SAMPLING FROM A CHANGING POPULATION[1, 2]

By Reinhold Baer

University of Illinois

1. Introduction. If, in sampling a certain population, it is impossible to take more than one sample at any given time, and if the population changes between any two samples, then we are confronted with the following mathematical situation. For every[3] t, $0 \leq t \leq 1$, there is given a distribution[4] (= population) $D(t)$. Let furthermore t_j be, for $0 < j \leq n$, a number between $(j - 1)/n$ and j/n; and assume that x_j is a sample taken from the population $D(t_j)$. We denote by T_n the set of the numbers $t_1, \cdots, t_n$ and by $O(T_n)$ the sample consisting of the x_j; and we assume that $O(T_n)$ is a random sample, i.e. that $x_1, \cdots, x_n$ are independent variables. The question arises to get information concerning the family $D(t)$ from the sample $O(T_n)$. It is clearly hopeless to try for information concerning an individual $D(t)$ or even some $D(t_j)$ or the statistics that may be derived from them. But we may hope for information in the mean, if we assume that the family $D(t)$ is in some sense continuous in t. To make this statement more precise we denote by $a(t)$ the average and by $M_i(t)$ the i-th moment of $D(t)$ around its average. We assume then that $a(t)$ and $M_i(t)$, for $i \leq 8$, exist and are continuous functions of t, and in section 7 we shall have to assume furthermore that $a(t)$ and $M_2(t)$ are functions of bounded variation. These hypotheses assure the existence of

$$\text{the mean average } a = \int_0^1 a(t)\,dt$$

$$\text{and the mean } i\text{-th moment } M_i = \int_0^1 M_i(t)\,dt$$

for $i \leq 8$. Clearly we may hope for information concerning a and M_i from the random sample $O(T_n)$. It is our object to discuss certain more or less well known statistics of the sample $O(T_n)$, and to determine their stochastic limits[5].

---

[1] Presented to the American Mathematical Society. September 15, 1945.

[2] The author is indebted to Dr. E. L. Welker for checking the results, in particular those rather obnoxious computations needed in sections 6 and 7 which the author did not incorporate into this paper.

[3] It constitutes a restriction of generality that we consider finite closed intervals only. But it is no further loss in generality to use the interval from 0 to 1, and this choice certainly simplifies notations.

[4] Comparatively little will be assumed of these distributions. These properties will be enumerated in Section 2.

[5] See [2] p. 81 and the criterion 2.d. of section 2.

As an illustration we mention the following results which will be obtained in the course of this investigation (among others):[6]

$\bar{x} = n^{-1} \sum_{j=1}^{n} x_j$ converges stochastically to the mean average a;

$s^2 = n^{-1} \sum_{j=1}^{n} (x_j - \bar{x})^2$ converges stochastically to $M_2 + \int_0^1 (a(t) - a)^2 \, dt$;

$d^2 = (2n)^{-1} \sum_{j=1}^{n-1} (x_j - x_{j+1})^2$ converges stochastically to the mean variance M_2.

It is clear that M_2 is the stochastic limit of s^2 if, and only if, $a(t)$ is constant. If $a(t)$ is not constant, then s^2 is not a consistent estimate[7] of M_2, and will have to be rejected—at least for large n—in favor of d^2 which is always a consistent estimate of M_2.

It was this last point that led us into this investigation. Recently the statistic d^2 has found much attention; and the question arose as to why the statistic s^2 should be rejected in favor of d^2. Reading the illuminating introduction of the fundamental paper [1], one sees that just such a situation as we have attempted to describe here in somewhat abstract terms has necessitated the use of d^2. Consequently our result may be considered a theoretical justification for this procedure.

Our other results will be discussed in their interrelation as they are obtained. It should be noted that all our results concern themselves with stochastic convergence, and thus they justify the use of a sample function as an estimate of some statistical number only for sufficiently large size n of the sample. Thus it is quite possible that for small n other functions provide better estimates. The practical applicability of our results depends, therefore, on a criterion for n to be sufficiently large, and unfortunately such a criterion is not yet available.

2. Notations and fundamental properties. We have not stated in the Introduction the hypotheses to which we subject the distributions under consideration. For our investigation we shall need only very few properties of distributions. Thus we are going to enumerate now some properties of distributions which we are going to use, and we shall assume throughout that these properties are satisfied. As will be seen these hypotheses are rather weak and are satisfied by a large class of distributions.

If x is any stochastic variable, then we denote by $E(x)$ its mathematical expectation, and the only properties of stochastic variables that concern us are properties of their expectations. $E(x)$ is a linear operation satisfying $E(1) = 1$.

[6] It should be noted that the stochastic limit of the following statistics would not be changed, if we substituted for the denominator n of s^2 the denominator $n - 1$ which is often used, and if we allowed the summation in the expression for d^2 to range from 1 to n, defining x_{n+1} as x_1.

[7] Wilks [2], p. 133.

If furthermore $x_1, \cdots, x_n$ are independent variables, and if the function f depends on some of these variables whereas g depends only on the others, then $E(fg) = E(f)E(g)$, and this property may serve as a definition of independence.

