THE EFFICIENCY OF THE MEAN MOVING RANGE

By Paul G. Hoel

University of California at Los Angeles

Summary. In studying the variation of a variable subject to erratic trend effects, it is customary to employ as a measure of variation a statistic that eliminates most of such effects. It is shown in this paper that the statistic $\mathrm{w} = \sum_1^{n-1} | x_{i+1} - x_i | \sqrt{\pi}/2(n - 1)$ is nearly as efficient as the statistic $\delta^2 = \sum_1^{n-1} (x_{i+1} - x_i)^2/(n - 1)$ that is customarily employed. The asymptotic variance of w is obtained by integration techniques; the proof of the asymptotic normality of w is based upon a theorem of S. Bernstein on the asymptotic distribution of sums of dependent variables. The method of proof is sufficiently general to prove the asymptotic normality of w, and of δ^2, for x having a distribution for which the third absolute moment exists.

1. Introduction. Let $x_1, x_2, \cdots, x_n$ denote a random sample of size n from a population with a continuous distribution function $f(x)$. If a measure of the variability of x is desired, it is customary to select the familiar statistic

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n - 1}, \tag{1}$$

or its positive square root s, as an estimate of the corresponding theoretical measure of variability.

If, however, it is known that the variable x is subject to trend effects and that $f(x)$ represents the distribution of x without such effects, then s^2 will not serve as a satisfactory measure of variability about the trend. In order to eliminate the influence of trends, it is helpful to employ statistics that capitalize on the time order relationships of the observations. There are several statistics of this type available, although most of them make no pretense of completely eliminating trend effects, even if the trend is linear.

Perhaps the best known among statistics of the desired type is the mean square successive difference,

$$\delta^2 = \frac{\sum_{i=1}^{n-1} (x_{i+1} - x_i)^2}{n - 1} \tag{2}$$

This measure of variation has been studied extensively in recent years. Among the results of these investigations is a determination [1] of the efficiency of $\delta^2/2$ as an estimate of σ^2 for a normally distributed variable when no trend exists.

A closely related measure of variation that is not so well known is the mean moving range of successive pairs of observations,

$$\text{(3)} \qquad \text{w} = \frac{\sum_{i=1}^{n-1} | x_{i+1} - x_i |}{n - 1}.$$

Although w appears [1] to have been used by ballisticians, very little seems to be known concerning the relative merits of δ^2 and w. Since w is considerably easier to calculate than δ^2, it would be preferred to δ^2 for applications in which computational advantages are important. However, one would hardly allow such advantages to dominate a choice unless δ^2 and w were about equally efficient as estimates of variation.

The purpose of this paper is to determine the efficiency of w and to study efficiency properties of generalizations of w.

2. Definition of efficiency. The definition that will be used in this paper [2] may be stated in the following manner. Let θ be a parameter, or a function of parameters, of the distribution function $f(x)$. Let T be a statistic for which there exists a number μ such that

$$t = \sqrt{n}\,(T - \theta)$$

is asymptotically normally distributed with zero mean and variance μ^2. Let T' be any other statistic for which there exists a number μ' such that

$$t' = \sqrt{n}\,(T' - \theta)$$

is asymptotically normally distributed with zero mean and variance μ'^2. Then T is said to be an efficient estimate of θ provided that $\mu \leq \mu'$ for all possible choices of T', and the efficiency of any particular T' is defined to be

$$\text{(4)} \qquad \mathcal{E}_{T'} = \left(\frac{\mu}{\mu'}\right)^2.$$

In order to determine the efficiency of a statistic, it is therefore necessary to first demonstrate its asymptotic normal distribution and then calculate its asymptotic variance. This order of procedure will be reversed in the following determination of the efficiency of w.

