POLYNOMIAL EXPANSIONS OF BIVARIATE DISTRIBUTIONS[1]

By G. K. Eagleson

University of Sydney

1. Introduction. Bivariate distributions subject to the condition of ϕ^2 boundedness, to be defined later, may be expanded in a canonical form (e.g. Lancaster [3]). In this paper, a class of bivariate distributions, whose canonical variables are the orthonormal polynomials of the marginal distributions, is exhibited. This class consists of the bivariate normal, a bivariate gamma, Poisson, negative binomial, binomial and hypergeometric.

The identity in Hermite-Chebyshev polynomials due to Runge [7] is generalised to this class of distributions, which have a particular type of generating function for the orthogonal polynomials. The completeness of the orthogonal polynomials of the same class is also proved.

The bivariate distributions are generated by considering sums of independent random variables, which are "additive" (i.e. closed under convolutions), where some of the variables are held in common. Pearson [6] generates the bivariate normal in this way and Cherian [1] a bivariate gamma.

The conditions on the parameters, in order that the distributions be ϕ^2 bounded, are obtained. The regressions are shown to be linear and the correlation coefficient is seen to be a satisfactory measure of dependence. Finally, a goodness of fit test for these bivariate distributions is outlined.

2. Distributions with a particular form of generating function for the orthogonal polynomials. Meixner [5] considered those distributions which have a generating function for their orthogonal polynomials of the form

$$G(t, x) = f(t)e^{xu(t)} = \sum_{n=0}^{\infty} P_n(x)t^n/n! \tag{2.1}$$

where $P_n(x) = x^n + a_{n,1}x^{n-1} + \cdots + a_{n,n}$, $f(t)$ is a power series in t with $f(0) = 1$ and $u(t)$ is a power series in t with $u(0) = 0$ and $u'(0) = 1$. Both $f(t)$ and $u(t)$ have real coefficients. We denote the functional inverse of $u(t)$ by $v(u)$, i.e. $v(u(t)) \equiv t$. When needed, subscripts will be used on $f(t)$ and $u(t)$ to indicate with which distribution they are associated. Thus

$$\begin{aligned}\int_{-\infty}^{\infty} f(t)f(s) \exp\{xu(t) + xu(s)\}\, d\psi(x) \\ = f(t)f(s)M(u(t) + u(s)) = \sum_{i=0}^{\infty} [(ts)^i/(i!)^2]c_i\end{aligned} \tag{2.2}$$

where $c_i = \int_{-\infty}^{\infty} P_i^2(x)\, d\psi(x)$, $\psi(x)$ is the distribution function of X, and $M(\theta) = E(e^{X\theta})$.

Received 27 November 1963.

[1] This research was undertaken while holding a C.S.I.R.O. Studentship at the University of Sydney.

THEOREM 2.1. *The moment generating function* (m.g.f.) *of* $\{P_i(x)\, d\psi(x)\}$ *is* $[f(v(\theta))]^{-1}\{[v(\theta)]^i/i!\}c_i$.

PROOF.

$$\int_{-\infty}^{\infty} e^{x\theta}f(t)e^{xu(t)}\, d\psi(x) = \int_{-\infty}^{\infty} \sum_{i=0}^{\infty} e^{\theta x}P_i(x)t^i/i!\, d\psi(x)$$

$$= [f(s)]^{-1}f(s)f(t)\int_{-\infty}^{\infty} \exp\{x(u(s) + u(t))\}\, d\psi(x), \qquad \text{where } u(s) = \theta,$$

$$= [f(s)]^{-1}\sum_{i=0}^{\infty} s^i t^i c_i/(i!)^2.$$

The theorem follows from equating coefficients of $t^i/i!$.

COROLLARY 1. *The* m.g.f. *of* $d\psi(x)$ *is* $[f(v(\theta))]^{-1}$ *as* $c_0 = 1$.