As stated in the Introduction we are going to study a family $D(t)$ of distributions, for $0 \leq t \leq 1$. If x is the stochastic variable of the distribution $D(t)$ for some fixed t, then we let

$$a(t) = E(x) \quad \text{and} \quad M_i(t) = E((x - a(t))^i).$$

We shall assume throughout that the average $a(t)$ and the variance $M_2(t)$ exist for every t, and that $a(t)$ and $M_2(t)$ are continuous functions of t. Moreover, when discussing $M_i(\tau)$, $1 \leq i \leq 4$, we shall assume that every $M_j(\tau)$ with $j \leq 2i$ is a continuous function of τ. Thus we are sure that the mean average a and the mean variance M_2, as defined in the Introduction, always exist, and the mean i-th moment M_i exists, whenever $M_i(t)$ is a continuous function of t.

Remark: If the mean i-th moment M_i exists for every i, then one may be tempted to consider as the mean of the family $D(t)$ a distribution D with average a and i-th moment M_i, provided such a distribution exists. But this has to be done with some caution. For suppose that every $D(t)$ is normal. Then $M_i(t) = 0$ for every odd i, implying $M_i = 0$ for odd i so that D would be symmetric. But $M_{2i}(t) = 1 \cdot 3 \cdots (2i - 1)M_2(t)^i$ and hence $M_{2i} = 1 \cdot 3 \cdots (2i - 1) \cdot \int_0^1 M_2(t)^i \, dt$, and the integral will be the i-th power of M_2 only if $M_2(t)$ is constant. Thus the mean distribution D of a continuous family of normal distributions need not be normal.

As in the Introduction we now let t_i be some number between $(i - 1)/n$ and i/n, and denote by x_i a sample taken from the distribution $D(t_i)$. We denote by T_n the set of the n numbers t_i and by $O(T_n)$ the sample consisting of the x_i. It will be assumed throughout that $O(T_n)$ is a random sample, i.e. we shall assume that $x_1, \cdots, x_n$ are independent variables.

We are not going to make any use of the customary definition of stochastic convergence[8] (and we shall therefore not restate it). Instead we are going to apply throughout the following criterion[9, 10]:

2.d. *The function $f(O(T_n))$ of the sample $O(T_n)$ converges stochastically to the number r, if*

$$\lim_{n \to \infty} E(f(O(T_n))) = r \text{ and } \lim_{n \to \infty} E([f(O(T_n)) - E(f(O(T_n)))]^2) = 0.$$

All the sample functions considered will be polynomials of the variables $x_1, \cdots, x_n$.

[8] Wilks [2], p. 81.

[9] Wilks [2], Theorem (A), p. 134.

[10] The validity of criterion 2.d. implies stochastic convergence in the customary sense. Thus, all results obtained in the present paper remain valid also when the customary definition of stochastic convergence is adopted.

3. The mean average. Though the discussion of this section is rather obvious, we give the details, since they may serve as a convenient introduction to the type of argument we have to use throughout.

THEOREM. *$\bar{x}$ converges stochastically to a.*

PROOF: We note first that $E(\bar{x}) = n^{-1} \sum_{j=1}^{n} E(x_j) = n^{-1} \sum_{j=1}^{n} a(t_j)$. Since t_j is between $(j-1)/n$ and j/n, and since n^{-1} is the length of this interval, it follows from the continuity of $a(t)$ that

$$\int_0^1 a(t)\,dt = \lim_{n\to\infty} n^{-1} \sum_{j=1}^{n} a(t_j);$$

and thus we have shown that $E(\bar{x})$ tends to a as n tends to infinity.

Next we find that

$$E((\bar{x} - E(\bar{x}))^2) = n^{-2} E\left(\left[\sum_{j=1}^{n} (x_j - a(t_j))\right]^2\right)$$
$$= n^{-2} \sum_{j=1}^{n} E((x_j - a(t_j))^2) = n^{-2} \sum_{j=1}^{n} M_2(t_j),$$

since $E((x_j - a(t_j))(x_h - a(t_h))) = E(x_j - a(t_j))E(x_h - a(t_h)) = 0$ for $j \neq h$. But $M_2(t)$ is, for $0 \leq t \leq 1$, a bounded non-negative function, showing that $E((\bar{x} - E(\bar{x}))^2)$ tends to 0 as n tends to infinity. Applying 2.d. we find that $\bar{x}$ converges stochastically to a, as we intended to show.

Remark: It is clear that the speed of the stochastic convergence of $\bar{x}$ to a depends on two factors:

(i) the goodness of $\bar{x}$ as an estimate of $E(\bar{x})$;

(ii) the speed of convergence of the sums $n^{-1} \sum_{j=1}^{n} a(t_j)$ to the integral $a = \int_0^1 a(t)\,dt$.

It is this difficulty which expresses itself in (ii) and which makes the present type of statistical estimation less effective than the one concerned with sampling from one distribution only. As to (i), it is again, as may be seen from the proof, of the order of magnitude $(M_2/n)^{\frac{1}{2}}$, (see Theorem 1, section 4).

It is probable that $\bar{x}$ is a better estimate of $E(\bar{x})$ than of a. But this does not help, since the former depends on the particular choice of T_n.

