The Evaluation of Probabilities in a Normal Multivariate Distribution, with Special Reference to the Correlation Ratio

By M. H. QUENOUILLE

(*Received* 10*th December*, 1946. *Read* 17*th January*, 1947.)

1. *Introduction.*

This paper deals with the multivariate normal distribution in the symmetrical case when the frequency function may be reduced to

$$f(x_1, x_2, \ldots, x_n) = \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}n}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} x_i^2\right).$$

This is the usual distribution occurring for a series of similar independent observations or for the firing of shots at a target.

The problem of evaluating the probability that an observation $(x_1\, x_2 \ldots x_n)$ falls within a given region presents no problem when the region is rectilinear, but when we wish to evaluate the probability of an observation falling within a given distance of any point P, several difficulties arise. It is this problem which is presented in this paper.

2. *Basis of the problem.*

If P is the point $(a_1, a_2, \ldots, a_n)$, then the problem is that of finding the distribution function of $\left[\sum_{i=1}^{n} (x_i - a_i)^2\right]^{\frac{1}{2}}$, which is equivalent to finding the distribution function of the correlation ratio R in large samples, *i.e.* when σ is known, but when the correlation ρ in the population is not zero. This problem has been investigated by other writers (Fisher 1928, Gårding 1943), and the frequency function of R in the present paper has been obtained previously. Particular applications of this theory are provided by the distribution of second moment of a set of observations about any point, and by the test of difference between observations from two Poisson series (Fisher, *loc. cit.*).

However, this problem was originally encountered in the estimation of the relative risks at different distances from an aiming point when each projectile has a zone of risk (in fact, each projectile may be considered to have a general symmetrical distribution of risk, not necessarily normal) and this has given rise to the following approach.

https://doi.org/10.1017/S0013091500002613 Published online by Cambridge University Press

3. *Case of Two Variables.*

The bivariate or plane case will be demonstrated first. Suppose O is the origin, and P is at distance σa $(= \sqrt{\Sigma a_i^2})$ from O. Now we require the probability that $\sum_{i=1}^{n} (x_i - a_i)^2 \leqq \sigma^2 z^2$, *i.e.* that X falls within the circle A with radius σz and centre P, where z is the multiple correlation coefficient. This is given by

$$\iint_A \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x_1^2 + x_2^2}{2\sigma^2}\right) dx_1\, dx_2.$$

Now if we make a change of variables to t and θ, so that t is the distance from X to P and θ is the angle OPX, then

$$\begin{aligned} x_1^2 + x_2^2 &= OX^2 \\ &= OP^2 + PX^2 - 2OP\,.\,PX \cos OPX \\ &= \sigma^2 (a^2 + t^2 - 2\,at \cos\theta) \end{aligned}$$

and $\quad dx_1\, dx_2 = \sigma^2 t\, dt\, d\theta.$

Thus the probability is given by the integral

$$\int_0^z \int_0^\pi \frac{t}{\pi} \exp\left\{-\tfrac{1}{2}(t^2 + a^2 - 2\,at\cos\theta)\right\} d\theta\, dt.$$

Now $\int_0^\pi \exp\,(at \cos\theta)\, d\theta = \pi I_0(at)$, where $I_n(x)$ is the Bessel function of order n, defined by $I_n(x) = i^n J_n(ix)$, so that the expression for the probability is reduced to $\int_0^z t \exp\left(-\frac{a^2+t^2}{2}\right) I_0(at)\, dt.$

4. *General case.*

For the general case involving n observations the integral obtained from the probability is

$$Z_n(z, a) = \int \cdots \int_V \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}n}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2\right) dx_1 \ldots dx_n,$$

where V is the solid enclosed by the hypersphere, radius σz, centre P.

By a similar change of co-ordinates this becomes

$$\int_0^z \int_0^\pi \frac{t\,(t\sin\theta)^{n-2}}{\Gamma(\frac{1}{2}n - \frac{1}{2})\, 2^{\frac{1}{2}n-1}\sqrt{\pi}} \exp\left\{-\tfrac{1}{2}(t^2 + a^2 - 2\,at\cos\theta)\right\} d\theta\, dt.$$

Now, for $a \neq 0$ and $n > 1$, we have the formula

https://doi.org/10.1017/S0013091500002613 Published online by Cambridge University Press

$$\int_0^\pi \sin^{n-2}\theta \exp(at\cos\theta)\,d\theta = \left(\frac{2}{at}\right)^{\frac{1}{2}n-1} \Gamma(\tfrac{1}{2})\,\Gamma(\tfrac{1}{2}n-\tfrac{1}{2})\,I_{\frac{1}{2}n-1}(at)$$ [1]

and so

$$Z_n(z, a) = \int_0^z t\left(\frac{t}{a}\right)^{\frac{1}{2}n-1} \exp\left(-\frac{a^2+t^2}{2}\right) I_{\frac{1}{2}n-1}(at)\,dt,$$

$$Z_n(z, 0) = \frac{1}{2^{\frac{1}{2}n-1}\Gamma(\frac{1}{2}n)} \int_0^z t^{n-1} e^{-\frac{1}{2}t^2}\,dt.$$

We observe that $Z_n(z, 0) = \lim\limits_{a \to 0+} Z_n(z, a)$.