COROLLARY 2. *The only frequency function* $g(x)$ *whose orthogonal polynomials are of the form* $g(x)P_i(x) = b_iD^i(g(x))$ *and which has the above form of generating function is the normal.*

PROOF. If $g(x)P_i(x) = b_iD^i(g(x))$, then the m.g.f. of $P_i(x)g(x)$ is $b_i\theta^i \times$ m.g.f. of $g(x)$, i.e. $u(\theta) = \theta$. But Meixner shows that the only distribution with $u(t) = t$ is the normal.

Meixner obtains the following recurrence relation for the orthogonal polynomials:

$$P_{n+1}(x) = (x - \mu + n\lambda)P_n(x) + n\{k_2 + (n-1)\kappa\}P_{n-1}(x), \qquad n = 0, 1, \cdots \tag{2.3}$$

where $k_2 < 0$, $\kappa \leqq 0$ and μ, which is the mean, can without loss of generality be made zero. He also obtains two differential equations for $f(t)$ and $v(u)$:

$$v'(u) = 1 - \lambda t - \kappa t^2 = (1 - \alpha t)(1 - \beta t) \tag{2.4}$$

$$f'(t)[f(t)]^{-1} = k_2t/[(1 - \lambda t - \kappa t^2)]. \tag{2.5}$$

LEMMA 2.1. *The constant* $-k_2$, *which appears in the recurrence relation for the orthogonal polynomials, is the variance of the weight function.*

PROOF. From the differential equation (2.5), determining $f(t)$ we see that $f(0) = 1$, $f'(0) = 0$ and $f''(0) = k_2$. Further $v(\theta) = \theta + v_2\theta^2 + v_3\theta^3 + \cdots$, for small enough θ. Thus the m.g.f. of the distribution $M(\theta) = [f(v(\theta))]^{-1} = 1 - \frac{1}{2}k_2\theta^2 + O(\theta^3)$ for small enough θ. The result follows immediately.

THEOREM 2.2. *Those distributions with the generating function of their orthogonal polynomials of the form* $f(t)e^{xu(t)}$ *are additive if and only if* $u(t)$ *is the same for all distributions in the class.*

PROOF. Necessity. We have, from (2.5), $\log f(t) = k_2w(t) + a$, where $w(t) = \int [t/(1 - \lambda t - \kappa t^2)]\, dt$ and a is a constant. If the distributions are additive, $\log f_X(t) + \log f_Y(t) = \log f_{X+Y}(t)$. Thus $w(t)$ must be the same for X and Y, except at most for a constant factor. Using the Lemma 2.1, we can show that the constant factor is unity. This implies that $1 - \lambda t - \kappa t^2$ and hence $v(u)$ and

$u(t)$ are the same for all distributions of the class. Sufficiency. This is obvious from the form of $\log f(t)$.

3. Generalisation of Runge's identity and completeness.

THEOREM 3.1. (Generalisation of Runge's identity) *If W_1 and W_2 are independent and additive variates whose orthogonal polynomials are generated by a function of the form $f(t)e^{w_i u(t)}$ and if $X = W_1 + W_2$ then*

$$P_n(x) = \sum_{i=0}^{n} \binom{n}{i} P_i(w_1)P_{n-i}(w_2).$$

PROOF. W_1 and W_2 additive implies that $u(t)$ is the same for W_1, W_2 and X. Also $f_X(t) = f_{W_1}(t)f_{W_2}(t)$. Thus

$$\sum_{n=0}^{\infty} P_n(x)t^n/n! = f_X(t)e^{xu(t)} = f_{W_1}(t)e^{w_1u(t)}f_{W_2}(t)e^{w_2u(t)}$$
$$= \sum_{n=0}^{\infty} P_n(w_1)t^n/n! \sum_{m=0}^{\infty} P_m(w_2)t^m/m!.$$

The theorem follows from equating coefficients of t^n.