4. The variance. THEOREM 1. *d^2 converges stochastically to M_2.*

PROOF: We note first that

$$E((x_j - x_{j+1})^2 = E([(x_j - a(t_j)) + (a(t_j) - a(t_{j+1})) + (a(t_{j+1}) - x_{j+1})]^2)$$
$$= M_2(t_j) + (a(t_j) - a(t_{j+1}))^2 + M_2(t_{j+1}),$$

since $E((x_j - a(t_j))(x_{j+1} - a(t_{j+1}))) = E(x_j - a(t_j))E(x_{j+1} - a(t_{j+1})) = 0$, $E(\text{const}) = \text{const}$ and $E((x_i - a(t_i))^2) = M_2(t_i)$. Hence

$$E(d^2) = (2n)^{-1}(A + B - C),$$

where $A = 2\sum_{j=1}^{n} M_2(t_j)$, $B = \sum_{j=1}^{n-1} (a(t_j) - a(t_{j+1}))^2$, $C = M_2(t_1) + M_2(t_n)$. Since t_j is a value between $(j-1)/n$ and j/n, and since n^{-1} is the length of this interval, it follows from the continuity of the function $M_2(t)$ that $M_2 = \int_0^1 M_2(t)\,dt = \lim_{n\to\infty} (2n)^{-1}A$. Since $M_2(t)$ is bounded as a continuous function, it follows that $(2n)^{-1}C$ tends to 0 as n tends to infinity. Finally we infer from the continuity of $a(t)$—which is used here for the first time to its full extent—that there exists to every given positive ϵ an integer $N = N(\epsilon)$ such that $(a(t') - a(t''))^2 < \epsilon$ for $|t' - t''| < (2N)^{-1}$. Thus for $N(\epsilon) < n$ we have $(a(t_j) - a(t_{j+1}))^2 < \epsilon$ and $(2n)^{-1}B < n\frac{1}{2n}\epsilon$. Hence $(2n)^{-1}B$ tends to 0 as n tends to infinity, and we have shown that

$E(d^2)$ tends to M_2 as n tends to infinity.

Next we note that

$$E((d^2 - E(d^2))^2) = E(d^4) - E(d^2)^2$$
$$= (2n)^{-2}\sum_{i,j} [E((x_i - x_{i+1})^2(x_j - x_{j+1})^2) - E((x_i - x_{i+1})^2)E((x_j - x_{j+1})^2)].$$

But if both i and $i+1$ are different from j and $j+1$, then $E((x_i - x_{i+1})^2(x_j - x_{j+1})^2) = E((x_i - x_{i+1})^2)E((x_j - x_{j+1})^2)$, and thus there are not more than $3n$ summands in the above summation that are not identically 0. These summands, however, depend only on $a(t_k)$, $M_2(t_k)$, $M_3(t_k)$ and $M_4(t_k)$, and they are therefore bounded. Thus $E((d^2 - E(d^2))^2)$ is equal to $(2n)^{-2}$ times a sum of not more than $3n$ summands which are bounded. Hence $E((d^2 - E(d^2))^2)$ tends to 0, as n tends to infinity. Now our theorem is an immediate consequence of the criterion 2.d.

THEOREM 2. *s^2 converges stochastically to $M_2 + \int_0^1 (a(t) - a)^2\,dt$.*

PROOF: We note first that $n(x_j - \bar{x}) = \sum_{h=1}^{n} (x_j - x_h)$ and that therefore $s^2 = n^{-3}\sum_{j=1}^{n}\sum_{h,k} (x_j - x_h)(x_j - x_k)$. Since $x_i - x_j = x_i - a(t_i) + a(t_i) - a(t_j) - (x_j - a(t_j))$, we find as usual that

$$E((x_j - x_h)^2) = M_2(t_j) + (a(t_j) - a(t_h))^2 + M_2(t_h),$$

and if $h \neq k$ we find that

$$E((x_j - x_h)(x_j - x_k)) = M_2(t_j) + (a(t_j) - a(t_h))(a(t_j) - a(t_k)).$$

Consequently

$$\sum_{h,k} E((x_j - x_h)(x_j - x_k)) = n^2M_2(t_j) + \sum_{h=1}^{n} M_2(t_h)$$
$$+ \sum_{h,k} (a(t_j) - a(t_h))(a(t_j) - a(t_k))$$
$$= n^2M_2(t_j) + \sum_{h=1}^{n} M_2(t_h) + \left[\sum_{h=1}^{n} (a(t_j) - a(t_h))\right]^2.$$

Consequently

$$E(s^2) = n^{-1} \sum_{j=1}^{n} M_2(t_j) + n^{-2} \sum_{h=1}^{n} M_2(t_h) + n^{-3} \sum_{j=1}^{n} \left[\sum_{h=1}^{n} (a(t_j) - a(t_h)) \right]^2.$$

As in the proof of Theorem 1 we see that the first of these sums tends to M_2 as n tends to infinity, and the second of these sums therefore tends to 0 as n tends to infinity. The last sum equals

$$n^{-3} \sum_{j,h,k} [a(t_j)^2 - a(t_j)(a(t_h) + a(t_k)) + a(t_h)a(t_k)]$$

$$= n^{-1} \sum_{j=1}^{n} a(t_j)^2 - 2n^{-2} \sum_{j,h} a(t_j)a(t_h) + n^{-2} \sum_{h,k} a(t_h)a(t_k)$$

$$= n^{-1} \sum_{j=1}^{n} a(t_j)^2 - \left[n^{-1} \sum_{j=1}^{n} a(t_j) \right]^2,$$

and this expression tends to $\int_0^1 a(t)^2\, dt - \left[\int_0^1 a(t)\, dt \right]^2$ as n tends to infinity. But

$$\int_0^1 a(t)^2\, dt - \left[\int_0^1 a(t)\, dt \right]^2 = \int_0^1 (a(t) - a)^2\, dt,$$

since $a = \int_0^1 a(t)\, dt$, and thus we have shown that $E(s^2)$ tends to $M_2 + \int_0^1 (a(t) - a)^2\, dt$ as n tends to infinity.

