the question of the existence of sufficient estimates. It may be difficult to state any general results about the concentration ellipsoid of biased estimates without postulating some relationships among the biases and/or their derivatives.

7. On Wald's equation and related results in sequential analysis. In section 4 we referred to a proof by Blackwell [1] of an equation due to Wald [5] which is fundamental in the Wald theory of sequential tests of statistical hypotheses. Here we shall give a perhaps simpler proof of this equation, and then prove several new and related results of general interest for sequential analysis.

The results of Theorems 7.2 and 7.3 below can be obtained by differentiation of Wald's fundamental identity of sequential analysis ([6], [7]). However, the conditions under which we obtain these results are less stringent than any so far found sufficient to establish the identity and the validity of differentiating it. Theorem 7.4 and its corollaries refer to sequential processes where the chance variables may have different distributions or even be dependent. In the future we hope to return to the question of finding all central moments of Z_n, the problem of generalizing the fundamental identity, and related questions.

For Theorems 7.1, 7.2, and 7.3 we shall assume a chance variable X whose cumulative distribution function $F(x)$ is subject only to whatever restrictions may be explicitly imposed on it in each theorem. We assume the existence of a general sequential process such as is described above, which is subject only to such restrictions as may be explicitly formulated in each theorem. The sequential process of course defines the chance variable n. Let $x_1, x_2, \cdots$ be successive independent observations on X. We define $Z_n = \sum_{i=1}^{n} x_i$. If $E(X)$ and $\sigma^2(X)$ exist we shall denote them by w and σ^2, respectively.

THEOREM 7.1 (Wald [5], Blackwell [1]). *Suppose w and En exist. Then*

$$E(Z_n - nw) = 0. \tag{7.1}$$

The following theorem, which is a sort of partial converse of Theorem 7.1, is proved concomitantly with Theorem 7.1:

THEOREM 7.1.1. *If EZ_n exists, and if either $P\{X > 0\} = 0$ or $P\{X < 0\} = 0$, then w and En both exist, and*

$$EZ_n = wEn.$$

Actually the same proof suffices for a somewhat stronger form of Theorem 7.1.1:

THEOREM 7.1.2. *If EZ_n exists, and if*

$$E(X_i \mid n = j) \geq 0 \qquad (\text{or } \leq 0)$$

for all positive integral j such that $P\{n = j\} \neq 0$, and all $i \leq j$, then w and En both exist, and

$$EZ_n = wEn\,.$$

THEOREM 7.2. *If $E\left(\sum_{i=1}^{n} \mid x_i - w \mid\right)^2$ exists, then σ^2 and En both exist, and*

$$E(Z_n - nw)^2 = \sigma^2 En\,. \tag{7.2}$$

We have

$$\begin{aligned} E(Z_n - nw) &= E\left(\sum_{i=1}^{n} (x_i - w)\right) = \sum_{j=1}^{\infty} \int_{R_j} \left(\sum_{i=1}^{j} (x_i - w)\right) \prod_{i=1}^{j} dF(x_i) \\ &= \sum_{j=1}^{\infty} \sum_{i=j}^{\infty} \int_{R_i} (x_j - w) \prod_{m=1}^{m=i} dF(x_m). \end{aligned} \tag{7.3}$$

Also

$$\sum_{i=j}^{\infty} \int_{R_i} (x_j - w) \prod_{m=1}^{m=i} dF(x_m) = P\{n \geq j\} E(x_j - w) = 0. \tag{7.4}$$

Hence

$$\sum_{j=1}^{\infty} \sum_{i=j}^{\infty} \int_{R_i} (x_j - w) \prod_{m=1}^{m=i} dF(x_m) = 0. \tag{7.5}$$

From this (7.1) follows.

