NUMERICAL INTEGRATION BY SYSTEMATIC SAMPLING

By P. A. P. MORAN

Received 19 February 1949; *and in revised form* 7 May 1949

Recent investigations by F. Yates (1) in agricultural statistics suggest a mathematical problem which may be formulated as follows. A function $f(x)$ is known to be of bounded variation and Lebesgue integrable on the range $-\infty < x < \infty$, and its integral over this range is to be determined. In default of any knowledge of the position of the non-negligible values of the function the best that can be done is to calculate the infinite sum

$$S(t) = \delta \sum_{s=-\infty}^{\infty} f(t+s\delta)$$

for some suitable δ and an arbitrary origin t, where s ranges over all possible positive and negative integers including zero. S is evidently of period δ in t and ranges over all its values as t varies from 0 to δ. Previous writers (Aitken (2), p. 45, and Kendall (3)) have examined the resulting errors for fixed t. (They considered only symmetrical functions, and supposed one of the lattice points to be located at the centre.) Here we do not restrict ourselves to symmetrical functions and consider the likely departure of $S(t)$ from J (the required integral) *when t is a random variable uniformly distributed in* $(0, \delta)$. It will be shown that $S(t)$ is distributed about J as mean value, with a variance which will be evaluated as a function of δ, the scale of subdivision.

As our interest is here in the results, in view of their application we shall not attempt to find conditions of validity more general than those stated, from which it follows in particular that

$$\sum_{s=-N}^{N} f(t+s\delta)$$

is bounded uniformly with respect to t and N. Writing $E(S)$ for the expected value of S in the above circumstances, we have

$$E(S) = \frac{1}{\delta}\int_0^{\delta} \delta \sum_{-\infty}^{\infty} f(t+s\delta)\,dt = \int_{-\infty}^{\infty} f(t)\,dt. \tag{1}$$

We also have

$$\operatorname{var}(S) = E(S^2) - \{E(S)\}^2,$$

where

$$E(S^2) = \delta \int_0^{\delta} \left[\sum_{-\infty}^{\infty} f(t+s\delta)\right]^2 dt$$

$$= \delta \sum_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(t)\,f(t+s\delta)\,dt. \tag{2}$$

These integrals exist since $f(x)$ is clearly also of class $L^2(-\infty, \infty)$. Formula (2) may be evaluated directly in particular cases, but it is often more convenient to use a formula involving the Fourier transform

$$\phi(t) = \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{\infty} e^{ixt} f(x)\,dx. \tag{3}$$

https://doi.org/10.1017/S0305004100025524 Published online by Cambridge University Press

Consider the function $$f_1(y) = \int_{-\infty}^{\infty} f(x) f(x+y)\, dx.$$

It is easy to see that this is an even continuous function of bounded variation, integrable $L(-\infty, \infty)$. Its Fourier cosine transform is

$$\begin{aligned} F_c(t) &= \sqrt{\left(\frac{2}{\pi}\right)} \int_0^{\infty} f_1(y) \cos yt\, dy \\ &= \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{\infty} f_1(y)\, e^{iyt}\, dy \\ &= \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{\infty} e^{iyt}\, dy \int_{-\infty}^{\infty} f(x) f(x+y)\, dx \\ &= \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{\infty} f(x)\, e^{-ixt}\, dx \int_{-\infty}^{\infty} f(x+y)\, e^{i(x+y)t}\, dy \\ &= \sqrt{(2\pi)}\, \phi(t) \phi(-t) = \sqrt{(2\pi)} |\phi(t)|^2. \end{aligned}$$

Then from the theorem on Poisson's formula given by Titchmarsh ((4), p. 61), we have

$$\alpha^{\frac{1}{2}} \sum_{-\infty}^{\infty} f_1(n\alpha) = \beta^{\frac{1}{2}} \sum_{-\infty}^{\infty} F_c(n\beta),$$

where $\alpha\beta = 2\pi$. Putting $\alpha = \delta$ we get

$$\delta \sum_{-\infty}^{\infty} f_1(n\delta) = 2\pi \sum_{-\infty}^{\infty} \left| \phi\left(\frac{2n\pi}{\delta}\right) \right|^2,$$

and thus

$$E(S^2) = 2\pi \sum_{-\infty}^{\infty} \left| \phi\left(\frac{2n\pi}{\delta}\right) \right|^2. \tag{4}$$

This formula is often more convenient than (2). Mr D. G. Kendall has pointed out to me that the above result can also be obtained by Fourier series. For $S(t)$ is periodic with period δ and we can write

$$S(t) = \sum_{m=-\infty}^{\infty} c_m \exp(-2\pi imt/\delta),$$

where

$$\begin{aligned} c_m &= \frac{1}{\delta} \int_0^{\delta} S(t) \exp(2\pi imt/\delta)\, dt \\ &= \sqrt{(2\pi)}\, \phi(2\pi m/\delta). \end{aligned}$$

Also Parseval's theorem gives

$$\frac{1}{\delta} \int_0^{\delta} \{S(t)\}^2\, dt = \sum_{m=-\infty}^{\infty} |c_m|^2,$$

and $|c_0|^2 = J^2$, so that

$$\operatorname{var}(S(t)) = 2\pi \sum_{m=-\infty}^{\infty}{}' \left| \phi\left(\frac{2\pi m}{\delta}\right) \right|^2,$$

the dash implying the omission of the term for which $m = 0$. The equivalence of the two methods is thus a result of Poisson's formula.