5. *A recurrence relation for* $Z_n(z, a)$, $a > 0$.

Integrating by parts and using

$$\frac{d}{dx}\left[x^m I_m(x)\right] = x^m I_{m-1}(x),$$

we get the relation

$$Z_n(z, a) = -\exp\left(-\frac{a^2+z^2}{2}\right)\left(\frac{z}{a}\right)^{\frac{1}{2}n-1} I_{\frac{1}{2}n-1}(az) + Z_{n-2}(z, a). \quad \text{(A)}$$

This relation, which will enable us to simplify greatly the reduction of the function $Z_n(z, a)$, shows us that the calculation of $Z_n(z, a)$ resolves itself into the two possibilities in which n is odd or even.

6. *Calculation of* $Z_n(z, a)$ *for odd* n.

If $n = 2m + 1$, we get from (A)

$$Z_{2m+1}(z, a) = Z_1(z, a) - \exp\left(-\frac{a^2+z^2}{2}\right)\sum_{r=0}^{m-1}\left(\frac{z}{a}\right)^{r+\frac{1}{2}} I_{r+\frac{1}{2}}(az) \quad \text{(B)}$$

where $\quad Z_1(z, a) = \displaystyle\int_0^z (at)^{\frac{1}{2}} \exp\left(-\frac{a^2+t^2}{2}\right) I_{-\frac{1}{2}}(at)\,dt.$

Now, for integral $m \geqq 0$, the formula

$$I_{m-\frac{1}{2}}(x) = \frac{2(2x)^{m-\frac{1}{2}}}{\sqrt{\pi}} \frac{d^m}{d(x^2)^m}\cosh x$$

gives

$$Z_1(z, a) = \int_0^z \frac{1}{\sqrt{2\pi}}\left[\exp -\frac{(a+t)^2}{2} + \exp -\frac{(a-t)^2}{2}\right] dt$$

$$= \int_{a-z}^{a+z} \frac{1}{\sqrt{2\pi}} \exp - t^2/2\,dt.$$

[1] For this formula see Whittaker and Watson, *Modern Analysis* (1940), p. 384, or, for a fuller account, Watson, *Bessel Functions* (1944), pp. 24 and 79.

https://doi.org/10.1017/S0013091500002613 Published online by Cambridge University Press

We see that although $Z_n(z, a)$ had only been proved to represent the probability for $n > 1$, it does, in fact, also represent this probability for $n = 1$.

We now have

$$Z_{2m+1}(z, a) = Z_1(z, a) - 2\exp\left(-\frac{a^2+z^2}{2}\right)\sum_{r=0}^{m-1}\frac{(2z^2)^{r+\frac{1}{2}}}{\sqrt{\pi}}\left[\frac{d^{r+1}}{d(x^2)^{r+1}}\cosh x\right]_{x=az}.$$

From this formula, given a, z and n, we can readily find the required probability, since $Z_1(z, a)$ can easily be found from a probit table, or from one of the many tables of the error integral, while the rest of the expression can, in general, be calculated from the ordinates of the normal distribution. For example,

$$Z_5(z, a) = Z_1(z, a) - \frac{1}{a^2\sqrt{2\pi}}\left[\left(a + z - \frac{1}{a}\right)\exp - \frac{(a-z)^2}{2} - \left(a - z - \frac{1}{a}\right)\exp - \frac{(a+z)^2}{2}\right].$$

When $a = 0$, the above forms do not hold and we get instead of (A)

$$Z_n(z, 0) = -\frac{z^{n-2}}{\Gamma(\frac{1}{2}n)\,2^{\frac{1}{2}n-1}}\exp - z^2/2 + Z_{n-2}(z, 0),$$

which may be proved by integration by parts in the same manner as (A), or deduced from (A) by making a tend to zero.