3. Variance of w. Let x be normally distributed with zero mean and unit variance. Then the mean of w, where w is given by (3), may be evaluated as follows:

$$\begin{aligned} E(\text{w}) &= E\,| x_2 - x_1 | \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty}\int | x_2 - x_1 | \, e^{-\frac{1}{2}(x_1^2+x_2^2)}\, dx_1\, dx_2 \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-(x_2^2/2)} \left[\int_{-\infty}^{x_2} (x_2 - x_1) e^{-(x_1^2/2)}\, dx_1 + \int_{x_2}^{\infty} (x_1 - x_2) e^{-(x_1^2/2)}\, dx_1 \right] dx_2 \end{aligned}$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-(x_2^2/2)}\left[x_2\int_{-\infty}^{x_2} e^{-(x_1^2/2)}\,dx_1 - x_2\int_{x_2}^{\infty} e^{-(x_1^2/2)}\,dx_1 + 2e^{-(x_2^2/2)}\right]dx_2$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-(x_2^2/2)}\cdot 2\left[x_2\int_{0}^{x_2} e^{-(x_1^2/2)}\,dx_1 + e^{-(x_2^2/2)}\right]dx_2 .$$

If integration by parts is performed on the first integral with

$$u = \int_0^{x_2} e^{-(x_1^2/2)}\,dx_1 \qquad \text{and} \qquad dv = x_2\, e^{-(x_2^2/2)}\,dx_2 ,$$

the uv term will vanish at both limits and $E(w)$ will reduce to

$$E(\mathrm{w}) = \frac{2}{\pi}\int_{-\infty}^{\infty} e^{-x_2^2}\,dx_2 = \frac{2}{\sqrt{\pi}} . \tag{5}$$

This result could have been obtained more easily by other methods, but some of the integrals involved will be needed later.

For the purpose of computing the second moment of w, it is convenient to separate the independent and dependent product terms of w^2. Since there are $2(n - 2)$ of the latter, $E(\mathrm{w}^2)$ may be expressed in the form

$$(n-1)^2E(\mathrm{w}^2) = (n-1)E\,|x_2 - x_1|^2 + 2(n-2)E\,|x_2 - x_1|\,|x_3 - x_2| + (n-2)(n-3)E^2\,|x_2 - x_1| .$$

But

$$E\,|x_2 - x_1|^2 = E(x_2 - x_1)^2 = E(x_2^2) + E(x_1^2) = 2.$$

Consequently, because of (5),

$$(n-1)^2E(\mathrm{w}^2) = 2(n-2)E\,|x_2 - x_1|\,|x_3 - x_2| + 2(n-1) + 4(n-2)(n-3)/\pi . \tag{6}$$

Now consider the evaluation of the product term

$$E\,|x_2 - x_1|\,|x_3 - x_2| = (2\pi)^{-\frac{3}{2}}\int\int_{-\infty}^{\infty}\int |x_2 - x_1|\,|x_3 - x_2|\,e^{-\frac{1}{2}(x_1^2+x_2^2+x_3^2)}\,dx_1\,dx_2\,dx_3 .$$

By means of the expressions that were used to give (5), this triple integral may be reduced in the following manner:

$$E\,|x_2 - x_1|\,|x_3 - x_2| = (2\pi)^{-\frac{3}{2}}\int_{-\infty}^{\infty}\int |x_3 - x_2|\,e^{-\frac{1}{2}(x_2^2+x_3^2)} \cdot 2\left[x_2\int_0^{x_2} e^{-(x_1^2/2)}\,dx_1 + e^{-(x_2^2/2)}\right]dx_3\,dx_2$$

$$= (2\pi)^{-\frac{3}{2}}\int_{-\infty}^{\infty} e^{-(x_2^2/2)}\cdot 4\left[x_2\int_0^{x_2} e^{-(x_1^2/2)}\,dx_1 + e^{-(x_2^2/2)}\right]^2 dx_2$$

$$= 4(2\pi)^{-\frac{3}{2}}\int_{-\infty}^{\infty} e^{-(x_2^2/2)}\left[x_2^2\left(\int_0^{x_2} e^{-(x_1^2/2)}\,dx_1\right)^2 + 2x_2e^{-(x_2^2/2)}\int_0^{x_2} e^{-(x_1^2/2)}\,dx_1 + e^{-x_2^2}\right]dx_2 .$$

These three integrals, without their constant factors, will be denoted by I_1, I_2, and I_3, respectively. I_1 may be evaluated by integrating by parts with

$$u = x_2 \left(\int_0^{x_2} e^{-(x_1^2/2)} \, dx_1 \right)^2 \quad \text{and} \quad dv = x_2 e^{-(x_2^2/2)} \, dx_2 .$$

