DISTRIBUTION OF ROUND-OFF ERRORS FOR RUNNING AVERAGES

R. E. GREENWOOD AND A. M. GLEASON

1. Statement of the problem. Let G_1, G_2, $\cdots$ be scores (positive integers) obtained in a sequence of plays in a certain game. For purposes of handicapping matches it is desired to use running averages, and on the hypothesis that the score of the last play is more significant than any prior score, the following formula is used for computing the running averages $\{S_n\}$:

$$S_{n+1} = \frac{(k-1)\,S_n + G_{n+1}}{k} \tag{1.1}$$

where k is a positive integer. Certain modifications in (1.1) may be necessary when $n < k$.

The running averages defined by (1.1) are not necessarily integers. It is therefore convenient to define a rounded running average (which will be integral) by the relation

$$T_{n+1} = \frac{(k-1)\,T_n + G_{n+1} + D}{k}. \tag{1.2}$$

It is convenient to use three set of values for D in the foregoing relation.

Case A. For k odd, $D = A \in \left\{\frac{-k+1}{2}, \frac{-k+3}{2}, \cdots, \frac{k-1}{2}\right\}$.

Case B. For k even, $D = B \in \left\{\frac{-k}{2} + 1, \frac{-k}{2} + 2, \cdots, \frac{k}{2}\right\}$.

Case C. For k even, $D = C \in \left\{\frac{-k}{2}, \frac{-k}{2} + 1, \cdots, \frac{k}{2}\right\}$.

For each $n \geq k$ define the error E_n by the relation

Received July 21, 1952.
Pacific J. Math. 3 (1953), 605-611

$$E_n = T_n - S_n. \tag{1.3}$$

(For $n < k$, the error would depend on the modifications made in relation (1.1).) For $n \geq k$, then,

$$E_{n+1} = T_{n+1} - S_{n+1} = \frac{(k-1)\, E_n + D}{k}. \tag{1.4}$$

For Case A, if at some stage $|E_n| \leq (k-1)/2$, then

$$|E_{n+1}| \leq \frac{(k-1)\,(k-1)/2 + (k-1)/2}{k} = \frac{k-1}{2}. \tag{1.5}$$

For cases B and C, if at some stage $|E_n| \leq k/2$, then by a similar procedure one obtains

$$|E_{n+1}| \leq k/2. \tag{1.6}$$

Thus the errors introduced by the rounding off process are bounded if $|E_k| \leq (k-1)/2$ or $k/2$ for the odd and even values of k respectively.

It is assumed that the scores $\{G_i\}$ are such that equal probability values are realistic. In case C, where there will sometimes be a choice for round-off, one might choose to round-off to the even integer. Thus, one would sometimes add $k/2$ and sometimes subtract $k/2$, corresponding to the two end-values with probabilities $1/(2k)$, while the intermediate values would have probabilities $1/k$. It is desired to find a limiting distribution for the error E_n; in this paper such limiting distributions are found for a few special cases.

Allowing one's intuition free rein, one sees that limiting distributions for the error E_n exist in all three cases. If such distributions exist, then relation (1.4) may be used to determine means and variances, if any. Thus

$$k\mu(E_{n+1}) = (k-1)\,\mu(E_n) + \mu(D), \tag{1.7}$$

$$k^2 \operatorname{Var}(E_{n+1}) = (k-1)^2 \operatorname{Var}(E_n) + \operatorname{Var}(D). \tag{1.8}$$

It is easy to verify that

$$\begin{aligned} \mu(A) &= 0, & \operatorname{Var}(A) &= (k^2-1)/12, \\ \mu(B) &= 1/2, & \operatorname{Var}(B) &= (k^2-1)/12, \\ \mu(C) &= 0, & \operatorname{Var}(C) &= (k^2+2)/12. \end{aligned}$$

Then for the limiting distributions E_A, E_B, E_C for the three cases one gets

$$\begin{aligned} \mu(E_A) &= 0, & \operatorname{Var}(E_A) &= (k^2-1)/12(2k-1), \\ \mu(E_B) &= 1/2, & \operatorname{Var}(E_B) &= (k^2-1)/12(2k-1), \\ \mu(E_C) &= 0, & \operatorname{Var}(E_C) &= (k^2+2)/12(2k-1), \end{aligned} \tag{1.10}$$

2. Distribution of the round-off error for $k = 2$, Case B. For the special value $k = 2$ and for Case B, one may take $E_1 \equiv 0$. Let $F_n(x)$ be the cumulative distribution for E_n, and let $\{f_{i,n}\}$ be the jumps in $F_n(x)$ at the points of discontinuity. One readily obtains the functions

$$F_2(x) = \begin{cases} 0, & x < 0, \\ 1/2, & 0 \le x < 1/2 \\ 1, & 1/2 \le x. \end{cases} \tag{2.1}$$

$$\{f_{i,2}\} = \{1/2 \text{ at } 0,\ 1/2 \text{ at } 1/2\}.$$

$$F_3(x) = \begin{cases} 0, \ x < 0, \\ j/4, \ (j-1)/4 \le x < j/4 & (j = 1, 2, 3), \\ 1, \ 3/4 \le x. \end{cases} \tag{2.2}$$

$$\{f_{i,3}\} = \{1/4 \text{ at } 0,\ 1/4 \text{ at } 1/4,\ 1/4 \text{ at } 1/2,\ 1/4 \text{ at } 3/4\}.$$