Then[1]

$$Z_{2m+1}(z, 0) = Z_1(z, 0) - \exp(-z^2/2)\sum_{r=0}^{m-1}\frac{1}{\Gamma(r+\frac{3}{2})}\left(\frac{z^2}{2}\right)^{r+\frac{1}{2}}$$

and $Z_1(z, 0) = \sqrt{\frac{2}{\pi}}\int_0^z \exp - t^2/2\, dt.$

7. *Calculation of $Z_n(z, a)$ for even n.*

As before, by using (A), we get the relationship

$$Z_{2m}(z, a) = Z_2(z, a) - \exp\left(-\frac{a^2+z^2}{2}\right)\sum_{r=1}^{m-1}\left(\frac{z}{a}\right)^r I_r(az) \qquad \text{(C)}$$

where $$Z_2(z, a) = \int_0^z t\exp\left(-\frac{a^2+t^2}{2}\right)I_0(at)\,dt.$$

---

[1] We note that, since $\lim_{n\to\infty} Z_n(z, 0) = 0$,

$$\frac{1}{\sqrt{2\pi}}\int_{-z}^{z}\exp - t^2/2\,dt = \exp - z^2/2\sum_{r=0}^{\infty}\frac{1}{\Gamma(r+\frac{3}{2})}\left(\frac{z^2}{2}\right)^{r+\frac{1}{2}}.$$

However, unless $Z_2(z, a)$ is known, we cannot calculate $Z_{2m}(z, a)$ as we calculated $Z_{2m+1}(z, a)$. We therefore use (A) again, noting that $\lim_{n\to\infty} Z_n(z, a) = 0$, and we get the series

$$Z_{2m}(z, a) = \exp\left(-\frac{a^2+z^2}{2}\right)\sum_{r=0}^{\infty}\left(\frac{z}{a}\right)^{m+r} I_{m+r}(az) \tag{D}$$

$$= 1 - \exp\left(-\frac{a^2+z^2}{2}\right)\sum_{r=1}^{\infty}\left(\frac{z}{a}\right)^{m-r} I_{m-r}(az), \tag{E}$$

where the latter expression is derived from the relation

$$\sum_{r=-\infty}^{\infty} p^r I_r(q) = \exp \tfrac{1}{2}q\left(p + \frac{1}{p}\right)$$

by inserting $p = z/a$, $q = az$.

For $a = 0$, we use (B), which reduces to

$$Z_{2m}(z, 0) = 1 - \exp(-z^2/2)\sum_{r=0}^{m-1}\frac{1}{r!}\left(\frac{z^2}{2}\right)^r.$$

This function is tabulated in Table 1 for values of m from 1 to 5 and z from 0.0 to 3·0.

Table 1

Table of $Z_{2m}(z, 0)$

z \ m	1	2	3	4	5
0.0	·00000	·00000	·00000	·00000	·00000
0·2	·01981	·00020	·00001	·00000	·00000
0·4	·07688	·00303	·00008	·00000	·00000
0·6	·16474	·01439	·00086	·00005	·00001
0·8	·27386	·04149	·00431	·00034	·00003
1·0	·39347	·09020	·01439	·00175	·00017
1·2	·51324	·16277	·03660	·00632	·00087
1·4	·62468	·25687	·07665	·01777	·00335
1·6	·72197	·36608	·13831	·04113	·01004
1·8	·80210	·48151	·22183	·08160	·02480
2·0	·86466	·59399	·32332	·14288	·05265
3·0	·98889	·93890	·82642	·65770	·46789

This function may be derived from the $I(u, p)$ function (tabulated by Pearson) using the relation $Z_{2m}(z, 0) = I(z^2/2\sqrt{m}, m-1.)$

8. *Calculation of* $Z_2(z, a)$.

$Z_{2m}(z, a)$ may be derived from (C) if $Z_2(z, a)$ and $I_r(az)$ $(r = 0, \ldots, m-1)$ are known. Now $I_r(x)$ has been tabulated for $r = 0$

https://doi.org/10.1017/S0013091500002613 Published online by Cambridge University Press

to 11 (a bibliography of tables may be found in Watson's *Bessel Functions*), so that the chief difficulty lies in calculating $Z_2(z, a)$. This may be easily carried out for z/a small or large, using the formulæ (D) and (E) with $m = 1$ while when $z = a$ these formulæ reduce to

$$2Z_2(z, z) = 1 - \exp(-z^2)\, I_0(z^2).$$

Two other formulæ that may be used for calculation may be obtained by expanding $I_0(az)$ in powers of az and integrating.

This gives

$$Z_2(z, a) = [1 - Z_2(a, 0)] \sum_{r=0}^{\infty} \frac{Z_{2r+2}(z, 0)}{r!} \left(\frac{a^2}{2}\right)^r$$

and

$$Z_2(z, a) = Z_2(z, 0) - \frac{z^2}{2}[1 - Z_2(z, 0)] \sum_{r=1}^{\infty} \frac{Z_{2r}(a, 0)}{r!} \left(\frac{z^2}{2}\right)^r.$$

These formulæ are useful for small values of z and a.

REFERENCES.

(1) R. A. Fisher, *Proc. Roy. Soc.* (A), 121 (1928), 654-673.

(2) L. Gårding, *Skandinavisk Aktuarietidskrift* (1941)), 185-202; reprinted in *Meddelanden Lunds Univ. Mat. Seminarium* 5 (1943).

STATISTICAL DEPARTMENT,
ROTHAMSTED EXPERIMENTAL STATION.