The uv term will vanish at both limits; consequently

$$\begin{aligned} I_1 &= \int_{-\infty}^{\infty} e^{-(x_2^2/2)} \left[2x_2 e^{-(x_2^2/2)} \int_0^{x_2} e^{-(x_1^2/2)} \, dx_1 + \left(\int_0^{x_2} e^{-(x_1^2/2)} \, dx_1 \right)^2 \right] dx_2 \\ &= 2 \int_{-\infty}^{\infty} x_2 e^{-x_2^2} \int_0^{x_2} e^{-(x_1^2/2)} \, dx_1 \, dx_2 + \int_{-\infty}^{\infty} e^{-(x_2^2/2)} \left(\int_0^{x_2} e^{-(x_1^2/2)} \, dx_1 \right)^2 dx_2 . \end{aligned} \tag{7}$$

The first of these two integrals may be evaluated in the same manner as the first integral preceding (5). The second integral may be evaluated by making the change of variable

$$u = \int_0^{x_2} e^{-(x_1^2/2)} \, dx_1 .$$

As a result of such manipulations,

$$I_1 = \frac{\sqrt{6\pi}}{3} + \frac{\pi\sqrt{2\pi}}{6} .$$

It will be observed that I_2 is the same as the first integral of (7) and that I_3 is available in tables; hence

$$\begin{aligned} &E \mid x_2 - x_1 \mid \mid x_3 - x_2 \mid \\ &\quad = 4(2\pi)^{-\frac{3}{2}} \left[\frac{\sqrt{6\pi}}{3} + \frac{\pi\sqrt{2\pi}}{6} + \frac{\sqrt{6\pi}}{3} + \frac{\sqrt{6\pi}}{3} \right] = \tfrac{1}{3} + \frac{2\sqrt{3}}{\pi} . \end{aligned} \tag{8}$$

If (8) is substituted in (6), $E(\mathrm{w}^2)$ will reduce to

$$E(\mathrm{w}^2) = \frac{2(n-2)}{(n-1)^2} \left[\tfrac{1}{3} + \frac{2\sqrt{3}}{\pi} \right] + \frac{2}{n-1} + \frac{4(n-2)(n-3)}{\pi(n-1)^2} . \tag{9}$$

Since $\sigma_{\mathrm{w}}^2 = E(\mathrm{w}^2) - E^2(\mathrm{w})$, (9) and (5) will yield the following desired variance of w,

$$\sigma_{\mathrm{w}}^2 = \frac{2}{(n-1)^2} \left[\left(\tfrac{4}{3} + \frac{2\sqrt{3} - 6}{\pi} \right) n + \left(\frac{10 - 4\sqrt{3}}{\pi} - \tfrac{5}{3} \right) \right] . \tag{10}$$

4. Efficiency of w. Now let x be normally distributed with mean m and variance σ^2. Then the mean of w as given by (5) will be multiplied by σ and the variance of w as given by (10) will be multiplied by σ^2; consequently $z = \mathrm{w}\sqrt{\pi}/2$ will serve as an unbiased estimate of σ. In the next section it will be shown that

$$t' = \sqrt{n}(z - \sigma)$$

possesses an asymptotic normal distribution. From (10) and section 2, it therefore follows that the asymptotic variance, μ'^2, that is needed to determine the efficiency of z is given by

$$\mu'^2 = \frac{\pi}{4} 2\left(\frac{4}{3} + \frac{2\sqrt{3} - 6}{\pi}\right) = \frac{2\pi}{3} + \sqrt{3} - 3.$$

Now it is known that for x normally distributed s, as defined by (1), is an efficient estimate of σ with $\mu^2 = \frac{1}{2}$; consequently, because of (4), the efficiency of z as an estimate of σ is given by

$$\mathcal{E}_z = \frac{1}{2\left(\frac{2\pi}{3} + \sqrt{3} - 3\right)} \doteq .605. \tag{11}$$

In [1] it was shown that for x normally distributed $\delta^2/2$ was an unbiased estimate of σ^2 and, assuming the normality of its asymptotic distribution, that the efficiency of $\delta^2/2$ as an estimate of σ^2 was 2/3. Thus, $z = \text{w}\sqrt{\pi}/2$ possesses very nearly the same efficiency as a measure of variation of a normal variable as $\delta^2/2$ does.

5. Asymptotic distribution of mean moving ranges. Although the efficiency obtained in the preceding section requires for its validity merely a demonstration that for x normally distributed w possesses an asymptotic normal distribution, it will be shown in this section that general mean moving ranges of a continuous variable x possess asymptotic normal distributions provided only that x possesses a third absolute moment.

