CALCULATING BIVARIATE ORTHONORMAL POLYNOMIALS BY RECURRENCE

Running Title: Bivariate Orthonormal Polynomials

BY J. C. W. RAYNER[1,2], O. THAS[3*], PETER PIPELERS[3] AND ERIC J. BEH[2]

Universities of Wollongong, Newcastle and Ghent

Summary

Emerson gave recurrence formulae for the calculation of orthonormal polynomials for univariate discrete random variables. He claimed that as these were based on the Christoffel-Darboux recurrence relation they were more efficient than those based on the Gram-Schmidt method. This approach was generalised by Rayner and colleagues to arbitrary univariate random variables. The only constraint was that the expectations needed are well-defined. Here the approach is extended to arbitrary bivariate random variables for which the expectations needed are well-defined. The extension to multivariate random variables is clear.

[*] Author to whom correspondence should be addressed.
[1] Centre for Statistical and Survey Methodology, School of Mathematics and Applied Statistics, University of Wollongong, NSW 2522, Australia and
[2] School of Mathematical and Physical Sciences, University of Newcastle, NSW 2308, Australia
[3] Department of Applied Mathematics, Biometrics and Process Control, B-9000 Gent, Belgium.
e-mail: olivier.thas@Ugent.be

Key words: Categorical data analysis; copulas; Emerson polynomials; orthonormal polynomials; smooth tests of goodness of fit.

1. Introduction

Bivariate random variables (X, Y) have orthonormal polynomials $\{h_{r,s}(x, y)\}$ if and only if

$$\begin{aligned} \mathrm{E}\{h_{r,s}(X, Y)\, h_{u,v}(X, Y)\} &= 0 \text{ for } (r, s) \neq (u, v) \text{ and} \\ &= 1 \text{ for } (r, s) = (u, v). \end{aligned}$$

Here E denotes expectation with respect to the distribution of (X, Y) and it is assumed that the expectations concerned exist. The couple (r, s) uniquely identifies every polynomial in $\{h_{r,s}(x, y)\}$, although there are different conventions for doing so. Subsequently it is assumed that $h_{0,0}(x, y) = 1$ for all x and y.

Example 1. Independent random variables. Suppose that X and Y are independent random variables, that $\{p_r(x)\}$ is a set of orthonormal functions on the distribution of X, and that $\{q_s(y)\}$ is a set of orthonormal functions on the distribution of Y. It follows that $h_{r,s}(x, y) = p_r(x) \times q_s(y)$ defines a set of orthonormal functions on the product distribution of X and Y. For $\mathrm{E}[\{p_r(X)q_s(Y)\}\{p_u(X)q_v(Y)\}] = \mathrm{E}\{p_r(X)p_u(X)\}\ \mathrm{E}\{q_s(Y)q_v(Y)\} = 1$ for $(r, s) = (u, v)$ and zero otherwise. If $p_r(x)$ and $q_s(y)$ are polynomials then the bivariate polynomials of degree n are $p_n(x) \times q_0(y), p_{n-1}(x) \times q_1(y), \ldots, p_0(x) \times q_n(y)$.

Example 2. Bivariate normal random variables. Suppose (X, Y) is bivariate normal with means μ_X, μ_Y, variances σ_X^2 and σ_Y^2, and correlation ρ. The standardised variables have covariance matrix

$$\mathbf{\Sigma} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \text{ for which } \mathbf{\Sigma}^{-0.5} = \frac{1}{2}\begin{pmatrix} (1+\rho)^{-0.5} + (1-\rho)^{-0.5} & (1+\rho)^{-0.5} - (1-\rho)^{-0.5} \\ (1+\rho)^{-0.5} - (1-\rho)^{-0.5} & (1+\rho)^{-0.5} + (1-\rho)^{-0.5} \end{pmatrix}.$$

Now if $Z = (Z_1, Z_2)^T = \mathbf{\Sigma}^{-0.5}\ ((X - \mu_X)/\sigma_X, (Y - \mu_Y)/\sigma_Y)^T$ then Z_1 are Z_2 are independent standard normal random variables. The orthonormal polynomials for the standard normal are the Hermite polynomials. The set of orthonormal polynomials for Z is, by the preceding example, the product set of Hermite polynomials. In terms of the original (X, Y) variables this set defines a set of orthonormal polynomials for the correlated bivariate normal. Both this example and the previous one generalise readily to the multivariate case. The resulting set of orthonormal polynomials is no less compact than that described by Withers (2000).