If j, h, k, p, q, r are integers between 1 and n, we put

$$(j, h, k; p, q, r) = E((x_j - x_h)(x_j - x_k)(x_p - x_q)(x_p - x_r)) - E((x_j - x_h)(x_j - x_k))E((x_p - x_q)(x_p - x_r)).$$

If neither j, h nor k is equal to any of the three integers p, q, r, it follows from the independence of the variables x_i that $(j, h, k; p, q, r) = 0$. Thus

$$E((s^2 - E(s^2))^2) = E(s^4) - E(s^2)^2 = n^{-6}\, \Sigma'(j, h, k; p, q, r),$$

where the summation is taken over all the values of j, h, k, p, q, r between 1 and n with the restriction that at least one of the three numbers j, h, k is equal to at least one of the three numbers p, q, r. This sum contains therefore not more than 3^3n^5 summands, and each of the summands is bounded, since they depend only on $a(t_i)$, $M_2(t_i)$, $M_3(t_i)$ and $M_4(t_i)$. Thus $E((s^2 - E(s^2))^2)$ is equal to n^{-6} times a sum of not more than 3^3n^5 summands which are bounded. Hence $E((s^2 - E(s^2))^2)$ tends to 0 as n tends to infinity. Now our theorem is an immediate consequence of the criterion 2.d.

Noting that $\int_0^1 (a(t) - a)^2\, dt$ is nothing but the variance of the function $a(t)$ (around its mean a), we obtain the following obvious consequence of Theorems 1 and 2.

COROLLARY: *$s^2 - d^2$ converges stochastically to the variance of $a(t)$.*

Remarks similar to those made in connection with the proof of the theorem of section 3 may be made now in regard to the theorems of this section.

By similar arguments it is possible to prove that the statistic $n^{-1} \sum_{i=1}^{n-1} x_i x_{i+1}$ converges stochastically to $\int_0^1 a(t)^2\,dt$.

5. The third moment. Put $d(3) = n^{-1} \sum_{j=1}^{n-2} (x_j - x_{j+1})^2 (x_{j+1} - x_{j+2})$. Then $d(3)$ is a function of the random sample $O(T_n)$.

THEOREM 1: *$d(3)$ converges stochastically to M_3.*

PROOF: It is readily seen that

$$E((x_j - x_{j+1})^2(x_{j+1} - x_{j+2})) = M_3(t_{j+1}) + (a(t_{j+1}) - a(t_{j+2}))(M_2(t_j) + (a(t_j) - (a(t_j) - a(t_{j+1}))^2 + M_2(t_{j+1})),$$

and in practically the same fashion as in the proof of Theorem 1 of section 4 one shows now that $E(d\,(3))$ tends to M_3 as n tends to infinity.

Furthermore we have

$$E((d(3) - E(d(3))^2) = E(d(3)^2) - E(d(3))^2 = n^{-2} \sum_{j,h} (j, h),$$

where

$$(j, h) = E((x_j - x_{j+1})^2(x_{j+1} - x_{j+2})(x_h - x_{h+1})^2(x_{h+1} - x_{h+2})) - E((x_j - x_{j+1})^2(x_{j+1} - x_{j+2}))E((x_h - x_{h+1})^2(x_{h+1} - x_{h+2})).$$

Clearly $(j, h) = 0$ whenever $j + 2 < h$ or $h + 2 < j$. Consequently there appear actually in the sum of all the (j, h) not more than $5n$ terms each of which is bounded by an absolute constant, since they depend only on $a(t_i)$, $M_2(t_i)$, $M_3(t_i)$, $M_4(t_i)$, $M_5(t_i)$ and $M_6(t_i)$. From this fact we infer as before that $E((d(3) - E(d(3))^2)$ tends to 0, as n tends to infinity, and our theorem is an immediate consequence of the criterion 2.d.

Remark 1. If $M_3(t)$, $M_2(t)$ and $a(t)$ are constant, it follows from the proof that

$$E(d(3)) = \frac{n-2}{n} M_3 ;$$

and thus $(n - 2)^{-1} \sum_{j=1}^{n-2} (x_j - x_{j+1})^2 (x_{j+1} - x_{j+2})$ is an unbiased estimate of M_3.

Remark 2. One might be tempted to use instead of $d(3)$ the following function:

$$n^{-1} \sum_{j=1}^{n-1} (x_j - x_{j+1})^3.$$

By an argument of a nature rather similar to the one used in the preceding proof one may show, however, that this statistic converges stochastically to 0.

