On the structure of the tetrachoric series

By M. A. HAMDAN

American University of Beirut

SUMMARY

Pearson (1900) introduced the tetrachoric series method for estimating the correlation between two non-measurable characters each with two levels. For characters with more than two levels, Ritchie-Scott (1918) suggested averaging all possible tetrachoric correlations. Using the theory of orthonormal functions, Lancaster & Hamdan (1964) suggested an alternative method essentially based on giving a weighting to each possible tetrachoric table. In this note, a special form of Lancaster & Hamdan's method is used to give an instructive derivation of the tetrachoric series.

Let a set of N observations (x_j, y_j) $(j = 1, 2, \ldots, N)$ be made on a bivariate normal variable (x, y) with coefficient of correlation ρ and density function $f(x, y, \rho)$ given by the Mehler identity (1866)

$$f(x, y, \rho) = \phi(x)\,\phi(y) \sum_{i=0}^{\infty} \rho^i H_i(x)\, H_i(y), \tag{1}$$

where $\phi(x)$ is the unit normal density function, and $\{H_i(x)\}$ is the set of standardized Hermite–Chebyshev polynomials orthonormal on the unit normal distribution. Let the total frequency N be divided into four parts a, b, c and d by two planes at right angles to the axes of x and y at distances h and k from the origin respectively, thus getting the 2×2 table

$$\begin{array}{c|c} a & b \\ \hline c & d \end{array}.$$

Let W be a random variable taking values $(p/q)^{\frac{1}{2}}$ and $-(q/p)^{\frac{1}{2}}$ with probabilities q and p, respectively. Then W is orthonormal on the two-point distribution. Define $u(x)$ by setting

$$p = (b+d)/N \quad \text{and} \quad W = u(x). \tag{2}$$

Define $v(x)$ by setting

$$p = (c+d)/N \quad \text{and} \quad W = v(x). \tag{3}$$

Now, if X is defined by

$$X = N^{-\frac{1}{2}} \sum_{j=1}^{N} u(x_j)\, v(y_j), \tag{4}$$

then under the hypothesis Ω_0 $(\rho = 0)$, X is asymptotically distributed as normal $(0, 1)$, by a theorem of Bernstein (1926). Under the hypothesis Ω_1 $(\rho \neq 0)$, X is a non-central normal variable, i.e. X^2 is asymptotically distributed as a chi-square variable with one degree of freedom, and noncentrality parameter

$$\lambda(\rho) = E^2(X \mid \Omega_1) = NE^2\{u(x)\, v(y) \mid \Omega_1\}. \tag{5}$$

To express $\lambda(\rho)$ in terms of ρ, let the functions $u(x)$ and $v(y)$ be represented as Fourier series in the sets $\{H_n(x)\}$ and $\{H_n(y)\}$,

$$u(x) = \sum_{n=0}^{\infty} a_n H_n(x), \quad \text{where} \quad a_0 = 0, \quad \sum_{n=1}^{\infty} a_n^2 = 1; \tag{6}$$

$$v(y) = \sum_{n=0}^{\infty} b_n H_n(y), \quad \text{where} \quad b_0 = 0, \quad \sum_{n=1}^{\infty} b_n^2 = 1. \tag{7}$$

By (6), (7), the Mehler identity (1) and the properties of the Hermite–Chebyshev polynomials (Szegö, 1959), we get

$$E\{u(x)\, v(y) \mid \Omega_1\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u(x)\, v(y) f(x, y, \rho)\, dx\, dy = \sum_{n=1}^{\infty} a_n b_n \rho^n, \tag{8}$$

so that

$$\lambda(\rho) = N\left\{\sum_{n=1}^{\infty} a_n b_n \rho^n\right\}^2. \tag{9}$$

The coefficients a_n and b_n are given by

$$\begin{aligned} a_n &= \int_{-\infty}^{\infty} u(x)\, H_n(x)\, \phi(x)\, dx \\ &= \{(b+d)/(a+c)\}^{\frac{1}{2}} \int_{-\infty}^{h} H_n(x)\, \phi(x)\, dx - \{(a+c)/(b+d)\}^{\frac{1}{2}} \int_{h}^{\infty} H_n(x)\, \phi(x)\, dx \\ &= (1/\sqrt{n})\, H_{n-1}(h)\, \phi(h) \left\{ \left(\frac{b+d}{a+c}\right)^{\frac{1}{2}} + \left(\frac{a+c}{b+d}\right)^{\frac{1}{2}} \right\} \end{aligned} \tag{10}$$

and similarly

$$b_n = (1/\sqrt{n})\, H_{n-1}(k)\, \phi(k) \left\{ \left(\frac{c+d}{a+b}\right)^{\frac{1}{2}} + \left(\frac{a+b}{c+d}\right)^{\frac{1}{2}} \right\}. \tag{11}$$

Substituting from (10) and (11) in (9), we get

$$\left\{\frac{\lambda(\rho)}{N}\right\}^{\frac{1}{2}} = \frac{N^2 \phi(h)\, \phi(k)}{\sqrt{\{(a+c)(b+d)(a+b)(c+d)\}}} \sum_{n=1}^{\infty} \frac{\rho^n}{n} H_{n-1}(h)\, H_{n-1}(k). \tag{12}$$

Finally, taking the value of the chi-square for the fourfold table, namely,

$$\frac{N(ad-bc)^2}{(a+c)(b+d)(a+b)(c+d)}$$

as an estimate of the noncentrality parameter $\lambda(\rho)$ and substituting in (12), we get

$$\frac{ad-bc}{N^2 \phi(h)\, \phi(k)} = \sum_{n=1}^{\infty} \frac{\rho^n}{n} H_{n-1}(h)\, H_{n-1}(k). \tag{13}$$

Noting that Pearson's tetrachoric functions are the standardized Hermite–Chebyshev polynomials, it becomes obvious that (13) is Pearson's tetrachoric series.

REFERENCES

BERNSTEIN, S. (1926). Sur l'extension du théorème limite du calcul des probabilités aux somes de quantites dépendantes. *Math. Annln* **97**, 1–79.

LANCASTER, H. O. & HAMDAN, M. A. (1964). Estimation of the correlation coefficient in contingency tables with possibly nonmetrical characters. *Psychometrika* **29**, 383–91.

MEHLER, F. G. (1866). Über die Entwicklung einer Funktion von beliebig vielen variablen nach Laplaceschen Funktionen höherer Ordnung. *J. Mathematik* **66**, 161–76.

PEARSON, K. (1900). Mathematical contribution to the theory of evolution. VII: On the correlation of characters not quantitatively measurable. *Phil. Trans.* A **195**, 1–47.

RITCHIE-SCOTT, A. (1918). The correlation coefficient of a polychoric table. *Biometrika* **12**, 93–133.

SZEGÖ, G. (1959). Orthogonal polynomials. *Colloquium Publ.* no. 23, *Am. Math. Soc.*

[*Received May* 1967. *Revised October* 1967]

A note on contingency-type bivariate distributions

By G. P. STECK

Sandia Corporation, Albuquerque, New Mexico

SUMMARY

Some results are proved about a family of bivariate distributions introduced by Plackett.

Let X and Y be random variables with distribution functions F and G and density functions f and g. Plackett (1965) defined a class of bivariate distribution functions for (X, Y) which are indexed by a parameter ψ measuring the association between X and Y as that root of the equation

$$\psi = \frac{H(1-F-G+H)}{(F-H)(G-H)}$$