THEOREM 3.2. *The orthogonal polynomials generated by a function of the form $f(t)e^{xu(t)}$ are a complete system with respect to the weight distribution.*

PROOF. If we denote the normalised system by $\{P_n^*(x)\}$ then a theorem due to Picone (see Sansone [8]) states that $\{P_n^*(x)\}$ is complete if and only if

$$\sum_{n=0}^{\infty} [\int_{-\infty}^{\infty} e^{isx}P_n^*(x)\, d\psi(x)]^2 = \int_{-\infty}^{\infty} d\psi(x) = 1.$$

From Theorem 2.1 we have

$$\int_{-\infty}^{\infty} e^{isx}P_n^*(x)\, d\psi(x) = [f(\alpha)]^{-1}\alpha^n \sqrt{c_n/n!}$$

where $is = u(\alpha)$, i.e., $\alpha = v(is)$. Then

$$\text{L.H.S.} = [f(\alpha)]^{-1} \times [\overline{f(\alpha)}]^{-1}\sum_{n=0}^{\infty}(\alpha\bar{\alpha})^n c_n/(n!)^2$$
$$= \int_{-\infty}^{\infty} \exp\{xu(\alpha) + xu(\bar{\alpha})\}\, d\psi(x) = 1$$

since $\overline{f(\alpha)} = f(\bar{\alpha})$ and $u(\bar{\alpha}) = \overline{u(\alpha)} = -is$ as both $f(t)$ and $u(t)$ are power series in t, with real coefficients.

4. Canonical expansions.

Definition. If $F(x, y)$ is a bivariate distribution function with marginal distributions $G(x)$ and $H(y)$ then we define

$$\phi^2 + 1 = \int \{dF/(dGdH)\}^2 dGdH = \int \Omega^2 dGdH \tag{4.1}$$

where $\Omega = \Omega(x, y)$ is the Radon-Nikodym derivative of $F(x, y)$ with respect to $G(x)H(y)$.

If $F(x, y)$ is ϕ^2 bounded then complete sets of orthonormal functions, $\{x^{(i)}\}$

and $\{y^{(i)}\}$, can be defined on the marginal distributions such that

$$dF(x, y) = \left\{1 + \sum_{r=1}^{\infty} \rho_r x^{(r)} y^{(r)}\right\} dG(x)\, dH(y) \qquad \text{a.e.} \tag{4.2}$$

and

$$\phi^2 = \sum_{r=1}^{\infty} \rho_r^2 . \qquad \text{(Lancaster [3])} \tag{4.3}$$

In this section we consider those bivariate distributions which are generated from three independent, additive variates. Thus let W_1, W_2 and W_3 be independently distributed and additive. Take $X = W_1 + W_2$, $Y = W_2 + W_3$. Then X and Y come from the same family as W_1, W_2 and W_3 but they are correlated. Their coefficient of correlation is

$$\rho = \operatorname{var} W_2/[(\operatorname{var}(W_1 + W_2)\operatorname{var}(W_2 + W_3))]^{\frac{1}{2}}. \tag{4.4}$$

Because of the particular way of obtaining these bivariate distributions, the coefficient of correlation is a satisfactory measure of dependence, as it is in the case of the bivariate normal. That is, $\rho = 0$ if and only if X and Y are independent, and $\rho = 1$ if and only if X and Y are linearly dependent.

THEOREM 4.1. *If, for a distribution we have:*

(i) *the orthogonal polynomials are generated by a function of the form* $f(t)e^{xu(t)}$,

(ii) *the distribution is additive and*

(iii) *a bivariate distribution is generated by using the additive property,*

then the matrix of correlations of the pairs of orthonormal polynomials on the marginals is diagonal. Further $\rho_{rr} = \rho_r$ *depends only on the normalising factor of the* rth *orthogonal polynomial.*