By induction one gets

$$F_{n+1}(x) = \begin{cases} 0, \ x < 0 \\ j/2^n, \ (j-1)/2^n \le x < j/2^n & (j = 1, \cdots, 2^n - 1), \\ 1, \ (2^n-1)/2^n \le x. \end{cases} \tag{2.3}$$

$$\{f_{i,n+1}\} = \{\text{jumps of } 1/2^n \text{ at points } j/2^n,\ j = 0, 1, \cdots, 2^n - 1\}.$$

In this simple example, heuristic considerations suggest that there is a limiting cumulative distribution function

$$F(x) = \begin{cases} 0, \ x < 0, \\ x, \ 0 \le x < 1, \\ 1, \ 1 \le x, \end{cases} \tag{2.4}$$

and its associated distribution function

$$f(x) = \begin{cases} 1, \ 0 < x < 1, \\ 0 \ \text{ elsewhere}. \end{cases} \tag{2.5}$$

In order to deal with continuous functions insofar as possible, it is convenient to take Fourier transforms of the jumps $\{f_{j,n}\}$. The finite Fourier transform may be defined by relations

$$(2.6)\qquad \phi_n(u) = \int_{-\infty}^{\infty} e^{iut}\, d\, F_n(t)$$

$$= \sum_{\text{all } j} f_{j,n} \exp(iuj).$$

Thus we get

$$(2.7)\qquad \phi_2(u) = 1/2 + (1/2)\exp(iu/2)$$

$$= \exp(iu/4)\cos(u/4) = \frac{1}{2}\exp(iu/4)\frac{\sin(u/2)}{\sin(u/4)},$$

$$(2.8)\qquad \phi_{n+1}(u) = \frac{1}{2^n}\sum_{j=0}^{j=2^n-1}\exp\left(\frac{iju}{2^n}\right)$$

$$= \frac{1}{2^n}\,\frac{1-\exp(iu)}{1-\exp(iu/2^n)}$$

$$= \frac{1}{2^n}\,\frac{\sin(u/2)}{\sin(u/2^{n+1})}\exp\left(iu\frac{2^n-1}{2^{n+1}}\right).$$

The sequence of transforms $\{\phi_n\}$ has a limit $\phi(u)$,

$$(2.9)\qquad \phi(u) = \frac{\sin u/2}{u/2}\exp(iu/2).$$

In order to transform back, it is convenient to use another definition of the Fourier transform,

$$(2.10)\qquad \phi(u) = \int_{-\infty}^{\infty} e^{iut} f(t)\, dt.$$

Then, whenever $f(x)$ is of class $L_2(-\infty, \infty)$ and of bounded variation in the neighborhood of t [**1**, p. 83, Theorem 58],

$$(2.11)\qquad \frac{1}{2}[f(t+0)+f(t-0)] = \lim_{\lambda\to\infty}\frac{1}{2\pi}\int_{-\lambda}^{\lambda} e^{-iut}\phi(u)\, du.$$

Direct computation of the inverse transform (using 2.11) of $\phi(u)$ as defined by (2.9) might be troublesome. However, the Fourier transform (2.10) of the supposed limiting distribution function of (2.5)

$$f(x) = \begin{cases} 1, & 0 < x < 1, \\ 0 & \text{elsewhere} \end{cases}$$

is just the limiting function $\phi(u)$ as given by (2.9). Since $f(x)$ is of class $L_2(-\infty, \infty)$ and is of bounded variation, the theorem quoted above enables one to identify (2.5) as the limiting distribution function of the error for Case B, except for the values $f(0)$ and $f(1)$ where f should be chosen as 1/2.

The use of the Fourier transform $\phi_n(u)$, (as defined by (2.6)), is equivalent to the use of the characteristic functions of the jump distributions $\{f_{j,n}\}$. With this interpretation, it is possible to use Lévy's theorem [2, p. 101-102] to the effect that convergence of $\phi_n(u)$ to $\phi(u)$ implies the convergence of $F_n(x)$ to the limiting form $F(x)$ given by (2.4) and that $\phi(u)$ is the characteristic function of the cumulative distribution function $F(x)$.

The mean and variance of $f(x)$ as given by (2.5) (with or without modifications at 0 and 1) are 1/2 and 1/12 respectively, and thus agree with the values called for by relations (1.10).