Put $s(3) = n^{-1}\sum_{j=1}^{n}(x_j - \bar{x})^3$. Then $s(3)$ is a function of the random sample $O(T_n)$. Furthermore let

$$F_3 = 3\left(\int_0^1 a(t)M_2(t)\,dt - aM_2 - a\int_0^1 a^2(t)\,dt\right) + 2a^3 + \int_0^1 a^3(t)\,dt.$$

THEOREM 2. *$s(3)$ converges stochastically to $M_3 + F_3$.*

PROOF: For fixed j, let $X(j) = \sum_{h=1}^{n}(x_j - a(t_j) + a(t_h) - x_h)$ and $A(j) = \sum_{h=1}^{n}(a(t_j) - a(t_h))$. Then

$$E(s(3)) = n^{-4}\sum_{j=1}^{n}E((X(j) + A(j))^3)$$
$$= n^{-4}\sum_{j=1}^{n}[E(X(j)^3) + 3A(j)E(X(j)^2) + A(j)^3],$$

since $E(X(j))$ is easily seen to be 0. We find furthermore that

$$E(X(j)^3) = (n-1)^3M_3(t_j) + E([\sum_{h\neq j}(a(t_h) - x_h)]^3)$$
$$= ((n-1)^3 + 1)M_3(t_j) - \sum_{h=1}^{n}M_3(t_h);$$
$$E(X(j)^2) = (n-1)^2M_2(t_j) + E(\sum_{h\neq 1}(a(t_h) - x_h)^2)$$
$$= ((n-1)^2 - 1)M_2(t_j) + \sum_{h=1}^{n}M_2(t_h).$$

Consequently

$$E(s(3)) = n^{-4}\Big[((n-1)^3 - n + 1)\sum_{j=1}^{n}M_3(t_j) + 3((n-1)^2 - 1)\sum_{j=1}^{n}A(j)M_2(t_j)$$
$$+ 3\sum_{j=1}^{n}A(j)\sum_{h=1}^{n}M_2(t_h) + \sum_{j=1}^{n}A(j)^3\Big].$$

Since furthermore $\sum_{j=1}^{n}A(j) = \sum_{j,h}(a(t_j) - a(t_h)) = 0,$

$$\sum_{j=1}^{n}A(j)M_2(t_j) = n\sum_{j=1}^{n}a(t_j)M_2(t_j) - \sum_{h=1}^{n}a(t_h)\sum_{j=1}^{n}M_2(t_j)$$

and

$$\sum_{j=1}^{n}A(j)^3 = \sum_{j=1}^{n}\left[na(t_j) - \sum_{h=1}^{n}a(t_h)\right]^3$$
$$= n^3\sum_{j=1}^{n}a(t_j)^3 - 3n^2\sum_{j=1}^{n}a(t_j)^2\sum_{h=1}^{n}a(t_h) + 3n\sum_{j=1}^{n}a(t_j)\left[\sum_{h=1}^{n}a(t_h)\right]^2$$
$$- n\left[\sum_{h=1}^{n}a(t_h)\right]^3,$$

it is easily verified that $E(s(3))$ tends to $M_3 + F_3$, as n tends to infinity.

To prove that $E((s(3) - E(s(3)))^2)$ tends to 0 as n tends to infinity, one proceeds as in the proofs of the preceding theorems, namely by verifying that this expectation is n^{-8} times a sum of not more than 4^4n^7 summands which are bounded, since they depend only on $a(t_i)$ and on the $M_m(t_i)$ for $1 < m < 7$. The proof of the theorem may then be completed by applying the criterion 2.d

It is readily seen that F_3 vanishes whenever $a(t)$ is constant. But from

$$F_3 = 3\left[\int_0^1 a(t)M_2(t)\,dt - aM_2\right] + \int_0^1 (a(t) - a)^3\,dt$$

we infer that F_3 vanishes too whenever $M_2(t)$ is constant and $a(t)$ is at the same time symmetric with regard to a, and more precisely: if $M_2(t)$ is constant, a necessary and sufficient condition for the vanishing of F_3 is the vanishing of the third moment of the function $a(t)$ around its mean. Thus we see that $d(3)$ is always a consistent statistic for M_3, though $s(3)$ is not.

6. The fourth moment. The results in this section will be stated without proof. Their proofs can be constructed on exactly the same lines as the proofs in sections 4 and 5.

$$(2n)^{-1}\sum_{j=1}^{n-1}(x_j - x_{j+1})^4, \qquad n^{-1}\sum_{j=2}^{n-1}(x_{j-1} - x_j)^3(x_{j+1} - x_j)$$

and

$$n^{-1}\sum_{j=2}^{n-1}(x_{j-1} - x_j)^2(x_{j+1} - x_j)^2$$

converge stochastically to $M_4 + 3\int_0^1 M_2(t)^2\,dt$.

$(4n)^{-1}\left[\sum_{j=1}^{n-1}(x_j - x_{j+1})^2\right]^2$ *converges stochastically to* $M_4 + M_2^2$.

$(4n)^{-1}\sum_{j=2}^{n-2}(x_{j-1} - x_j)^2(x_{j+1} - x_{j+2})^2$ *converges stochastically to* $\int_0^1 M_2(t)^2\,dt$.