Subsequently we use a different indexing convention from that used in the first two examples. A bivariate polynomial of degree n is of the form $\sum_{i,j} a_{i,j} x^i y^j$ with $i, j = 0, 1, 2, ...$, $i + j \leq n$, and with $a_{i,\, n-i} > 0$ for at least one i. Note that a *homogeneous* bivariate polynomial of degree n includes only those terms with $i + j = n$. Any bivariate polynomial of degree n could be considered to be a sum of homogeneous bivariate polynomials of degree 0, 1, ..., n. We say $h_{n,s}(x, y)$ is the sth bivariate orthonormal polynomial of degree n with $n = 0, 1, 2, ...$ and $s = 0, 1, ... , n$ if $h_{n,s}(x, y)$ is both a bivariate polynomial of degree n and $\{h_{n,s}(x, y)\}$ is orthonormal. Thus n indicates the degree and s is a unique counting index.

In section 2 we give recurrence formulae for the construction of orthonormal polynomials for bivariate distributions. At the nth step a basis of bivariate polynomials is constructed and then a linear transformation produces the orthonormal polynomials of degree n. These degree n orthonormal polynomials are not unique.

In section 3 the use of the bivariate orthonormal polynomials to construct smooth tests of goodness of fit is discussed. Emphasis is given to testing for copulas. The conclusion in section 4 focuses mainly on categorical data analysis.

At http://www.biomath.ugent.be/~othas/recurrence/ R code that implements the recurrence relations is given for three distributions. The first is essentially Example 3. The distributions on the second example are related to 3 × 3 and 4 × 4 contingency tables, and the final example produces orthonormal polynomials for a copula; see Section 3.2. The interested reader is invited to modify the code for examples of their choice. The code verifies that the functions are both orthogonal and normalised, and writes the orthonormal functions to R functions.

2. Bivariate recurrence formulae

Subsequently we will need vectors $\mathbf{d}_r$ containing functions of degree r in the variables x and y, then move to vectors $\mathbf{u}_r$ that contain the corresponding orthogonal polynomials, then vectors $\mathbf{h}_r$ that contain the corresponding orthonormal polynomials. The initial choices are sensible but just one set of possibilities; similarly the outcomes are not necessarily unique. First, given our definition of $h_{0,0}(x, y)$, we take $\mathbf{d}_0 = \mathbf{u}_0 = \mathbf{h}_0 = 1$.

Henceforth $\mathbf{d}_r$ will contain degree r terms in x and y in a manner to be defined shortly. We take $\mathbf{d}_1 = (x, y)^T$ and construct the degree one orthonormal polynomials. Incidentally, since $\mathbf{u}_r$ contains the orthogonal polynomials of degree r and $\mathbf{h}_r$ contains the orthonormal polynomials of degree r, $\mathbf{d}_r$, $\mathbf{u}_r$ and $\mathbf{h}_r$ are all $(r + 1) \times 1$ vectors. Also we write $\mathbf{M}_r = \text{cov}(\mathbf{u}_r)$, which is $(r + 1) \times (r + 1)$.

It will not be assumed that the random variables X and Y are standardised. We take $\mathbf{u}_1 = \mathbf{d}_1 - \text{E}(\mathbf{d}_1)$, $\mathbf{M}_1 = \text{cov}(\mathbf{u}_1)$ and $\mathbf{h}_1 = \mathbf{M}_1^{-0.5}\ \mathbf{u}_1$. Clearly $\mathbf{h}_1$ is orthogonal to $\mathbf{h}_0$ since $\text{E}(\mathbf{h}_1) = 0$ and $\text{cov}(\mathbf{h}_1) = \mathbf{I}_2$, the 2×2 identity matrix.

If $\mathbf{L}$ is any orthogonal matrix the elements of $\mathbf{Lh}_1$ are also degree one polynomials orthogonal to $h_{0,0}$ since $\text{E}(\mathbf{Lh}_1) = \mathbf{L}\text{E}(\mathbf{h}_1) = 0$ and mutually orthonormal since $\text{cov}(\mathbf{Lh}_1) = \mathbf{I}_2$. So the orthonormal polynomials of degree zero and one are not unique, and uniqueness cannot be expected for the orthonormal polynomials of any degree. This is consistent with the bivariate normal, since our orthogonal polynomials are different from those given by Withers (2000). However the elements of $\mathbf{h}_1$ are, in a sense, symmetric, and this symmetry of form is sensible.