Let r_i denote the range of the observations from x_i to x_{i+k-1}. Then the variable

$$W = \frac{r_1 + r_2 + \cdots + r_{n-k+1}}{n - k + 1} \tag{12}$$

will represent a generalized mean moving range, of which w will be a special case when $k = 2$.

A proof of the asymptotic property of W can be constructed as an application of a general theorem of S. Bernstein [3]. Since his theorem is long and involves much explanation of notation, a simplified version of it that is sufficient to cover this application, and indeed many similar applications, will be given.

Let $y_1, y_2, \cdots, y_m$ denote m variables for which the third absolute moments are bounded and let

$$S_m = y_1 + y_2 + \cdots + y_m.$$

Then Bernstein's theorem implies that if there exist constants c_1, c_2, c_3, and c_4 such that

$$c_1 m < \sigma^2_{S_m} < c_2 m, \tag{a}$$

and

(b) y_i and y_{i+g} are independently distributed for

$$g > c_3 m^{c_4}, \; c_4 < \tfrac{1}{2},$$

then

$$\frac{S_m - E(S_m)}{\sigma_{S_m}}$$

possesses an asymptotic normal distribution with zero mean and unit variance.

Consider the application of this theorem to $R = (n - k + 1)W$. The variance of R may be expressed in compact form by means of the techniques of section 3. Since r_i is the range of k consecutive observations, it is clear that

$$E(r_i r_{i+g}) = E^2(r_1)$$

if $g \geq k$. Furthermore, for subscripts for which it is defined, $E(r_i r_{i+g})$ will be independent of i. These two properties may be used to collect terms in the expansion of $E(R^2)$ to give

$$E(R^2) = (n - k + 1)E(r_1^2) + 2\sum_{i=0}^{k-2}(n - k - i)E(r_1 r_{2+i}) + (n - 2k + 1)(n - 2k + 2)E^2(r_1).$$

Consequently,

$$\sigma_R^2 = (n - k + 1)E(r_1^2) + 2\sum_{i=0}^{k-2}(n - k - i)E(r_1 r_{2+i}) + [n(1 - 2k) + (k - 1)(3k - 1)]E^2(r_1). \tag{13}$$

From the definition of the correlation coefficient and the fact that a correlation coefficient cannot exceed one, it follows that

$$\begin{aligned} E(r_1 r_{2+i}) &\leq E(r_1)E(r_{2+i}) + \sigma_{r_1}\sigma_{r_{2+i}} \\ &\leq E^2(r_1) + \sigma_{r_1}^2 . \end{aligned}$$

If this inequality is applied to (13),

$$\begin{aligned} \sigma_R^2 &\leq (n - k + 1)E(r_1^2) + (k - 1)(2n - 3k + 2)[E^2(r_1) + \sigma_{r_1}^2] + [n(1 - 2k) + (k - 1)(3k - 1)]E^2(r_1) \\ &\leq (n - k + 1)[E(r_1^2) - E^2(r_1)] + (k - 1)(2n - 3k + 2)\sigma_{r_1}^2 \\ &\leq [n(2k - 1) - (k - 1)(3k - 1)]\sigma_{r_1}^2 \\ &\leq 2k\sigma_{r_1}^2(n - k + 1). \end{aligned}$$

Thus, for a fixed k the right inequality in (a) of Bernstein's modified theorem is satisfied.

For the purpose of demonstrating that the left inequality in (a) is also satisfied, consider the following application of Schwarz's inequality. Let

$$G(x_p, \cdots, x_k) = \int \cdots \int r_p f(x_{k+1}) \cdots f(x_{k+p-1})\, dx_{k+1} \cdots dx_{k+p-1}, \tag{14}$$

where $f(x)$ denotes the distribution function of the variable x and the range of integration in this and subsequent integrals is from $-\infty$ to ∞. Since r_p and f are continuous non-negative functions, this integral is a positive function of the indicated variables. Then, denoting $G(x_p, \cdots, x_k)$ by G, it follows from Schwarz's inequality that

$$\begin{aligned} I &= \left[\int \cdots \int r_1 f(x_1) \cdots f(x_k)\, dx_1 \cdots dx_k\right]^2 \\ &= \left[\int \cdots \int \{r_1 f(x_1) \cdots f(x_k) G\}^{\frac{1}{2}} \{r_1 f(x_1) \cdots f(x_k) G^{-1}\}^{\frac{1}{2}}\, dx_1 \cdots dx_k\right]^2 \\ &\leq \int \cdots \int r_1 f(x_1) \cdots f(x_k) G\, dx_1 \cdots dx_k \int \cdots \int r_1 f(x_1) \cdots f(x_k) G^{-1} \\ &\qquad dx_1 \cdots dx_k . \end{aligned} \tag{15}$$