Suppose now that the conditions of Theorem 7.2 are fulfilled. We have

$$\begin{aligned} E(Z_n - nw)^2 &= \sum_{j=1}^{\infty} \int_{R_j} \left(\sum_{i=1}^{j} (x_i - w) \right)^2 \prod_{m=1}^{m=j} dF(x_m) \\ &= \sum_{j=1}^{\infty} \sum_{i=j}^{\infty} \int_{R_i} (x_j - w)^2 \prod_{m=1}^{m=i} dF(x_m) \\ &\quad + 2 \sum_{j=2}^{\infty} \sum_{s=1}^{j-1} \sum_{i=j}^{\infty} \int_{R_i} (x_s - w)(x_j - w) \prod_{m=1}^{m=i} dF(x_m). \end{aligned} \tag{7.6}$$

Let $s < j$ be any two positive integers. Then

$$\sum_{i=j}^{\infty} \int_{R_i} (x_s - w)(x_j - w) \prod_{m=1}^{m=i} dF(x_m) = 0. \tag{7.7}$$

Hence

$$\sum_{j=2}^{\infty} \sum_{s=1}^{j-1} \sum_{i=j}^{\infty} \int_{R_i} (x_s - w)(x_j - w) \prod_{m=1}^{m=i} dF(x_m) = 0. \tag{7.8}$$

In a similar manner we obtain

$$\sum_{i=j}^{\infty} \int_{R_i} (x_j - w)^2 \prod_{m=1}^{m=i} dF(x_m) = \sigma^2 P\{n \geq j\}. \tag{7.9}$$

From (7.6), (7.8), and (7.9) it therefore follows that

$$E(Z_n - nw)^2 = \sigma^2 \sum_{j=1}^{\infty} P\{n \geq j\} = \sigma^2 \sum_{j=1}^{\infty} jP\{n = j\} = \sigma^2 En \tag{7.10}$$

which is the desired result.

It remains to prove the validity of rearranging the series in (7.3) and (7.6). First, we have

$$\sum_{i=j}^{\infty} \int_{R_i} | x_j - w | \prod_{m=1}^{m=i} dF(x_m) = P\{n \geq j\} E | X - w |. \tag{7.11}$$

Hence it follows that

$$(7.12)\quad \begin{aligned}\sum_{j=1}^{\infty}\sum_{i=j}^{\infty}\int_{R_i} |x_j - w| \prod_{m=1}^{m=i} dF(x_m) &= \sum_{j=1}^{\infty} P\{n \geq j\} E|X - w| \\ &= E|X - w| \sum_{j=1}^{\infty} jP\{n = j\} = E|X - w| En.\end{aligned}$$

This justifies the rearrangement of terms in the series in (7.3). Second, the series (7.6) is dominated by the series

$$(7.13)\quad \begin{aligned}&\sum_{j=1}^{\infty}\sum_{i=j}^{\infty}\int_{R_i} (x_j - w)^2 \prod_{m=1}^{m=i} dF(x_m) \\ &\qquad + 2\sum_{j=2}^{\infty}\sum_{s=1}^{j-1}\sum_{i=j}^{\infty}\int_{R_i} |x_s - w|\cdot|x_j - w| \prod_{m=1}^{m=i} dF(x_m)\end{aligned}$$

all of whose terms are positive. The series (7.13) converges because

$$(7.14)\quad E\left(\sum_{i=1}^{n} |x_i - w|\right)^2 < +\infty.$$

Hence the rearrangement of the series (7.6) is valid.

In the sequel we require certain sets $R_j'(j = 1, 2, \cdots)$ which we shall define now. Let R_{ij}^*, $i \leq j$, be the totality of all points $(x_1, \cdots, x_j)$ such that

$$(7.15)\quad (x_1, \cdots, x_i) \in R_i.$$

Let R^j be the j-dimensional Euclidean space. Then

$$(7.16)\quad R_j' = R^j - \sum_{i=1}^{j} R_{ij}^*.$$

We shall now prove:

THEOREM 7.3. *Suppose that* $E\left[\sum_{i=1}^{n} |x_i - w|\right]^3$ *and* $En\left[\sum_{i=1}^{n} |x_i - w|\right]$ *exist.*[2] *Then*

$$(7.17)\quad E(Z_n - nw)^3 = w_3En + 3\sigma^2En(Z_n - nw)$$

where

$$w_3 = E(X - w)^3$$

exists.

---

[2] The author has succeeded in proving that the existence of $E\left[\sum_{i=1}^{n} |x_i - w|\right]^3$ implies the existence of $E\left[n\sum_{i=1}^{n} |x_i - w|\right]$. The proof will be published subsequently in connection with other results.