We now consider some typical cases. If $f(x)$ is the normalized Gaussian function

$$\frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}x^2}, \tag{5}$$

https://doi.org/10.1017/S0305004100025524 Published online by Cambridge University Press

we have $E(S) = 1$. Now

$$\int_{-\infty}^{\infty} f(x)\, f(x+s\delta)\, dx = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2-\frac{1}{2}(x+s\delta)^2}\, dx$$

$$= \frac{1}{2\sqrt{\pi}} e^{-\frac{1}{4}s^2\delta^2},$$

and so

$$E(S^2) = \frac{\delta}{2\sqrt{\pi}} \sum_{-\infty}^{\infty} e^{-\frac{1}{4}s^2\delta^2}. \tag{6}$$

But we also have

$$\phi(t) = \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}t^2},$$

and so

$$E(S^2) = \sum_{p=-\infty}^{\infty} \exp\left[-\frac{4p^2\pi^2}{\delta^2}\right]. \tag{7}$$

The equivalence of this to (6) is a well-known formula in the theory of theta functions.

Yates (1) gives a numerical investigation of this case. He supposes $f(x)$ given by (5) and takes $\delta = 2$. He then calculates S to seven decimal places for $t = 0, 0{\cdot}1, \ldots, 1{\cdot}9$. The resulting values are, in view of the large value of δ, surprisingly close to unity, and a numerical estimate of var (S) may be found by calculating the variance of these values from unit mean. This variance was found to be 0·0001034460. Using (7), we find that the theoretical value for the variance is

$$2(e^{-\pi^2} + e^{-4\pi^2} + e^{-9\pi^2} + \ldots). \tag{8}$$

Taking the value $2e^{-\pi^2}$ as a first approximation we obtain 0·0001034476, which agrees with the above to five significant places, the discrepancy being almost entirely due to the experimentally obtained variance being based on only twenty values. Formula (7) is here more useful than (6), because so long as $\delta < 2\sqrt{\pi} = 3{\cdot}5449$, the series (7) converges more rapidly than (6). It is unlikely that we should ever want to use a δ greater than this. Moreover, in most cases (8) is very well represented by its first term.

Similar results hold for other functions. Consider the rectangular function

$$f(x) = l^{-1} \quad (-\tfrac{1}{2}l \leqslant x \leqslant \tfrac{1}{2}l),$$

$$= 0 \text{ elsewhere.}$$

Then

$$\phi(t) = \frac{1}{l\sqrt{(2\pi)}}\int_{-\frac{1}{2}l}^{\frac{1}{2}l} e^{ixt}\, dt = \frac{1}{\sqrt{(2\pi)}} \frac{\sin \frac{1}{2}lt}{\frac{1}{2}lt} \quad \text{for} \quad t \neq 0,$$

$$= 1 \quad \text{for} \quad t = 0,$$

and so

$$2\pi \sum_{-\infty}^{\infty} \left|\phi\left(\frac{2p\pi}{\delta}\right)\right|^2 = 1 + \frac{2\delta^2}{l^2\pi^2} \sum_{1}^{\infty} \frac{\sin^2 lp\pi/\delta}{p^2}. \tag{9}$$

On the other hand,

$$\int_{-\infty}^{\infty} f(x)\, f(x+s\delta)\, dx = l^{-2}(l - |s\delta|) \quad \text{if} \quad |s\delta| \leqslant l,$$

$$= 0 \quad \text{elsewhere.}$$

Then

$$\delta \sum_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x)\, f(x+s\delta)\, dx = \delta \sum_{s=-p}^{s=p} l^{-2}(l - |s\delta|),$$

https://doi.org/10.1017/S0305004100025524 Published online by Cambridge University Press

where $l = p\delta + q$ and $0 \leqslant q < \delta$, so that p is the integral part of $l\delta^{-1}$. This equals

$$\delta \sum_{s=-p}^{s=p} l^{-2}(p\delta + q - |s\delta|) = l^{-2}(l^2 - q^2 + q\delta) \,, \tag{10}$$

$$= 1 \quad \text{when} \quad q = 0.$$

The equality of (9) and (10) is the Poisson formula

$$1 + \delta^2 l^{-2}[\{l\delta^{-1}\} - \{l\delta^{-1}\}^2] = 1 + \frac{2\delta^2}{l^2\pi^2} \sum_1^\infty \frac{\sin^2 lp\pi/\delta}{p^2},$$

where $\{l\delta^{-1}\}$ is the fractional part of $l\delta^{-1}$. In this case (10) is the more convenient formula to use.

