THE MODIFIED MEAN SQUARE SUCCESSIVE DIFFERENCE AND RELATED STATISTICS[1]

By Seymour Geisser

University of North Carolina

1. Introduction. In estimating the variance of a normal population one uses the statistic $s^2 = (n - 1)^{-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$ because of its optimum properties. In certain cases where there is an indeterminable trend in the data, it has been thought useful to estimate the variance by another statistic, namely the mean square successive difference, the mean of the squared first differences, studied by J. von Neumann *et al.* [5], which eliminates a good deal of the trend and under some conditions is less biased than s^2. An explicit form of the exact distribution of this statistic seems, at least for the present, too difficult to obtain. However, by applying a device analogous to one used by Durbin and Watson [1], that is, by dropping from the mean square successive difference the middle term for an even number of observations and the two middle terms for the odd case, we find that the quadratic form has double roots, thus enabling us to obtain exact distributions in terms of elementary functions. In addition we define analogues of the Student t and the Fisher F using similarly modified statistics and derive their exact distributions when the observations are independent.

The results of this paper are mainly the exact distributions of these statistics and were given at the April 1955 meetings of the Institute of Mathematical Statistics. A short while after, these same results, independently derived, were published by A. R. Kamat [3]. Since Kamat has already published the exact distributions of these statistics and the motivation for them, it would be inappropriate to rederive them here; hence we shall only state the results and give that material that Kamat had not considered in his paper.

2. The modified mean square successive difference. Let x_i be $N(0, \sigma^2)$ and let $x_1, \cdots, x_{2m}$ be independent. We define the modified mean square successive difference to be

$$\delta_0^2 = 4^{-1}(m - 1)^{-1} \sum_{\substack{i=1 \\ i \neq m}}^{2m-1} (x_{i+1} - x_i)^2. \tag{2.1}$$

The exact density of δ_0^2 is

$$\begin{aligned} \mathcal{P}_m(\delta_0^2) = \frac{4^{m-2}(m-1)}{m\sigma^2} \sum_{k=1}^{m-1} (-1)^{k+1} \sin^2 \frac{k\pi}{m} \\ \cdot \cos^{2m-6} \frac{k\pi}{2m} \exp\left[-\frac{(m-1)\sec^2 \frac{k\pi}{2m}}{2\sigma^2} \right] \end{aligned} \tag{2.2}$$

Received June 28, 1955.

[1] Work under contract with the Office of Naval Research NR 042 031, for investigations in statistics and probability at Chapel Hill. Reproduction in whole or in part is permitted for any purpose of the United States government.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Mathematical Statistics.* www.jstor.org

and the cumulative distribution function is

$$P_m(\delta_0^2) = 1 - 2^{2m-3}m^{-1} \sum_{k=1}^{m-1} (-1)^{k+1} \sin^2 \frac{k\pi}{m} \cdot \cos^{2m-4} \frac{k\pi}{2m} \exp\left[- \frac{(m-1)\sec^2 \frac{k\pi}{2m}}{2\sigma^2} \right]. \tag{2.3}$$

We shall also show that this statistic is asymptotically normal by showing that the mean square successive difference δ^2 is asymptotically normal using the central limit theorem for dependent random variables of Hoeffding and Robbins [2].

Let

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

and let

$$w_i = (x_{i+1} - x_i)^2 - 2\sigma^2.$$

Now $\mathcal{E}w_i = 0$ and

$$\mathcal{E}w_i w_{i+j} = \begin{cases} 0 & \text{if } j > 1, \\ 2\sigma^4 & \text{if } j = 1, \\ 8\sigma^4 & \text{if } j = 0. \end{cases} \tag{2.4}$$