PROOF. We know, Lancaster [4], that a bivariate distribution is completely characterised a.e. by its marginal distributions and the matrix of correlations of any pair of complete sets of orthonormal functions on the marginal distributions. By Theorem 3.2, the systems of orthogonal polynomials are complete on the marginals. Thus the bivariate distribution is completely determined by the distributions of x and y and $\rho_{rs} = E(P_r^*(x)P_s^*(y))$. But

$$[c_r^{(X)}c_s^{(Y)}]^{\frac{1}{2}}\rho_{rs} = \iiint \left(\sum_{i=0}^{r} \binom{r}{i} P_i(w_1)P_{r-i}(w_2)\right)\left(\sum_{j=0}^{s} \binom{s}{j} P_j(w_2)P_{s-j}(w_3)\right) \times d\psi_1(w_1)\, d\psi_2(w_2)\, d\psi_3(w_3)$$

by Theorem 3.1, where $\psi_i(w_i)$ is the distribution function for W_i, $i = 1, 2, 3$. That is,

$$\rho_{rs} = \delta_{rs}c_r^{(2)}/[c_r^{(X)}c_s^{(Y)}]^{\frac{1}{2}}, \tag{4.5}$$

which proves the theorem.

Thus, if the bivariate distribution is ϕ^2 bounded, its frequency function is a.e. equal to the series

$$d\psi_x(x)\, d\psi_y(y) \left\{1 + \sum_{r=1}^{\infty} \rho_r P_r^*(x) P_r^*(y)\right\}, \tag{4.6}$$ and

$$\phi^2 = \sum_{r=1}^{\infty} \rho_r^2 . \tag{4.7}$$

COROLLARY. *For the above bivariate distributions the regression of $P_i^*(x)$ on $P_i^*(y)$ is linear, for all i.*

PROOF. This is an immediate consequence of the form of the canonical variables.

The above method, however, is not the only way of obtaining bivariate distributions. In fact, there exist an infinity of bivariate distributions with the same marginal distributions.

5. Examples. If $1 - \lambda t - \kappa t^2 = (1 - \alpha t)(1 - \beta t)$ then the form of α and β determines the type of weight distribution. Meixner shows that there only exist five types of distributions with the given form of generating function. We determine the canonical correlations of the corresponding bivariate distribution for each of these five types.

Type I $\qquad \alpha = \beta = \lambda = \kappa = 0.$

This is the normal distribution with

$$d\psi(x) = (2\pi\sigma)^{-\frac{1}{2}} \exp[-\tfrac{1}{2}x^2/\sigma^2]\, dx, \qquad -\infty < x < \infty.$$

The generating function for the orthogonal polynomials is

$$G(t, x) = \exp[xt - \tfrac{1}{2}\sigma^2 t^2], \qquad k_2 = -\sigma^2$$

and the normalising constants are $c_i = (\sigma^2)^i i!$. Now $\rho = \operatorname{corr}(X, Y) = \sigma_2^2/[(\sigma_1^2 + \sigma_2^2)(\sigma_2^2 + \sigma_3^2)]^{\frac{1}{2}}$, and the rth canonical correlation is

$$\rho_r = c_r^{(2)}/[c_r^{(X)} c_r^{(Y)}]^{\frac{1}{2}} = \rho^r. \tag{5.1}$$

The canonical expansion is just the Mehler expansion of the bivariate normal.

Type II $\qquad \alpha = \beta \neq 0.$

This is a generalised gamma distribution with

$$d\psi(x) \propto (-x - k_2/\alpha)^{-1-k_2/\alpha^2} \exp[x/\alpha], \qquad -\infty < x < -k_2/\alpha,$$

where $\alpha > 0$. If we let $k_2 = -p\alpha^2$, then $Y = -X/\alpha$ is a gamma variable with parameter p.