The two integrals of this inequality will be denoted by I_α and I_β, respectively. If the value of G given by (14) is substituted in I_α, it will be observed that

$$I_\alpha = \int \cdots \int r_1 r_p f(x_1) \cdots f(x_{k+p-1})\, dx_1 \cdots dx_{k+p-1} . \tag{16}$$

Now I_β may be written in the form

$$I_\beta = \int \cdots \int f(x_p) \cdots f(x_k) G^{-1}\left[\int \cdots \int r_1 f(x_1) \cdots f(x_{p-1})\, dx_1 \cdots dx_{p-1}\right] dx_p \cdots dx_k .$$

Since the x_i possess the same distribution function and r_1 is the range of the variables from x_1 to x_k, the integral in brackets is equivalent to the integral defining G in (14); hence

$$I_\beta = \int \cdots \int f(x_p) \cdots f(x_k) G^{-1} G\, dx_p \cdots dx_k = 1. \tag{17}$$

If (16) and (17) are applied to inequality (15), they will yield the inequality

$$\left[\int \cdots \int r_1 f(x_1) \cdots f(x_k)\, dx_1 \cdots dx_k\right]^2 \leq \int \cdots \int r_1 r_p f(x_1) \cdots f(x_{k+p-1})\, dx_1 \cdots dx_{k+p-1} .$$

In statistical language, this inequality states that

$$E^2(r_1) \leq E(r_1 r_p),$$

or, what is equivalent, that

$$E^2(r_1) \leq E(r_i r_j). \tag{18}$$

If (18) is applied to (13),

$$\sigma_R^2 \geq (n - k + 1)E(r_1^2) + (k - 1)(2n - 3k + 2)E^2(r_1) + [n(1 - 2k) + (k - 1)(3k - 1)]E^2(r_1)$$

$$\geq (n - k + 1)[E(r_1^2) - E^2(r_1)]$$

$$\geq \sigma_{r_1}^2(n - k + 1).$$

Thus, for a fixed k the left inequality in (a) of the theorem is also satisfied, and it merely remains to be shown that condition (b) is satisfied.

For k fixed, r_i and r_{i+g} will be independently distributed provided that $g \geq k$. But if $c_3 > k$, then $c_3(n - k + 1)^{c_4} > k$ for $0 < c_4 < \frac{1}{2}$ because $n - k + 1 > 1$; consequently r_i and r_{i+g} will be independently distributed for $g > c_3(n - k + 1)^{c_4}$, where $0 < c_4 < \frac{1}{2}$. Thus, conditions (a) and (b) are both satisfied by R. Since $R = (n - k + 1)W$, it therefore follows that

$$\frac{W - E(W)}{\sigma_W} \tag{19}$$

possesses an asymptotic normal distribution with zero mean and unit variance provided only that x possesses a continuous distribution function for which the third absolute moment exists. The existence of the third absolute moment for x insures the existence of the same moment for r_i.

If $k = 2$, W reduces to w, and therefore the validity of (11) is assured.

6. Other asymptotic distributions. The only property of the range employed in the proof of the preceding section was its positive nature; consequently the proof is applicable to moving means of other dependent statistics that are positive and possess third absolute moments.

For example, the preceding proof can be applied to δ^2 to show that δ^2 possesses an asymptotic normal distribution provided only that the sixth moment of x exists. In the study [1] of the efficiency of δ^2 for x normally distributed, no proof was given of its asymptotic property. The preceding proof could be used in studying the efficiency of δ^2, or obvious generalizations of it, as measure of variation for non-normal populations. The normality of the asymptotic distribution of the serial correlation coefficient could also be verified by means of this proof.

REFERENCES

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

[2] S. S. WILKS, "*Mathematical Statistics*," Princeton University Press, pp. 134–135.

[3] S. BERNSTEIN, "Sur l'extension du théorème limite du calcul des probabilités aux sommes de quantités dépendantes," *Math. Annalen*, Vol. 97 (1926–27), p. 14.