The set $w_1, w_2, \cdots, w_{n-1}$ is a 1-dependent sequence; i.e., the set $(w_1, w_2, \cdots, w_r)$ is independent of the set $(w_{r+2}, \cdots, w_{n-1})$ for $r = 1, 2, \cdots, n-2$. Let

$$P_i = \mathcal{E}w_{i+1}^2 + 2\mathcal{E}w_{i+1}w_i \qquad \text{for } i = 1, 2. \tag{2.5}$$

Therefore $P_i = 12\sigma^4$ for all i. Since all the other conditions of the Hoeffding-Robbins theorem are satisfied, we have for every real a and b,

$$\lim_{n\to\infty} \Pr[\sigma^2 + 3^{1/2}a\sigma^2(n-1)^{-1} < \delta^2 < \sigma^2 + 3^{1/2}b\sigma^2(n-1)^{-1}] = F(b) - F(a),$$

where $F(x)$ is the cumulative normal density function,

$$(2\pi)^{-\frac{1}{2}} \int_{-\infty}^{x} e^{-\frac{1}{2}t^2}\, dt.$$

Since δ_0^2 is in reality the sum of two δ^2's, it is also asymptotically normal.

3. Moments. We shall now evaluate the moments of δ_0^2 by integrating (2.2). Hence

$$\mathcal{E}\delta_0^{2r} = \sigma^{2r}2^{2m-3+r}r!m^{-1}(m-1)^{-r}S,$$

where

$$S = \sum_{k=1}^{m-1} (-1)^{k-1} \cos^{2(m+r-2)} \frac{k\pi}{2m} \sin^2 \frac{k\pi}{m}$$

$$= 4\sum_{k=1}^{m-1} (-1)^{k-1} \cos^{2(m+r-1)} \frac{k\pi}{2m} - 4\sum_{k=1}^{m-1} (-1)^{k-1} \cos^{2(m+r)} \frac{k\pi}{2m}$$

$$= 4(S_1 - S_2); \tag{3.1}$$

$$S_1 = \sum_{k=1}^{m-1} (-1)^{k-1} \cos^{2(m+r-1)} \frac{k\pi}{2m};$$

$$S_2 = \sum_{k=1}^{m-1} (-1)^{k-1} \cos^{2(m+r)} \frac{k\pi}{2m}.$$

Now according to Schwatt ([4], p. 222),

$$S_1 = 4^{1-m-r} \sum_{\alpha=1}^{m+r-1} \binom{2m-2r-2}{m+r-1-\alpha} \left[1 + \frac{(-1)^{m-2} \cos \frac{(2m-1)\alpha\pi}{2m}}{\cos \frac{\alpha\pi}{2m}} \right] + (1-(-1)^{m-1})2^{1-2m-2r} \binom{2m+2r-2}{m+r-1}; \tag{3.2}$$

similarly, S_2 is the same as S_1 with $m + r$ in place of $m + r - 1$.

Now for m even,

$$S_1 = 4^{1-m-r} \left\{ \sum_{\alpha=1}^{m+r-1} \binom{2m+2r-2}{m+r-1-\alpha} \left[1 + \frac{\cos (2m-1) \frac{\alpha\pi}{2m}}{\cos \frac{\alpha\pi}{2m}} \right] + \binom{2m+2r-2}{m+r-1} \right\}$$

$$= 4^{1-m-r} \left\{ \sum_{\alpha=1}^{m+r-1} \binom{2m+2r-2}{m+r-1-\alpha} + \sum_{\alpha=0}^{m+r-1} \binom{2m+2r-2}{m+r-1-\alpha} \cdot \left[\frac{\cos (2m-1) \frac{\alpha\pi}{2m}}{\cos \frac{\alpha\pi}{2m}} \right] \right\}$$

$$= 4^{1-m-r}(S_{11} + S_{12}),$$

where S_{11} is equal to the first sum above and S_{12} is equal to the second sum above. Now

$$S_{11} = \sum_{\alpha=1}^{m+r-1} \binom{2m+2r-2}{m+r-1-\alpha} = \frac{1}{2}\left[2^{2m+2r-2} - \binom{2m+2r-2}{m+r-1}\right].$$