The generating function for the orthogonal polynomials is

$$G(t, x) = (1 - \alpha t)^{k_2/\alpha^2} \exp[k_2 t/\alpha(1 - \alpha t) + xt/(1 - \alpha t)]$$

and the normalising constants are $c_i = i!\, \alpha^{2i}\Gamma(p + i)/\Gamma(p)$. If $X/\alpha = W_1 + W_2$, and $Y/\alpha = W_2 + W_3$, where W_i is a gamma variable with parameter a_i, then

$$\begin{aligned}(5.2)\quad \rho_r = {} & [\Gamma(a_2 + r)/\Gamma(a_2)][\Gamma(a_1 + a_2) \\ & \cdot\Gamma(a_2 + a_3)/\Gamma(a_1 + a_2 + r)\Gamma(a_2 + a_3 + r)]^{\frac{1}{2}} \leqq \rho_{r-1}\end{aligned}$$

with equality only in the degenerate case $a_1 = a_3 = 0$.

Type III $\qquad \alpha \neq 0, \qquad \beta = \kappa = 0.$

This is a generalised Poisson distribution where $d\psi(x)$ is a step function with saltuses at the points $x_n = -k_2/\alpha - \alpha n$, $n = 0, 1, \cdots$ and

$$\psi(x_n + 0) - \psi(x_n - 0) = \exp[k_2/\alpha^2](-k_2/\alpha^2)^n/n!.$$

The generating function for the orthogonal polynomials is

$$G(t, x) = (1 - \alpha t)^{-k_2/\alpha^2 - x/\alpha} \exp[-k_2 t/\alpha]$$

and the normalising constants are $c_i = i!(-k_2)^i$. The rth canonical correlation is

$$(5.3) \qquad \rho_r = \rho^r.$$

Type IV $\qquad \alpha \neq \beta, \qquad \kappa \neq 0, \qquad$ both α and β real.

Without loss of generality, let $|\alpha| > |\beta|$. For $\alpha\beta > 0$, this is a negative binomial distribution where $d\psi(x)$ is a step function with saltuses at the points $x_n = -k_2/\alpha - (\alpha - \beta)n$, $n = 0, 1, \cdots$ and if $k_2/\alpha\beta = -a$

$$\psi(x_n + 0) - \psi(x_n - 0) = (1 - \beta/\alpha)^a(-\beta/\alpha)^n \binom{-a}{n}.$$

The generating function for the orthogonal polynomials is given by

$$(\alpha - \beta) \log G(t, x) = (k_2/\beta + x) \log(1 - \beta t) - (k_2/\alpha + x) \log(1 - \alpha t)$$

and the normalising constants are $c_i = \alpha^i\beta^i i!\Gamma(a + i)/\Gamma(a)$. The rth canonical correlation is

$$\begin{aligned}(5.4)\quad \rho_r = {} & [\Gamma(a_2 + r)/\Gamma(a_2)] \\ & \cdot[\Gamma(a_1 + a_2)\Gamma(a_2 + a_3)/\Gamma(a_1 + a_2 + r)\Gamma(a_2 + a_3 + r)]^{\frac{1}{2}}.\end{aligned}$$

For $\alpha\beta < 0$ and $k_2/\alpha\beta$ an integer, this distribution is the positive binomial (see Gonin [2]). Thus we can obtain for a bivariate binomial, whose marginals have the same parameter p, the canonical form in which

$$\begin{aligned}(5.5)\quad \rho_r = {} & [\Gamma(a_2 + 1)/\Gamma(a_2 - r + 1)] \\ & \cdot[\Gamma(a_1 + a_2 - r + 1)\Gamma(a_2 + a_3 - r + 1)/\Gamma(a_1 + a_2 + 1)\Gamma(a_2 + a_3 + 1)]^{\frac{1}{2}},\end{aligned}$$

where now $a = k_2/\alpha\beta$.

Type V $\quad \alpha \neq \beta, \qquad \kappa \neq 0, \qquad$ both α and β complex conjugates.

Without loss of generality, let $\mathscr{I}(\alpha) > \mathscr{I}(\beta)$. This is a hypergeometric distribution with

$d\psi(x) \propto (-\beta/\alpha)^{x/(\beta-\alpha)}\cdot\Gamma((\beta x + k_2)/\beta(\beta - \alpha))\cdot\Gamma((\alpha x + k_2)/\alpha(\alpha - \beta))$, $-\infty < x < \infty$ and where $|\arg(-\beta/\alpha)| < \Pi$.