If $f(x)$ is the probability density in the Cauchy distribution,

$$f(x) = \frac{1}{\pi(1+x^2)},$$

and the results are even simpler. We then have

$$\phi(t) = \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{\infty} \frac{e^{ixt}\,dx}{\pi(1+x^2)} = \frac{e^{-|t|}}{\sqrt{(2\pi)}}$$

and

$$E(S^2) = 2\pi \sum_{-\infty}^{\infty} \left|\phi\left(\frac{2p\pi}{\delta}\right)\right|^2$$

$$= 1 + 2\sum_1^\infty e^{-4p\pi/\delta} = \frac{1+e^{-4\pi/\delta}}{1-e^{-4\pi/\delta}}. \tag{11}$$

Alternatively, we have

$$\int_{-\infty}^{\infty} f(x)\,f(x+s\delta)\,dx = \frac{1}{\pi^2}\int_{-\infty}^{\infty} \frac{dx}{(1+x^2)\,(1+(x+s\delta)^2)}$$

$$= \frac{2}{\pi(4+s^2\delta^2)}.$$

Thus

$$E(S^2) = \frac{2\delta}{\pi}\left\{\tfrac{1}{4} + 2\sum_1^\infty \frac{1}{4+s^2\delta^2}\right\} = \frac{1+e^{-4\pi/\delta}}{1-e^{-4\pi/\delta}},$$

which is another well-known example of Poisson's formula (Titchmarsh (4), p. 64).

The two-variable analogue of the above problem also has both theoretical and practical interest. A case which has arisen in practice is that in which $f(x, y)$ is a two-dimensional Gaussian function,

$$f(x,y) = \frac{1}{2\pi\sqrt{(1-\rho^2)}} \exp\left[-\frac{1}{2(1-\rho^2)}\{x^2 - 2\rho xy + y^2\}\right],$$

where ρ is the correlation coefficient. Let $0 < \delta_1, \delta_2 < \infty$, and consider the sum

$$S = \delta_1\delta_2 \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} f(t_1 + m\delta_1, t_2 + n\delta_2)$$

an approximation to

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1,$$

https://doi.org/10.1017/S0305004100025524 Published online by Cambridge University Press

where t_1, t_2 are random variable distributed independently in rectangular distributions on $(0, \delta_1)$, $(0, \delta_2)$ respectively. Then, as before,

$$E(S) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1,$$

and

$$E(S^2) = E\left\{\delta_1\delta_2 \sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} f(t_1+m\delta_1,\, t_2+n\delta_2)\right\}^2$$

$$= \delta_1\delta_2 \sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x+m\delta_1,\, y+n\delta_2)\,dx\,dy. \tag{12}$$

Write

$$\phi(u,v) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{i(xu+yv)} f(x,y)\,dx\,dy$$

$$= \frac{1}{2\pi}\exp\left[-\tfrac{1}{2}(u^2+2\rho uv+v^2)\right].$$

Then, arguing in the same way as in the proof of Poisson's formula in the one-variable case, we find that

$$E(S^2) = \sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} \exp\left[-4\pi^2\left\{\frac{m^2}{\delta_1^2}+\frac{2\rho mn}{\delta_1\delta_2}+\frac{n^2}{\delta_2^2}\right\}\right], \tag{13}$$

which, in most practical cases, converges faster than (12).

An interesting theoretical case occurs when $f(x, y)$ is equal to unity (say) at all the points inside some geometrical figure (e.g. the circle $x^2+y^2 = R^2$) and is zero elsewhere. The theory then connects up with the starting point of D. G. Kendall's paper (5) on the estimation of the area of a general oval curve by counting the number of lattice points which fall in it when a lattice of random position is superimposed on the figure.

REFERENCES

(1) Yates, F. *Philos. Trans.* A, 241 (1948), 345–77.
(2) Aitken, A. C. *Statistical mathematics* (Edinburgh: Oliver and Boyd, 1939).
(3) Kendall, D. G. *Quart. J. Math.* 13 (1942), 172–84.
(4) Titchmarsh, E. C. *Theory of Fourier integrals* (Oxford, 1937).
(5) Kendall, D. G. *Quart. J. Math.* 19 (1948), 1–26.

Institute of Statistics
Oxford University

https://doi.org/10.1017/S0305004100025524 Published online by Cambridge University Press