Further

$$\frac{\cos (2m-1)\dfrac{\alpha\pi}{2m}}{\cos \dfrac{\alpha\pi}{2m}} = (-1)^{\alpha}$$

for all integral values of α except $\alpha = (2\gamma - 1)m, \gamma = 1, 2, \cdots$. However

$$\lim_{x\to(2\gamma-1)m} \frac{\cos (2m-1)\dfrac{x\pi}{2m}}{\cos \dfrac{x\pi}{2m}} = 1 - 2m.$$

We shall evaluate S_{12} for $r \leqq 3m - 1$

$$S_{12} = \sum_{\alpha=0}^{m+r-1} \binom{2m+2r-2}{m+r-1-\alpha} (-1)^{\alpha} - 2m\binom{2m+2r-2}{r+1}$$

$$= \frac{1}{2}\binom{2m+2r-2}{m+r-1} - 2m\binom{2m+2r-2}{r-1}.$$

Therefore

$$S_1 = 4^{1-m-r}\left\{2^{2m+2r-3} - 2m\binom{2m+2r-2}{r-1}\right\}$$

$$= \frac{1}{2} - 2m4^{1-m-r}\binom{2m+2r-2}{r-1}$$

and similarly

$$S_2 = \frac{1}{2} - 2m4^{-(m+r)}\binom{2m+2r}{r}.$$

Therefore

$$S = m4^{2-(m+r)}(2m^2 - m - r)\cdot\frac{(2m+2r-2)!}{r!(2m+r)!}.$$

Substituting in (3.1) we get

$$\mathcal{E}\delta_0^{2r} = \sigma^{2r}(m-1)^{-r}2^{1-r}(2m^2 - m - r)\cdot\frac{(2m+2r-2)!}{(2m+r)!} \tag{3.3}$$

for $r \leqq 3m - 1$.

Similarly for m odd the same type derivation is carried out with the same result (3.3).

4. Variance Ratio Analogue. Let us now consider two independent random sample of sizes $2m_1$ and $2m_2$ whose values are $x_i (i = 1, 2, \cdots, 2m_1)$ and $y_j (j = 1, 2, \cdots, 2m_2)$. These provide estimates δ_{01}^2 and δ_{02}^2 of the variances of the population. We wish to consider whether the samples may be regarded as drawn from the same normal population of variance σ^2. If we consider the ratio $\psi = \delta_{01}^2/\delta_{02}^2$, we have an analogue of Fisher's F.

Since δ_{01}^2 is independent of δ_{02}^2 the probability is given by

$$h(\psi) = \int_0^\infty \delta_{02}^2 p_{m_1}(\psi\delta_{02}^2)p_{m_2}(\delta_{02}^2)\, d(\delta_{02}^2). \tag{4.1}$$

Using this *quotient convolution formula* we get the density of ψ to be

$$\begin{aligned} h(\psi) = 4^{m_1+m_2-3}(m_1 - 1)(m_2 - 1)m_1^{-1}m_2^{-1} \sum_{k=1}^{m_1-1} \sum_{t=1}^{m_2-1} (-1)^{t+k} \cos^{2m_1-6} \frac{k\pi}{2m_1} \\ \cdot \cos^{2m_2-6} \frac{t\pi}{2m_2} \sin^2 \frac{k\pi}{m_1} \sin^2 \frac{t\pi}{m_2} \left[\psi(m_2 - 1) \sec^2 \frac{t\pi}{2m_2} \right. \\ \left. + (m_1 - 1) \sec^2 \frac{k\pi}{2m_2} \right]^{-2}. \end{aligned} \tag{4.2}$$

5. An analogue of the Student *t*. We will now give the distribution of

$$\xi = 2m(\bar{x} - \mu) \,/\, \delta_0^2, \tag{5.1}$$

where $\mathcal{E}x_i = \mu$ and $(2m)^{-1} \sum_{i=1}^{2m} x_i = \bar{x}$.