We now construct the order two polynomials. Define $\mathbf{d}_2 = (x \times (\mathbf{h}_1)_1, x \times (\mathbf{h}_1)_2, y \times (\mathbf{h}_1)_2)^T$ and $\mathbf{u}_2 = \{\mathbf{d}_2 - \text{E}(\mathbf{d}_2)\} + \mathbf{A}_1\mathbf{h}_1$. For $\mathbf{u}_2$ to be orthogonal to $\mathbf{h}_0$ requires $0 = \text{E}(\mathbf{u}_2)$, which is obviously true since $\text{E}[\{\mathbf{d}_2 - \text{E}(\mathbf{d}_2)\}] = 0$ and $\text{E}(\mathbf{h}_1) = 0$. Orthogonality to $\mathbf{h}_1$ requires

$$0 = \text{cov}(\mathbf{u}_2, \mathbf{h}_1) = \text{cov}(\mathbf{d}_2, \mathbf{h}_1) + \mathbf{A}_1, \text{ so that } \mathbf{A}_1 = -\text{cov}(\mathbf{d}_2, \mathbf{h}_1).$$

Hence

$$\mathbf{u}_2 = \mathbf{d}_2 - \text{E}(\mathbf{d}_2) - \text{cov}(\mathbf{d}_2, \mathbf{h}_1)\mathbf{h}_1.$$

Now routine calculations show that $\mathbf{M}_2 = \text{var}(\mathbf{d}_2) - \text{cov}(\mathbf{d}_2, \mathbf{h}_1)\ \text{cov}(\mathbf{h}_1, \mathbf{d}_2)$ so $\mathbf{h}_2 = \mathbf{M}_2^{-0.5}\mathbf{u}_2$ is fully specified.

At the nth step we have constructed $\mathbf{h}_0, \mathbf{h}_1, \ldots, \mathbf{h}_{n-1}$ mutually orthonormal and define both $\mathbf{d}_n = (x \times (\mathbf{h}_{n-1})_1, x \times (\mathbf{h}_{n-1})_2, \ldots, x \times (\mathbf{h}_{n-1})_n, y \times (\mathbf{h}_{n-1})_n)^T$ and $\mathbf{u}_n = \{\mathbf{d}_n - \mathrm{E}(\mathbf{d}_n)\} + \mathbf{A}_{n-1}\mathbf{h}_{n-1} + \ldots + \mathbf{A}_1\mathbf{h}_1$. The term $\{\mathbf{d}_n - \mathrm{E}(\mathbf{d}_n)\}$ gives a vector of polynomials of degree n while the remaining terms allow us to complete the construction so that $\mathbf{u}_n$ is orthogonal to all previous orthonormal polynomials. By an argument parallel to that in Rayner et al. (2008), in $\mathbf{u}_n$ it is sufficient to include only terms involving $\mathbf{h}_{n-1}$ and $\mathbf{h}_{n-2}$. For

$$0 = \mathrm{cov}(\mathbf{u}_n, \mathbf{h}_j) = \mathrm{cov}(\mathbf{u}_n, \mathbf{d}_{\underline{n}}) + \mathbf{A}_j \text{ for } j \leq n - 3$$

and to evaluate $\mathrm{cov}(\mathbf{u}_n, \mathbf{h}_j)$ we need to evaluate terms like $\mathrm{E}\{X \times (\mathbf{h}_{n-1})_r \times (\mathbf{h}_j)_s\} = \mathrm{E}\{(\mathbf{h}_{n-1})_r \times X \times (\mathbf{h}_j)_s\}$. For $j \leq n - 3$ $X \times (\mathbf{h}_j)_s$ is expressible as a linear combination of elements of the basis vectors $\mathbf{h}_0, \mathbf{h}_1, \ldots, \mathbf{h}_{n-2}$, which must be orthogonal to $\mathbf{h}_{n-1}$. Thus $\mathbf{A}_j = 0$ and

$$\mathbf{u}_n = \{\mathbf{d}_n - \mathrm{E}(\mathbf{d}_n)\} + \mathbf{A}_{n-1}\mathbf{h}_{n-1} + \mathbf{A}_{n-2}\mathbf{h}_{n-2}.$$