PROOF: We have

$$E(Z_n - nw)^3 = \sum_{j=1}^{\infty} \int_{R_j} \left[\sum_{i=1}^{j} (x_i - w)\right]^3 \prod_{m=1}^{j} dF(x_m)$$

$$= \sum_{j=1}^{\infty} \int_{R_j} \sum_{i=1}^{j} (x_i - w)^3 \prod_{m=1}^{j} dF(x_m)$$

$$(7.18) \qquad + 3 \sum_{j=2}^{\infty} \int_{R_j} \sum_{i=2}^{j} \sum_{s=1}^{i-1} (x_s - w)(x_i - w)^2 \prod_{m=1}^{j} dF(x_m)$$

$$+ 3 \sum_{j=2}^{\infty} \int_{R_j} \sum_{i=2}^{j} \sum_{s=1}^{i-1} (x_s - w)^2(x_i - w) \prod_{m=1}^{j} dF(x_m)$$

$$+ 6 \sum_{j=3}^{\infty} \int_{R_j} \sum_{i=3}^{j} \sum_{s=2}^{i-1} \sum_{t=1}^{s-1} (x_t - w)(x_s - w)(x_i - w) \prod_{m=1}^{j} dF(x_m).$$

Considering the first term in the right member of (7.18), it follows that

$$\sum_{j=1}^{\infty} \int_{R_j} \left[\sum_{i=1}^{j} (x_i - w)^3\right] \prod_{m=1}^{j} dF(x_m)$$

$$(7.19) \qquad = \sum_{i=1}^{\infty} \sum_{j=i}^{\infty} \int_{R_j} (x_i - w)^3 \prod_{m=1}^{j} dF(x_m)$$

$$= \sum_{i=1}^{\infty} w_3 P\{n \geq i\}$$

$$= \sum_{i=1}^{\infty} i w_3 P\{n = i\} = w_3 En.$$

All the rearrangements of terms in the operations involved in the proof of Theorem 7.3 are legitimate because the various series are absolutely convergent.

As for the second term in the right member of (7.18), we have

$$\sum_{j=2}^{\infty} \int_{R_j} \sum_{i=2}^{j} \sum_{s=1}^{i-1} (x_s - w)(x_i - w)^2 \prod_{m=1}^{j} dF(x_m)$$

$$(7.20) \qquad = \sum_{s=1}^{\infty} \sum_{i=s+1}^{\infty} \sum_{j=i}^{\infty} \int_{R_j} (x_s - w)(x_i - w)^2 \prod_{m=1}^{j} dF(x_m)$$

$$= \sigma^2 \sum_{s=1}^{\infty} \sum_{i=s+1}^{\infty} \int_{R'_{i-1}} (x_s - w) \prod_{m=1}^{i-1} dF(x_m)$$

$$= \sigma^2 \sum_{s=1}^{\infty} \sum_{i=s}^{\infty} \int_{R'_i} (x_s - w) \prod_{m=1}^{i} dF(x_m).$$

We now operate on $En(Z_n - nw)$, and obtain

$$(7.21) \qquad En(Z_n - nw) = \sum_{j=1}^{\infty} \int_{R_j} j \sum_{i=1}^{j} (x_i - w) \prod_{m=1}^{j} dF(x_m)$$

$$= \sum_{j=1}^{\infty} \sum_{i=j}^{\infty} \int_{R_i} i(x_j - w) \prod_{m=1}^{i} dF(x_m).$$

We observe that

$$\sum_{i=j}^{\infty} \int_{R_i} i(x_j - w) \prod_{m=1}^{i} dF(x_m)$$
$$(7.22) \qquad = j \sum_{i=j}^{\infty} \int_{R_i} (x_j - w) \prod_{m=1}^{i} dF(x_m)$$
$$+ \sum_{s=j+1}^{\infty} \sum_{i=s}^{\infty} \int_{R_i} (x_j - w) \prod_{m=1}^{i} dF(x_m).$$

To evaluate the left member of (7.22), we proceed as follows: It is easy to see that

$$(7.23) \qquad \sum_{i=j}^{\infty} \int_{R_i} (x_j - w) \prod_{=1}^{i} dF(x_m) = 0.$$