From these facts one easily deduces that M_4 *is the stochastic limit of*

$$n^{-1}\left[\frac{1}{2}\sum_{j=1}^{n-1}(x_j - x_{j+1})^4 - \frac{3}{4}\sum_{j=2}^{n-2}(x_{j-1} - x_j)^2(x_{j+1} - x_{j+2})^2\right],$$

and that $\int_0^1 (M_2(t) - M^2)^2\,dt$ *is the stochastic limit of*

$$(2n)^{-1}\left[\sum_{j=1}^{n-1}(x_j - x_{j+1})^4 - \frac{1}{2}\left\{\sum_{j=1}^{n-1}(x_j - x_{j+1})^2\right\}^2 - \sum_{j=2}^{n-2}(x_{j-1} - x_j)^2(x_{j+1} - x_{j+2})^2\right].$$

7. Efficiency. If $f = f(O(T_n))$ is a function of the random sample $O(T_n)$, and if f converges stochastically to a number r, then

$$\lim_{n\to\infty} nE((f - r)^2)$$

may be considered as some sort of a measure for the efficiency[11, 12] of the statistic f as an estimate of r, provided, of course, the limit exists.

THEOREM 1. *If the function $a(t)$ is of bounded variation, then*

$$\lim_{n\to\infty} nE((\bar{x} - a)^2) = M_2 .$$

PROOF: Clearly

$$nE((\bar{x} - a)^2) = n^{-1}E\left(\left[\sum_{j=1}^{n} (x_j - a)\right]^2\right)$$
$$= n^{-1}\sum_{j=1}^{n} M_2(t_j) + n^{-1}\left[\sum_{j=1}^{n} (a(t_j) - a)\right]^2.$$

Now $\sum_{j=1}^{n} (a(t_j) - a) = \sum_{j=1}^{n} a(t_j) - na = \sum_{j=1}^{n}\left[a(t_j) - n\int_{(j-1)/n}^{j/n} a(t)dt\right].$

Since $a(t)$ is a continuous function, there exists a number u_j such that

$$(j - 1)/n \leq u_j \leq j/n, \quad \text{and} \quad \int_{(j-1)/n}^{j/n} a(t)dt = n^{-1}a(u_j).$$

Thus

$$\sum_{j=1}^{n} (a(t_j) - a) = \sum_{j=1}^{n} (a(t_j) - a(u_j)).$$

But both t_j and u_j are between $(j - 1)/n$ and j/n, and $a(t)$ is of bounded variation. Hence there exists a constant A which depends on $a(t)$ only and not on n or T_n such that

$$\left[\sum_{j=1}^{n} (a(t_j) - a)\right]^2 \leq A \text{ for every choice of } T_n.$$

The contention of our theorem is a fairly immediate consequence of these facts.

This theorem and its proof may serve as an additional substantiation of the remarks appended to section 3.

Remark: If we had assumed only the continuity of $a(t)$ instead of its being of bounded variation, we could have tried to argue as follows: Since $a(t)$ is continuous, there exists to every positive number ϵ an integer $N(\epsilon)$ such that $| a(t') - a(t'') | < \epsilon$ for $| t' - t'' | < N(\epsilon)^{-1}$. Hence we would find that for $N(\epsilon) < n$ we have

$$n^{-1}\left[\sum_{j=1}^{n} (x_j - a)\right]^2 < n\epsilon^2;$$

and this inequality is certainly insufficient for proving that the left side of the inequality tends to 0 as n tends to infinity.

THEOREM 2: *If the functions $a(t)$ and $M_2(t)$ are both of bounded variation, then* $\lim_{n\to\infty} nE((d^2 - M_2)^2) = M_4$.

---

[11] Wilks [2], p. 134/135.

[12] or a measure for the asymptotic variance of the function f.

PROOF: In the course of the proof of Theorem 1 of section 4 we have shown that $E(d^2) = (2n)^{-1}(A + B - C)$, where

$$A = 2 \sum_{j=1}^{n} M_2(t_j), B = \sum_{j=1}^{n-1} (a(t_j) - a(t_{j+1}))^2, C = M_2(t_1) + M_2(t_n).$$

Since $M_2(t)$ is bounded, it is clear that $n^{-\frac{1}{2}}C$ tends to 0 as n tends to infinity. Since $a(t)$ is of bounded variation, there exists a constant B^* such that $B \leq B^*$ for every choice of T_n, and hence $n^{-\frac{1}{2}}B$ tends to 0 as n tends to infinity.[13] Furthermore we have

$$\sum_{j=1}^{n} M_2(t_j) - nM_2 \sum_{j=1}^{n} \left[M_2(t_j) - n \int_{(j-1)/n}^{j/n} M_2(t)dt \right].$$

Because of the continuity of $M_2(t)$ there exist numbers v_j such that

$$(j - 1)/n \leq v_j \leq j/n, \quad \text{and} \quad M_2(v_j) = n \int_{(j-1)/n}^{j/n} M_2(t)dt.$$

Consequently

$$\sum_{j=1}^{n} M_2(t_j) - nM_2 = \sum_{j=1}^{n} [M_2(t_j) - M_2(v_j)].$$

But $M_2(t)$ is a function of bounded variation, and thus we may infer, as in the proof of Theorem 1, that $n^{\frac{1}{2}}[(2n)^{-1}A - M_2]$ tends to 0 as n tends to infinity. Combining all the facts we see that $n^{\frac{1}{2}}[E(d^2) - M_2]$ tends to 0 as n tends to infinity, and hence we have shown that $n[E(d^2) - M_2]^2$ tends to 0, as n tends to infinity.