To evaluate the $\mathbf{A}_j$ note that orthogonality requires $\mathrm{E}(\mathbf{u}_n) = 0$, which again is clearly true, and, for $j = n - 1$ and $n - 2$

$$0 = \mathrm{cov}(\mathbf{u}_n, \mathbf{h}_j) = \mathrm{cov}(\mathbf{d}_n, \mathbf{h}_j) + \mathbf{A}_j, \text{ so } \mathbf{A}_j = -\,\mathrm{cov}(\mathbf{d}_n, \mathbf{h}_j) \text{ for } j = n - 1 \text{ and } n - 2.$$

Hence

$$\mathbf{u}_n = \mathbf{d}_n - \mathrm{E}(\mathbf{d}_n) - \mathrm{cov}(\mathbf{d}_n, \mathbf{h}_{n-1})\, \mathbf{h}_{n-1} - \mathrm{cov}(\mathbf{d}_n, \mathbf{h}_{n-2})\, \mathbf{h}_{n-2}. \tag{1}$$

We could now calculate $\mathbf{M}_n = \mathrm{cov}(\mathbf{u}_n)$ numerically and put $\mathbf{h}_n = \boldsymbol{\Sigma}_n^{-0.5}\,\mathbf{u}_n$.

Example 3. Two independent standard normal random variables. From Examples 1 and 2 we know two sets of bivariate orthonormal polynomials independent of the recurrence developed above. We now apply the recurrence to this situation.

If Z_1 and Z_2 are independent standard normal random variables put $(Z_1, Z_2)^T = \mathbf{d}_1 = \mathbf{u}_1 = \mathbf{h}_1$ and $\mathbf{M}_1 = \mathbf{I}_2$. Applying our construction, $\mathbf{d}_2 = (Z_1^2, Z_1Z_2, Z_2^2)^T$ and $E(\mathbf{d}_2) = (1, 0, 1)^T$. Now using the independence and that Z_1 and Z_2 have zero means and third moments,

$$\mathrm{cov}(\mathbf{d}_2, \mathbf{h}_1) = \begin{pmatrix} \mathrm{cov}(Z_1^2, Z_1) & \mathrm{cov}(Z_1^2, Z_2) \\ \mathrm{cov}(Z_1Z_2, Z_1) & \mathrm{cov}(Z_1Z_2, Z_2) \\ \mathrm{cov}(Z_2^2, Z_1) & \mathrm{cov}(Z_2^2, Z_2) \end{pmatrix} = \begin{pmatrix} E[Z_1^3] & E[Z_1^2Z_2] \\ E[Z_1^2Z_2] & E[Z_1Z_2^2] \\ E[Z_1Z_2^2] & E[Z_2^3] \end{pmatrix} = \mathbf{0}.$$

Next, $\mathbf{u}_2 = (Z_1^2 - 1, Z_1Z_2, Z_2^2 - 1)^T$ from which $\mathbf{M}_2 = \mathbf{\Sigma}_{22} = \mathrm{var}(\mathbf{d}_2) = \mathrm{diag}(2, 1, 2)$. Hence

$$\mathbf{h}_2 = ((Z_1^2 - 1)2^{-0.5}, Z_1Z_2, (Z_2^2 - 1)2^{-0.5})^T = (g_2(Z_1), g_1(Z_1)g_1(Z_2), g_2(Z_2))^T$$

if we write $g_r(z)$ for the Hermite polynomial of order r. This is consistent with Example 1.

At the next step we apply our construction again. This gives

$$\mathbf{d}_3 = (Z_1g_2(Z_1), Z_1g_1(Z_1)g_1(Z_2), Z_1g_2(Z_2), Z_2g_2(Z_2))^T.$$