Moreover, when $s > j$,

$$(7.24) \qquad \sum_{i=s}^{\infty} \int_{R_i} (x_j - w) \prod_{m=1}^{i} dF(x_m) = \int_{R'_{s-1}} (x_j - w) \prod_{m=1}^{s-1} dF(x_m).$$

Hence

$$(7.25) \qquad \sum_{i=j}^{\infty} \int_{R_i} i(x_j - w) \prod_{m=1}^{i} dF(x_m) = \sum_{s=j}^{\infty} \int_{R'_s} (x_j - w) \prod_{m=1}^{s} dF(x_m).$$

Therefore

$$(7.26) \qquad En(Z_n - nw) = \sum_{j=1}^{\infty} \sum_{s=j}^{\infty} \int_{R'_s} (x_j - w) \prod_{m=1}^{s} dF(x_m).$$

It remains now to consider the third term of the right member of (7.18). We have

$$\sum_{j=2}^{\infty} \int_{R_j} \sum_{i=2}^{j} \sum_{s=1}^{i-1} (x_s - w)^2 (x_i - w) \prod_{m=1}^{j} dF(x_m).$$
$$(7.27) \qquad = \sum_{s=1}^{\infty} \sum_{i=s+1}^{\infty} \sum_{j=i}^{\infty} \int_{R_j} (x_s - w)^2 (x_i - w) \prod_{m=1}^{j} dF(x_m).$$

Now, suppose that in the expression

$$(7.28) \qquad V_{sij} = \int_{R_j} (x_s - w)^2 (x_i - w) \prod_{m=1}^{j} dF(x_m)$$

where $j \geq i > s$, we integrate with respect to all x_m for which $m \geq i$. Then it is not difficult to see that

$$(7.29) \qquad \sum_{j=i}^{\infty} V_{sij} = 0$$

for all s and i such that $1 \leq s < i$. Hence from (7.27)

$$(7.30) \qquad \sum_{j=2}^{\infty} \int_{R_j} \sum_{i=2}^{j} \sum_{s=1}^{i-1} (x_s - w)^2 (x_i - w) \prod_{m=1}^{j} dF(x_m) = 0.$$

In a similar way it is shown that the fourth term of the right member of (7.18) is zero.

The desired result (7.17) is a direct consequence of (7.18), (7.19), (7.20), (7.26), and (7.30).

Consider now an infinite sequence of chance variables $x_1, x_2, \cdots$, which need not have the same distribution and which may be dependent (in which case they must satisfy the obvious consistency relationships). We take successive observations on these chance variables and define a sequential process as above, which is subject only to such restrictions as we shall explicitly state. Let Z_n maintain its previous definition.

THEOREM 7.4. *Suppose that*

$$\nu_i = E(X_i \mid n \geq i) \tag{7.31}$$

exists for all positive integral i for which $P\{n \geq i\} \neq 0$. In those cases write

$$\nu_i' = E(| X_i - \nu_i | \mid n \geq i). \tag{7.32}$$

Suppose also that the series

$$\sum_{i=1}^{\infty} (\nu_1' + \cdots + \nu_i')P\{n = i\} \tag{7.33}$$

converges. Then

$$E\left[Z_n - \sum_{i=1}^{n} \nu_i\right] = 0. \tag{7.34}$$

It is regrettable but unavoidable that the mean values ν_i and ν_i' entering into (7.33) and (7.34) be conditional. The fundamental reason is that the sequential process may drastically modify the distribution of dependent chance variables, so that their distribution for our purposes can only be considered in conjunction with the sequential process itself. Consider the following example:

$$P\{X_1 = -1\} = \tfrac{1}{2}, \qquad P\{X_1 = 1\} = \tfrac{1}{2}$$

$$P\{X_2 = -2 \mid X_1 = -1\} = \tfrac{1}{2}$$

$$P\{X_2 = -1 \mid X_1 = -1\} = \tfrac{1}{2}$$

$$P\{X_2 = 1 \mid X_1 = 1\} = \tfrac{1}{2}$$

$$P\{X_2 = 2 \mid X_1 = 1\} = \tfrac{1}{2}.$$

We have $E(X_2) = 0$. Suppose we define the following sequential process: If $X_1 = -1, n = 1$, and if $X_1 = 1, n = 2$. It is then clear that for our purposes X_2 can take no negative values and the fact that $E(X_2) = 0$ is of no use to us.