As in the proof of Theorem 1 of section 4 we note next that

$$E(d^4) - E(d^2)^2 = (2n)^{-2} \sum_{i,j} (i,j),$$

where $(i, j) = E((x_i - x_{i+1})^2(x_j - x_{j+1})^2) - E((x_i - x_{i+1})^2)E((x_j - x_{j+1})^2)$, and that $(i, j) = 0$, if either $i + 1 < j$ or $j + 1 < i$. Next we observe that

$$\begin{aligned}(i, j) = {} & E((x_i - a(t_i) + a(t_{i+1}) - x_{i+1})^2(x_j - a(t_j) + a(t_{j+1}) - x_{j+1})^2) \\ & - E((x_i - a(t_i) + a(t_{i+1}) - x_{i+1})^2)E((x_j - a(t_j) + a(t_{j+1}) - x_{j+1})^2) \\ & + (a(t_i) - a(t_{i+1}))(i, j)' + (a(t_j) - a(t_{j+1}))(i, j)'',\end{aligned}$$

where the expressions $(i, j)'$ and $(i, j)''$ are bounded (by a number independent of i, j, n or T).

Consequently we have

$$\begin{aligned}(i, i) = {} & M_4(t_i) + 6M_2(t_i)M_2(t_{i+1}) + M_4(t_{i+1}) - (M_2(t_i) + M_2(t_{i+1}))^2 \\ & + (a(t_i) - a(t_{i+1}))(i, i)^* \\ = {} & M_4(t_i) + M_4(t_{i+1}) + M_2(t_i)^2 + M_2(t_{i+1})^2 \\ & - 2(M_2(t_i) - M_2(t_{i+1}))^2 + (a(t_i) - a(t_{i+1}))(i, i)^*,\end{aligned}$$

where $(i, i)^* = (i, i)' + (i, i)''$ is bounded by a bound independent of i, n, T_n.

[13] A remark similar to the one made just before stating Theorem 2 may be made here and below about the indispensability of the hypothesis that $a(t)$ and $M_2(t)$ be of bounded variation.

Likewise we find that

$$\begin{aligned}(i, i+1) &= M_2(t_i)M_2(t_{i+1}) + M_2(t_i)M_2(t_{i+2}) + M_4(t_{i+1}) + M_2(t_{i+1})M_2(t_{i+2}) \\ &\quad - (M_2(t_i) + M_2(t_{i+1}))(M_2(t_{i+1}) + M_2(t_{i+2})) \\ &\quad + (a(t_i) - a(t_{i+1}))(i, i+1)' + (a(t_{i+1}) - a(t_{i+2}))(i, i+1)'' \\ &= M_4(t_{i+1}) - M_2(t_{i+1})^2 \\ &\quad + (a(t_i) - a(t_{i+1}))(i, i+1)' + (a(t_{i+1}) - a(t_{i+2}))(i, i+1)''\end{aligned}$$

Hence

$$\begin{aligned}(i, i) + 2(i, i+1) &= M_4(t_i) + 3M_4(t_{i+1}) + (M_2(t_i) \\ &\quad - M_2(t_{i+1}))(3M_2(t_{i+1}) - M_2(t_1)) + (a(t_i) \\ &\quad - a(t_{i+1}))(i, i)^+ + (a(t_{i+1}) \\ &\quad - a(t_{i+2}))(i, i+1)'',\end{aligned}$$

where $(i, i)^+ = (i, i)' + (i, i)'' + (i, i+1)'$ is bounded by a bound independent of i, n, T. Considering that

$$\sum_{i,j} (i,j) = \sum_{i=1}^{n-1} (i, i) + 2\sum_{i=1}^{n-2} (i, i+1),$$

it is now deduced from the continuity of the functions $a(t)$, $M_2(t)$ and $M_4(t)$ that $n[E(d^4) - E(d^2)^2]$ tends to M_4, as n tends to infinity. We note finally that $E((d^2 - M_2)^2) = E((d^2 - E(d^2))^2) + (E(d^2) - M_2)^2$, and the theorem is an immediate consequence of the facts we have deduced.

THEOREM 3. *If the functions $a(t)$ and $M_2(t)$ are both of bounded variation, then*

$$\begin{aligned}&\lim_{n\to\infty} nE((s^2 - M_2 - \int_0^1 (a(t) - a)^2\,dt)^2) \\ &= M_4 - \int_0^1 M_2(t)^2\,dt + 4\int_0^1 (a(t)M_3(t) - aM_3)\,dt + 4\int_0^1 M_2(t)(a(t) - a)^2\,dt.\end{aligned}$$

