Some Properties of the Bivariate Normal Distribution Considered in the Form of a Contingency Table

Author(s): H. O. Lancaster

Source: *Biometrika*, Jun., 1957, Vol. 44, No. 1/2 (Jun., 1957), pp. 289-292

Published by: Oxford University Press on behalf of Biometrika Trust

Stable URL: https://www.jstor.org/stable/2333274

For all n except 101 the differences were checked by the method of interpolation recommended in the introduction to Pearson & Hartley's tables (1954, p. 13). For $n = 101$ the main term in (3) was found from the Wilson–Hilferty normal approximation.

The table shows in its comparison of (3) with Kathirgamatamby's results that the main term

$$F_{n-1} = \Pr(\chi^2_{n-1} \geqslant \chi^2_{n-1,\alpha}/a),$$

which depends only on $a = \kappa_2/\kappa_1$ and not on the higher cumulants, is a good approximation to the power for most practical purposes when $n = 51$ and 101. It is less than 2·5 % higher than Kathirgamatamby's figure even for μ as low as 5. If more accuracy is required the second term gives an extremely good correction down to $\mu = 5$ and in most cases down to $\mu = 1$.

We have no independent calculation by which to check the accuracy of the correction when $n = 6$, 10 and 20 and μ has the same values as before, although it is to be expected that it will not be as good as for $n = 51$ and 101.

My thanks are due to Miss B. I. Harley for checking the main part of the calculations, which were in some places in need of correction.

REFERENCES

Bartlett, M. S. (1935). *Proc. Camb. Phil. Soc.* **31**, 223.
Bateman, G. I. (1950). *Biometrika*, **37**, 59.
Cramér, H. (1946). *Mathematical Methods of Statistics*. Princeton University Press.
Gayen, A. K. (1949). *Biometrika*, **36**, 353.
Kathirgamatamby, N. (1953). *Biometrika*, **40**, 225.
Pearson, E. S. & Hartley, H. O. (1954). *Biometrika Tables for Statisticians*. **1**. Cambridge University Press.

Some properties of the bivariate normal distribution considered in the form of a contingency table

By H. O. LANCASTER

School of Public Health and Tropical Medicine, Sydney

1. In considering the properties of contingency tables and bivariate frequency distributions Karl Pearson (1904) showed that if a bivariate normal distribution is classified in a two-way table the contingency and the correlation parameter are related by the expression

$$\phi^2 \equiv \chi^2/N = \rho^2/(1-\rho^2). \tag{1}$$

It is assumed that N, the number of observations, is large and that the class intervals are very narrow —the equation, in fact, represents a limiting property. For a p-way classification based on a multivariate normal distribution of dimension p Pearson also showed that

$$\phi^2 = 1/\sqrt{(RR')} - 1, \tag{2}$$

where R is the determinant of the correlation matrix R and R' is the determinant of the matrix $(2\,\mathbf{1} - \mathbf{R})$.

It was on the basis of these results that Pearson proposed his coefficient of contingency

$$C = \left(\frac{\phi^2}{1+\phi^2}\right)^{\frac{1}{2}}. \tag{3}$$

In the limiting normal case this is equal to the correlation parameter ρ. More generally, it may be regarded as an invariative property of a distribution, being unchanged by any arbitrary, not necessarily linear transformation of the scale of either or both variates.

2. In 1940 Fisher considered contingency tables from the point of view of discriminant analysis. Suppose that 'scores', i.e. arbitrary variate values, are assigned to the rows and also to the columns of a contingency table: what are the best scores to assign to the rows so that a linear function of them will best differentiate the classes determined by the columns, and vice versa? This turns out to be a problem in maximizing the correlation between the scores and the required correlations are those known as

'canonical' in the sense of Hotelling (1936). The work was continued and developed by Maung (1941). In particular, Maung quotes a result by Fisher which gives the observed frequency in terms of the canonical correlations; in fact, if the frequency is a_{ij} with marginal *totals* $a_{i.}$, $a_{.j}$ and total $a_{..}$, and if the canonical correlations are $R_1, R_2, \ldots, R_{m-1}$, we have

$$a_{ij} = \frac{a_{i.}a_{.j}}{a_{..}}\left\{1+\sum_1^{m-1}(x_k y_k R_k)\right\}, \tag{4}$$

where x and y are the assigned scores corresponding to the given cell.