If, however, the chance variables $X_1, X_2, \cdots$ are independent, this difficulty disappears, and we have the following.

COROLLARY 1 TO THEOREM 7.4. *If the chance variables $X_1, X_2, \cdots$ are independent, we have Theorem 7.4 with $\nu_i = E(X_i)$, and $\nu_i' = E \mid X_i - \nu_i \mid$.*

If further all the X_i have the same distribution, we see that Theorem 7.1 is a special case of Theorem 7.4, since the convergence of the series (7.33) is then a consequence of the existence of w and En. From this argument we see, however, that it is not necessary that all the X_i have the same distribution, and we may write the following generalization of Theorem 7.1:

COROLLARY 2 TO THEOREM 7.4. *Let the X_i be independent with, in general, different distributions. Suppose, however, that all ν_i are equal, and all ν_i' are equal, except perhaps for those i such that $P\{n \geq i\} = 0$. Suppose further that En exists. Then (7.1) holds.*

Among possible fields of application of Theorem 7.4 are sequential tests of composite statistical hypotheses, and the random walk of a particle governed by probability distributions which are functions of time and the position of the particle. The extension of this theorem to vector chance variables is straightforward. The extension to higher moments may present difficulties. We hope to return to some of these questions in the future.

PROOF OF THEOREM 7.4. This is very elementary. We have

$$\begin{aligned} E\left(Z_n - \sum_{i=1}^{n} \nu_i\right) &= \sum_{j=1}^{\infty} \int_{R_j} \left[\sum_{i=1}^{j} (x_i - \nu_i)\right] dF(x_1, \cdots, x_j) \\ &= \sum_{j=1}^{\infty} \sum_{i=j}^{\infty} \int_{R_i} (x_j - \nu_j)\, dF(x_1, \cdots, x_i). \qquad (7.35) \\ &= \sum_{j=1}^{\infty} P\{n \geq j\} E(X_j - \nu_j \mid n \geq j) = 0. \end{aligned}$$

The rearrangement of the series is valid because

$$\begin{aligned} \sum_{j=1}^{\infty} \sum_{i=j}^{\infty} \int_{R_i} \mid x_j - \nu_j \mid dF(x_1, \cdots, x_i) &= \sum_{j=1}^{\infty} \nu_j' P\{n \geq j\} \qquad (7.36) \\ &= \sum_{j=1}^{\infty} (\nu_1' + \cdots + \nu_j') P\{n = j\} \end{aligned}$$

which converges by (7.33).

REFERENCES

[1] DAVID BLACKWELL, "On an equation of Wald," *Annals of Math. Stat.*, Vol. 17 (1946), pp. 84–87.

[2] H. CRAMÉR, *Mathematical Methods of Statistics*, Princeton Univ. Press, 1946.

[3] M. A. GIRSHICK, FREDERICK MOSTELLER, AND L. J. SAVAGE, "Unbiased estimates for certain binomial sampling problems," *Annals of Math. Stat.*, Vol. 17 (1946), pp. 13–23.

[4] H. CRAMÉR, "A contribution to the theory of statistical estimation," *Skandinavisk Aktuarietidskrift*, Vol. 29 (1946), pp. 85–94.

[5] A. WALD, "Sequential tests of statistical hypotheses," *Annals of Math. Stat.*, Vol. 16 (1945), pp. 117–186.

[6] A. WALD, "On cumulative sums of random variables," *Annals of Math. Stat.*, Vol. 15 (1944), pp. 283–296.

[7] A. WALD, "Differentiation under the expectation sign of the fundamental identity in sequential analysis," *Annals of Math. Stat.*, Vol. 17 (1946).

[8] C. R. RAO, "Information and the accuracy attainable in the estimation of statistical parameters," *Bull. Calcutta Math. Soc.*, Vol. 37, No. 3 (Sept., 1945), pp. 81–91.

[9] A. C. AITKEN AND H. SILVERSTONE "On the estimation of statistical parameters," *Proc. Roy. Soc. Edinburgh*, Vol. 61 (1941), pp. 56–62.