3. Distribution of round-off errors for $k = 2$, Case C. Case C has symmetry noticeably lacking in Case B. For convenience, take $E_1 \equiv 0$ as before. Let $G_n(x)$ and $\{g_{j,n}\}$ be the cumulative and point-wise distribution functions. For this case

$$G_2(x) = \begin{cases} 0, & x < -1/2, \\ 1/4, & -1/2 \leq x < 0, \\ 3/4, & 0 \leq x < 1/2, \\ 1, & 1/2 \leq x, \end{cases} \tag{3.1}$$

$$\{g_{j,2}\} = \{1/4 \text{ at } -1/2,\ 1/2 \text{ at } 0,\ 1/4 \text{ at } 1/2\}.$$

Designate the finite Fourier transform (2.6) by $\psi_2(u)$. Then

$$\begin{aligned} \psi_2(u) &= (1/4)\exp(-iu/2) + 1/2 + (1/4)\exp(iu/2) \\ &= (1/4)[\exp(iu/4) + \exp(-iu/4)]^2 = \cos^2(u/4). \end{aligned} \tag{3.2}$$

This may be written in the form

$$\psi_2(u) = (1/4)[x + 1/x]^2 \quad \text{where} \quad x = \exp(iu/4). \tag{3.3}$$

Notice that to get $\{g_{j,3}\}$ from $\{g_{j,2}\}$ and the set $\{C\}$, $\{C\} = \{-1, 0, 1\}$ with probabilities $\{1/4, 1/2, 1/4\}$ respectively, one merely takes $1/4$ of the set $\{g_{j,2}\}$ on a smaller range at one end of the new range, $1/2$ of the set $\{g_{j,2}\}$ on a smaller range at the middle, and $1/4$ of the set $\{g_{j,2}\}$ on a smaller range at the other end of the new range. In effect, one goes from $\psi_2(u)$ to $\psi_3(u)$ by replacing x by x^2, multiplying by

$$[(1/4)x^2 + 1/2 + 1/(4x^2)] = (1/4)[x + 1/x]^2.$$

and then identifying $x = \exp(iu/8)$.

By this rule, one gets

$$\psi_3(u) = (1/4)^2 (x + 1/x)^2 (x^2 + 1/x^2)^2 = \cos^2(u/4)\cos^2(u/8). \tag{3.4}$$

Proceeding by induction, one gets

$$\psi_{n+1}(u) = \cos^2(u/4)\cos^2(u/8)\cdots\cos^2(u/2^{n+1}). \tag{3.5}$$

The sequence of transforms $\{\psi_n(u)\}$ has a limit,

$$\psi(u) = \lim_{n\to\infty} \psi_n(u) = \frac{\sin^2(u/2)}{(u/2)^2}, \tag{3.6}$$

by use of a well-known infinite product.

Direct computation of the inverse transform of (3.6) may be troublesome. However, it may be verified quite readily that if

$$g(x) = \begin{cases} 1 + x, & -1 < x < 0, \\ 1 - x, & 0 \le x < 1, \\ 0, & \text{elsewhere}, \end{cases} \tag{3.7}$$

the Fourier transform of $g(x)$ is just $\psi(u)$ of (3.6). Then, by use of (2.11), it follows that $g(x)$ as defined above may be taken as the pointwise distribution function for the limiting distribution E_C.

Direct computations show that

$$\mu(E_C) = 0, \quad \mathrm{Var}(E_C) = 1/6,$$

which values are in agreement with relations (1.10).

4. Conclusion. For higher values of k, the limits of the Fourier transforms may be difficult to obtain.

A somewhat more general problem would be to take

$$S_{n+1} = \frac{(k-m)\, S_n + mG_{n+1}}{k} \tag{4.1}$$

instead of (1.1), where k and m are both positive integers. In effect, however, this merely allows the k in (1.1) to be a positive rational number instead of a positive integer.

An equivalent statement of the problem would be to consider the distribution of $M(d)$, where

$$M(d) = \frac{1}{k} \sum_{i=0}^{\infty} d_i \left(\frac{k-1}{k}\right)^i, \tag{4.2}$$

and where $\{d_i\}$ is selected from the set D according to the value of k and the end-point choice. For the expansion of $M(d)$ is

$$M(d) = (1/k)\{d_0 + (k-1)/k\{d_1 + (k-1)/k\{d_2 + \cdots\}\}\}, \tag{4.3}$$

and this is just the scoring used in (1.4) but with reversed numerical ordering. Thus for $k = 2$ and Case B, M is uniformly distributed on $(0, 1)$, while for Case C, M has a house-top distribution on $(-1, 1)$.

REFERENCES

1. E. C. Titchmarsh, *Theory of Fourier integrals*, 2nd edition, Oxford, 1948.

2. M. G. Kendall, *The advanced theory of statistics*, volume 1, 4th edition, London, 1948.

UNIVERSITY OF TEXAS
HARVARD UNIVERSITY