3. In this note I derive some further theorems in this field and link together some hitherto disconnected results. It will be shown that (3) is, in the limit, equivalent to the well-known Mehler identity or tetrachoric series:

$$(2\pi)^{-1}(1-\rho^2)^{-\frac{1}{2}}\exp\{-\tfrac{1}{2}(x^2-2\rho xy+y^2)/(1-\rho^2)\} = (2\pi)^{-1}\exp\{-\tfrac{1}{2}(x^2+y^2)\}\left\{1+\sum_1^\infty \psi_i(x)\,\psi_i(y)\,\rho^i\right\}, \tag{5}$$

where $\psi_i(x)$ is the Hermite–Tchebycheff polynomial of order i.

A maximal property of the bivariate normal distribution

4. We may consider the variables standardized so as to have unit variance and take $0<\rho<1$.

Theorem. *Let x and y be jointly distributed in the bivariate normal distribution with correlation ρ. If now a transformation, $x' = x'(x)$, $y' = y'(y)$, is made to any new variables x' and y' such that*

$$(2\pi)^{-\frac{1}{2}}\int_{-\infty}^{\infty} x'^2\exp-\tfrac{1}{2}x^2\,dx \quad \textit{and} \quad (2\pi)^{-\frac{1}{2}}\int_{-\infty}^{\infty} y'^2\exp-\tfrac{1}{2}y^2\,dy$$

are finite, then the correlation of the new variables is less in absolute value than ρ. That is, ρ is the maximum canonical correlation.

Under the conditions of the theorem, the new variables may be expressed in a series of standardized Hermite–Tchebycheff polynomials,

$$x' = a_0 + a_1\psi(x) + a_2\psi_2(x) + \ldots \tag{6}$$

such that $$(2\pi)^{-\frac{1}{2}}\int_{-\infty}^{\infty}\left\{\left(x'-a_0-\sum_1^\infty a_i\psi_i(x)\right)^2\exp(-\tfrac{1}{2}x^2)\,dx\right\}$$

is arbitrarily small and $\sum_1^\infty a_i^2$ convergent. Moreover, the correlation is unaffected by either a linear change of scale or a change of origin, so that without loss of generality we can write,

$$\left.\begin{aligned} x' &= \sum_1^\infty a_i\psi_i(x), \quad \sum_1^\infty a_i^2 = 1,\\ y' &= \sum_1^\infty b_i\psi_i(y), \quad \sum_1^\infty b_i^2 = 1,\end{aligned}\right\} \tag{7}$$

where the $\psi_i(x)$ obey the orthogonality relations

$$(2\pi)^{-\frac{1}{2}}\int_{-\infty}^{\infty}\psi_i(x)\,\psi_j(x)\exp-\tfrac{1}{2}x^2\,dx = \delta_{ij}. \tag{8}$$

By a consideration of the expectation of $\exp(tx-\frac{1}{2}tx^2+uy-\frac{1}{2}u^2)$, we find

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\psi_i(x)\,\psi_j(y)\,f(x,y)\,dx\,dy = \delta_{ij}\rho^i, \tag{9}$$

where $f(x,y)$ is the density function of the bivariate normal population.

The variance of x' and y' are easily found to be unity by the orthogonal properties of the ψ_j. We have then

$$\operatorname{corr}(x',y') = \sum_1^\infty a_i b_i \rho^i, \tag{10}$$

and this is less than $|\rho|$ unless $a_1 = b_1 = 1$. $|\rho|$ is therefore the maximum canonical correlation under a very general class of transformations. This was proved by Maung (1941) by an alternative method.