Since δ_0 is invariant under a translation, it is independent of the numerator and we may again apply the quotient convolution formula. Hence the density of ξ is

$$\begin{aligned} v(\xi) = 4^{m-2}(m - 1)m^{-1} \sum_{k=1}^{m-1} (-1)^{k+1} \cos^{2m-6} \frac{k\pi}{2m} \sin^2 \frac{k\pi}{m} \\ \cdot \left[\xi^2 + (m - 1) \sec^2 \frac{k\pi}{2m} \right]^{-3/2}, \end{aligned}$$

where $-\infty < \xi < \infty$.

It can be shown that $\xi \to N(0, 1)$ by considering $\xi^2 = 2m(\bar{x} - \mu)^2 / \delta_0^2$. If we divide the numerator and the denominator of ξ^2 by σ^2, the resultant denominator converges in probability to 1 as m increases and the numerator is a chi-square variable with 1 degree of freedom for all m. Hence ξ^2 converges to a χ_1^2 variable, and since ξ is symmetric, it tends to a $N(0, 1)$ variable.

6. Tables. The application of these statistics to control charts has been discussed by Kamat [3]. We shall give a table of the upper and lower .025 points of δ_0^2 and the two tailed .05 points of ξ which is symmetric. In the table, $n = 2m$ is the number of observations. The values for $n \leqq 20$ have been computed directly from the cumulative distribution functions. For $n > 20$ we find the values by an approximate procedure. We let

$$P_m = P_\infty + a_1m^{-1} + a_2m^{-2} + \cdots, \tag{6.1}$$

$$P_m \sim P_\infty + a_1m^{-1} + a_2m^{-2} + a_3m^{-3}. \tag{6.2}$$

Since we have exact values for P_m for $n \leqq 20$, we choose three of these values and get three simultaneous equations in a_1, a_2, and a_3. We then use the values for a_1, a_2, and a_3 in (6.2) to extend Table I.

TABLE I

	δ_0^2/σ^2	δ_0^2/σ^2	ξ
n	upper .025	lower .025	.05 level
4	3.689	.026	4.30
6	3.071	.106	2.82
8	2.694	.172	2.55
10	2.458	.225	2.39
12	2.294	.269	2.34
14	2.172	.306	2.31
16	2.078	.338	2.29
18	2.001	.366	2.27
20	1.938	.391	2.25
22*	1.866	.405	2.22
24*	1.810	.423	2.20
26*	1.761	.442	2.18
28*	1.718	.463	2.16
30*	1.677	.483	2.14
32*	1.645	.502	2.13
34*	1.612	.520	2.12
36*	1.583	.538	2.11
38*	1.558	.554	2.10
40*	1.534	.570	2.09
42*	1.511	.585	2.09
44*	1.491	.599	2.08
46*	1.472	.612	2.08
48*	1.454	.624	2.07
50*	1.437	.637	2.06

* For these n's the value calculated in the table is approximate.

7. Acknowledgment. I am extremely grateful to Professor Harold Hotelling for guidance and help in this research.

REFERENCES

[1] J. Durbin and G. S. Watson, "Exact tests of serial correlation using non-circular statistics," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 446–451.

[2] W. Hoeffding and H. Robbins, "The central limit theorem for dependent random variables," *Duke Math. J.*, Vol. 15 (1948), pp. 773–780.

[3] A. R. Kamat, "Modified mean square successive difference with an exact distribution," *Sankhya*, Vol. 15, Pt. 3 (1955), pp. 295–302.

[4] I. J. Schwatt, *An Introduction to the Operations with Series*, University of Pennsylvania Press, Philadelphia (1924).

[5] J. von Neumann, R. H. Kent, H. R. Bellinson, and B. I. Hart, "The mean square successive difference," *Ann. Math. Stat.*, Vol. 12 (1941), pp. 153–162.