The generating function of the orthogonal polynomials is the same as that of Type IV. The normalising constants are $c_i = (\alpha\beta)^i i!\Gamma(a + i)/\Gamma(a)$, where $a = -k_2/\alpha\beta$. The rth canonical correlation is

$$\rho_r = [\Gamma(a_2 + r)/\Gamma(a_2)] \cdot[\Gamma(a_1 + a_2)\Gamma(a_2 + a_3)/\Gamma(a_1 + a_2 + r)\Gamma(a_2 + a_3 + r)]^{\frac{1}{2}}. \tag{5.6}$$

ϕ^2 boundedness. The canonical correlations in the five kinds of bivariate distribution are of two types.

Type (i). $\rho_r = \rho^r$—the distribution is ϕ^2 bounded, provided $|\rho| \neq 1$.

Type (ii).

$$\rho_r = [\Gamma(a_2 + r)/\Gamma(a_2)][\Gamma(a_1 + a_2)\Gamma(a_2 + a_3)/\Gamma(a_1 + a_2 + r)\Gamma(a_2 + a_3 + r)].^{\frac{1}{2}}$$

(N.B. The expansion of the bivariate positive binomial is a terminating series and so the distribution is ϕ^2 bounded.) It is difficult to find necessary and sufficient conditions on the a_i in order that the distribution be ϕ^2 bounded. However, it is necessary that $|\rho| \neq 1$ (i.e. $a_1 > 0$, $a_3 > 0$) and by comparing $\sum_{r=1}^{\infty} \rho_r^2$ with $\sum_{r=1}^{\infty} r^{-(1+\delta)}$ $(\delta > 0)$ it can be seen that $(a_1 + a_3) > 1$, $a_1 > 0$, $a_3 > 0$ is a sufficient condition for ϕ^2 boundedness.

Goodness of fit tests. As the canonical expansions of the above distributions are known and the orthonormal polynomials are easily calculated, a goodness of fit test can be constructed using the method of Lancaster [3] for the bivariate normal. The total χ^2 is partitioned into a sum of χ^2 due to the fitting of the marginal distributions, χ^2 due to the regression of the lower orthonormal polynomials in X on those of the same degree in Y and a residual χ^2. The residual χ^2 is due to the regression of the orthonormal polynomials in X on those of different degree in Y and may be tested for significance.

Acknowledgments. I wish to acknowledge the help and guidance of Professor H. O. Lancaster in this work.

Note added in proof. The condition $(a_1 + a_3) > 1$ is both a necessary and sufficient condition for ϕ^2 boundedness.

REFERENCES

[1] CHERIAN, K. C. (1941). A bivariate correlated gamma type distribution function. *J. Indian Math. Soc.* **5** 133–144.

[2] GONIN, H. T. (1961). The use of orthogonal polynomials of the positive and negative binomial frequency functions in curve fitting by Aitken's method. *Biometrika* **48** 115–123.

[3] LANCASTER, H. O. (1958). The structure of bivariate distributions. *Ann. Math. Statist.* **29** 719–736.

[4] LANCASTER, H. O. (1963). Correlations and canonical forms of bivariate distributions. *Ann. Math. Statist.* **34** 532–538.

[5] Meixner, J. (1934). Orthogonale Polynomsysteme mit einer besonderen Gestalt der erzeugenden Funktion. *J. London Math. Soc.* **9** 6–13.

[6] Pearson, K. (1896). Mathematical contributions to the theory of evolution. III. Regression, heredity and panmixia. *Philos. Trans. Roy. Soc. London Ser. A* **187** 253–318.

[7] Runge, C. (1914). Über eine besondere Art von Integralgleichungen. *Math. Ann.* **75** 130–132.

[8] Sansone, G. (1959). *Orthogonal Functions.* Interscience, New York.