5. THEOREM. *The values to be assigned to the canonical variables corresponding to the canonical correlations in descending order are $\psi_i(x)$ and $\psi_i(y)$. The canonical correlations are the powers of ρ.*

We have already shown that the maximum canonical correlation corresponds with $\psi_1(x) = x$ and $\psi_1(y) = y$. Let us take a second set of values x'' and y'', such that

$$E(x'') = E(y'') = 0 \quad \text{and} \quad E(xx'') = E(yy'') = 0 \quad \text{and} \quad E(x'')^2 = 1 = E(y'')^2.$$

We may write
$$\left.\begin{aligned} x'' &= \sum_1^\infty c_i \psi_i(x), \\ y'' &= \sum_1^\infty d_i \psi_i(y). \end{aligned}\right\} \tag{11}$$

The conditions of the theorem again enable us to set $\sum_{i=1}^\infty c_i^2$ and $\sum_{d=1}^\infty d_i^2$ equal to unity.

Further $E(xx'') = 0$ forces c_1 to be zero and similarly d_1 is zero.

Now
$$\operatorname{corr}(x'', y'') = \sum_2^\infty c_i d_i \rho^i, \tag{12}$$

and this is maximal in absolute value only if $c_2 = d_2 = 1$ and all other $c_i = d_i = 0$.

This process can be extended by induction and proves the theorem.

COROLLARY 1. If a choice of variables can be made so that a joint bivariate normal distribution results from a contingency table, then the canonical correlations are powers of the greatest of them and the sets of canonical variables are the standardized Hermite–Tchebycheff polynomials.

COROLLARY 2. The roots of the determinantal equation usually solved in the treatment of this problem would be powers of the greatest of them but for disturbances due to sampling errors and difficulties caused by grouping.

COROLLARY 3. In the normal case the concept of minimum correlation has no validity. The lowest of the non-zero correlations will be ρ^{m-1} approximately and deviations from this value will be due again to sampling and difficulties caused by grouping.

FISHER'S IDENTITY AND MEHLER'S IDENTITY

6. Consider the identity (5) in the limiting case when the contingency table with elements $a_{ij}/a_{..}$ becomes the frequency function $f(x, y)\,dx\,dy$. We have

$$f(x,y)\,dx\,dy = \left\{1 + \sum_1^\infty \psi_i(x)\,\psi_i(y)\,\rho^i\right\}(2\pi)^{-1}\exp\{-\tfrac{1}{2}(x^2+y^2)\}\,dx\,dy, \tag{5 bis}$$

which is the Mehler series.

THE RELATIONSHIP BETWEEN χ^2 AND CORRELATION

7. In reconsidering the partition of χ^2 in a contingency table Lancaster (1953) used the Hermite–Tchebycheff polynomials to avoid the infinitely fine subdivision used by Pearson. In the notation of this paper he derived

$$\begin{aligned} \chi^2/a_{..} &= R_1^2 + R_2^2 + \dots + R_{m-1}^2 \\ &\sim \rho^2 + \rho^4 + \dots + \rho^{2m-2} = \rho^2(1-\rho^{2m-2})/(1-\rho^2). \end{aligned} \tag{13}$$

Evidently
$$\chi^2/a_{..} \geqslant R_1^2, \tag{14}$$

and R_1 is greater than any other observed correlation. A more general result is that in a complex contingency table
$$\chi^2/a_{..} \geqslant \sum_{i<j} r_{ij}^2, \tag{15}$$

where r_{ij} are the observed correlations between any variables. For if we partition χ^2 making use of the direct product of matrices (Lancaster, 1951) we may take the first row of each to be of the form $\sqrt{w_j}$ and the second row $x_j\sqrt{w_j}$, where x is suitably normalized, the second row being orthogonal to the first. The remaining rows may be filled in with due regard to the orthogonal conditions. In the observed case, the set of canonical variables for x, when the correlation between x and y is being maximized, may be different from the set when the correlation between x and z is being maximized. So that we cannot assert that $\chi^2/a_{..}$ is greater than the sum of the squares of the canonical correlations.

We can obtain a result analogous to (2):

$$\log(1+\phi^2) = -\tfrac{1}{2}\log(RR') \quad \text{(Pearson notation)}$$
$$= -\tfrac{1}{2}\log|1+\mathbf{P}| - \tfrac{1}{2}\log|1-\mathbf{P}|, \tag{16}$$

where $\mathbf{P}$ is the matrix with elements, $p_{ij} = r_{ij}$ for $i \neq j$, and $p_{ii} = 0$. If $\phi^2 < 1$, we may expand both sides using the expansion of Durbin & Watson (1950) on the right.