PROOF. Since $a(t)$ and $M_2(t)$ are of bounded variation, we show—as in the proofs of the two preceding theorems—that

$$n^{\frac{1}{2}}(n^{-1}\sum_{j=1}^{n} a(t_j) - a),\ n^{\frac{1}{2}}(n^{-1}\sum_{j=1}^{n} a(t_j)^2 - \int_0^1 a(t)^2\,dt), \quad \text{and} \quad n^{\frac{1}{2}}(n^{-1}\sum_{j=1}^{n} M_2(t_j) - M_2)$$

all tend to 0, as n tends to infinity. In the proof of Theorem 2 of section 4 we computed $E(s^2)$. Using this result we obtain:

$$\begin{aligned}&n^{\frac{1}{2}}(E(s^2) - M_2\int_0^1 (a(t) - a)^2\,dt) \\ &\quad = n^{\frac{1}{2}}(n^{-1}\sum_{j=1}^{n} M_2(t_j) - M_2) + n^{-\frac{1}{2}}n^{-1}\sum_{j=1}^{n} M_2(t_j) \\ &\qquad + n^{\frac{1}{2}}(n^{-1}\sum_{j=1}^{n} a(t_j)^2 - \int_0^1 a(t)^2\,dt) \\ &\qquad + n^{\frac{1}{2}}\left(a^2 - \left[n^{-1}\sum_{j=1}^{n} a(t_j)\right]^2\right),\end{aligned}$$

where one should remember the identity $\int_0^1 (a(t) - a)^2\,dt = \int_0^1 a(t)^2\,dt - a^2$. But

$$n^{\frac{1}{2}}\left(a^2 - \left[n^{-1}\sum_{j=1}^{n} a(t_j)\right]^2\right) = n^{\frac{1}{2}}\left(a - n^{-1}\sum_{j=1}^{n} a(t_j)\right)\left(a + n^{-1}\sum_{j=1}^{n} a(t_j)\right),$$

where the last factor on the right is bounded by a bound independent of n and T_n. Hence it follows that

$$n\left(E(s^2) - M_2 - \int_0^1 (a(t) - a)^2\,dt\right)^2 \text{ tends to 0, as } n \text{ tends to infinity.}$$

By a computation of great length and little interest one shows that

$$\begin{aligned} nE((s^2 - E(s^2))^2) = n^{-3}\Bigg[&(n-1)^2\sum_{j=1}^{n} M_4(t_j) + 4n(n-1)\sum_{j=1}^{n} M_3(t_j)a(t_j) \\ &- 4(n-1)\sum_{j=1}^{n} M_3(t_j)\sum_{h=1}^{n} a(t_h) + 2\left[\sum_{j=1}^{n} M_2(t_j)\right]^2 \\ &- (n^2 - 2n + 3)\sum_{j=1}^{n} M_2(t_j)^2 + 4n^2\sum_{j=1}^{n} M_2(t_j)a(t_j)^2 \\ &- 8n\sum_{j=1}^{n} a(t_j)\sum_{h=1}^{n} a(t_h)M_2(t_h) \\ &+ 4\left[\sum_{j=1}^{n} a(t_j)\right]^2\sum_{h=1}^{n} M_2(t_h)\Bigg]. \end{aligned}$$

It is readily seen that this expression tends to

$$\begin{aligned} M_4 + 4\int_0^1 M_3(t)a(t)\,dt - 4M_3 a - \int_0^1 M_2(t)^2\,dt + 4\int_0^1 M_2(t)a(t)^2\,dt \\ - 8a\int_0^1 a(t)M_2(t)\,dt + 4a^2 M_2, \end{aligned}$$

and now it is clear how to complete the proof of our theorem.

COROLLARY 1. *If $a(t)$ is constant and $M_2(t)$ of bounded variation, then* $\lim_{n\to\infty} nE((s^2 - M_2)^2) = M_4 - \int_0^1 M_2(t)^2\,dt.$

This is an almost immediate consequence of Theorem 3, since $a(t) = a$, if $a(t)$ is constant.

It has been shown in section 4 that d^2 is always a consistent estimate of M_2 whereas s^2 is a consistent estimate of M_2 if, and only if, $a(t)$ is constant. Theorem 1 and Corollary 1 offer a basis for comparing the efficiency of these two statistics. Since

$$0 < M_2(t)^2 < M_4(t) \text{ for every } t$$

(apart from trivial exceptions), we infer from Theorem 1 and Corollary 1 the following fact.

COROLLARY 2. *If $a(t)$ is constant and $M_2(t)$ of bounded variation, then*

$$\lim_{n\to\infty} \frac{E((s^2 - M_2)^2)}{E((d^2 - M_2)^2)} = 1 - \frac{\int_0^1 M_2(t)^2\,dt}{M_4},$$

and this expression is always positive and smaller than 1.

Thus we may say roughly that for large n the estimate s^2 of M_2 is more efficient than the estimate d^2, in case both may be used.[14] We do, however, not offer any information of the necessary size of n. Neither do we claim that for small n it might not happen that d^2 gives a good estimate and s^2 a poor one.

REFERENCES

[1] J. VON NEUMANN, R. H. KENT, H. R. BELLINSON AND B. I. HART, "The mean square successive difference," *Annals of Math. Stat.*, V. 12 (1941), pp. 153–162.

[2] S. S. WILKS, *Mathematical Statistics*, Princeton, N. J., 1943.

[14] It has been pointed out before that s^2 is a consistent estimate of M_2 if, and only if, $a(t)$ is constant, and thus the efficiency of s^2 and d^2 as estimates of M_2 may be compared only if $a(t)$ is constant.