We have

$$\phi^2 - \tfrac{1}{2}\phi^4 + \tfrac{1}{3}\phi^6 \ldots = \tfrac{1}{2}\operatorname{tr}\mathbf{P}^2 + \tfrac{1}{4}\operatorname{tr}\mathbf{P}^4 \ldots,$$

and so

$$\phi^2 > \tfrac{1}{2}\operatorname{tr}\mathbf{P}^2 = \sum_{i<j} r_{ij}^2. \tag{17}$$

Even for $\phi^2 > 1$ this still holds, for

$$1+\phi^2 = \exp\{\tfrac{1}{2}\operatorname{tr}\mathbf{P}^2 + \tfrac{1}{4}\operatorname{tr}\mathbf{P}^4 \ldots\}$$
$$\geqslant 1 + \tfrac{1}{2}\operatorname{tr}\mathbf{P}^2,$$
$$\phi^2 \geqslant \sum_{i<j} r_{ij}^2. \tag{18}$$

There are various determinantal conditions on the r_{ij} derived by Pearson (1904) but it appears that they can all be summarized by saying that $(1+\mathbf{P})$ and $(1-\mathbf{P})$ of (16) are both positive definite. Pearson (1904) noted that the first is necessary for the system to be a correlation system and the second condition is necessary for ϕ^2 to have a finite limit.

Discussion

8. This paper relates some hitherto disconnected results. A maximal property of the bivariate distribution is proved. Further, in the theoretical case, we have derived the interesting result that the sets of canonical variables are the standardized Hermite–Tchebycheff polynomials and the canonical correlations the corresponding power of $|\rho|$. This result raises a doubt as to whether the common procedure of considering the roots of the determinantal equation separately can be justified if the hypothesis specifies a bivariate normal population. It would appear that once the first set of canonical variables are obtained, little further is gained by examining other sets since they differ from certain polynomial functions of the first set of canonical variables only because of sampling difficulty. Furthermore, a consideration of the hypothesis of the existence of one canonical correlation as in Williams (1952) shows that a value of the canonical correlation gives a limit to the range of the canonical variables; for the modified hypothesis is

$$p_{ij} = p_{i.}p_{.j}(1+\rho x_j y_i), \tag{19}$$

and $p_{ij} \geqslant 0$, so that

$$x_j y_i \geqslant -\rho^{-1} \tag{20}$$

for any pair of y_i and x_j. But both x and y must take some positive values and some negative values under the maximization procedure, so that (20) can only hold if neither x nor y have an infinite range. It appears that these difficulties have not been thoroughly explored as yet.

This paper is published with the permission of the Director-General of Health, Canberra.

REFERENCES

Durbin, J. & Watson, G. S. (1950). Testing for serial correlation in least squares regression. I. *Biometrika*, **37**, 409.

Fisher, R. A. (1940). The precision of discriminant functions. *Ann. Eugen., Lond.*, **10**, 422.

Fisher, R. A. (1941). Cited by Maung (1941).

Hotelling, H. (1936). Relations between two sets of variables. *Biometrika*, **28**, 321.

Lancaster, H. O. (1951). Complex contingency tables treated by the partition of χ^2. *J. R. Statist. Soc.* B, **13**, 242.

Lancaster, H. O. (1953). A reconciliation of χ^2. *Sankhyā*, **13**, 1.

Maung, K. (1941). Measurement of association in a contingency table with special reference to the pigmentation of hair and eye colours of Scottish school children. *Ann. Eugen., Lond.*, **11**, 189.

Pearson, K. (1904). Mathematical contributions to the theory of evolution. XIII. On the theory of contingency and its relation to association and normal correlation. *Drap. Co. Mem. Biom. Ser.* no. 1.

Williams, E. J. (1952). Use of scores for the analysis of association in contingency tables. *Biometrika*, **39**, 274.