Since, for example, $Z_1g_2(Z_1) = Z_1^3 - Z_1$, $E\{Z_1g_2(Z_1)\} = 0$ and $E(\mathbf{d}_3) = \mathbf{0}$. We could also use arguments about odd and even functions, as the odd order Hermite polynomials are odd functions, and both the even odd order Hermite polynomials and the standard normal pdf are even functions. We now require cov($\mathbf{d}_3$, $\mathbf{h}_2$) and cov($\mathbf{d}_3$, $\mathbf{h}_1$). First, using arguments about odd and even functions, cov($\mathbf{d}_3$, $\mathbf{h}_2$) = 0. The details are omitted here. Next

$$\text{cov}(\mathbf{d}_3, \mathbf{h}_1) = \begin{pmatrix} \text{cov}(Z_1 g_2(Z_1), Z_1) & \text{cov}(Z_1 g_2(Z_1), Z_2) \\ \text{cov}(Z_1 g_1(Z_1) g_1(Z_2), Z_1) & \text{cov}(Z_1 g_1(Z_1) g_1(Z_2), Z_2) \\ \text{cov}(Z_1 g_2(Z_2), Z_1) & \text{cov}(Z_1 g_2(Z_2), Z_2) \\ \text{cov}(Z_2 g_2(Z_2), Z_1) & \text{cov}(Z_2 g_2(Z_2), Z_2) \end{pmatrix}$$

$$= \begin{pmatrix} \mathrm{E}\{Z_1^2 g_2(Z_1)\} & \mathrm{E}\{Z_1 g_2(Z_1)\}E(Z_2) \\ \mathrm{E}\{Z_1^2 g_1(Z_1)\}E\{g_1(Z_2)\} & \mathrm{E}\{Z_1 g_1(Z_1)\}E\{Z_2 g_1(Z_2)\} \\ \mathrm{E}(Z_1^2)E\{g_2(Z_2)\} & \mathrm{E}(Z_1)E\{Z_2 g_2(Z_2)\} \\ \mathrm{E}(Z_1)E\{Z_2 g_2(Z_2)\} & \mathrm{E}\{Z_2^2 g_2(Z_2)\} \end{pmatrix} = \begin{pmatrix} 2^{0.5} & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 2^{0.5} \end{pmatrix}$$

since, for example, $g_2(z_1) = (z_1^2 - 1)2^{0.5}$, $\mathrm{E}\{Z_1^2 g_2(Z_1)\} = \{E(Z_1^4) - E(Z_1^2)\}2^{0.5} = 2^{0.5}$ and $\mathrm{E}\{Z_1 g_1(Z_1)\} = E(Z_1^2) = 1$. It is routine to show that E($\mathbf{d}_3$) = 0 so that

$$\mathbf{u}_3 = \begin{pmatrix} z_1 g_1(z_1) \\ z_1 g_1(z_1) g_1(z_2) \\ z_1 g_2(z_2) \\ z_2 g_2(z_2) \end{pmatrix} - \begin{pmatrix} z_1 2^{0.5} \\ z_2 \\ 0 \\ z_2 2^{0.5} \end{pmatrix} = \begin{pmatrix} g_3(z_1) 3^{0.5} \\ g_2(z_1) g_1(z_2) 2^{0.5} \\ g_1(z_1) g_2(z_2) 2^{0.5} \\ g_3(z_2) 3^{0.5} \end{pmatrix}.$$

This uses the functional forms of $g_1(z)$, $g_2(z)$ and $g_3(z)$. Again it is routine to show that $\mathbf{M}_3$ = diag(3, 2, 2, 3) and $\mathbf{h}_3$ is as anticipated in Example 2.

We now consider the order n polynomials. The construction puts

$$\mathbf{h}_n = (z_1 g_{n-1}(z_1), z_1 g_{n-2}(z_1)\, g_1(z_2), z_1 g_{n-3}(z_1) g_2(z_2), \ldots, z_1 g_{n-1}(z_2), z_2 g_{n-1}(z_2))^T.$$

We have from previously that $E(\mathbf{d}_1) = 0$, $E(\mathbf{d}_2) = (1, 0, 1)^T$ and $E(\mathbf{d}_3) = 0$. Again using the independence and arguments about odd and even functions, all elements of $E(\mathbf{d}_n)$ apart from the first and last are zero. From Rayner et al. (2008), quoting Abramowitz & Stegun (1970, Chapter 22), the unnormed Hermite polynomials $\{He_r(x)\}$ satisfy $He_r(x) = x\, He_{r-1}(x) - (r-1)\, He_{r-2}(x)$ from which $E\{X\, He_{r-1}(X)\} = E\{He_r(X)\} + (r-1)\, E\{He_{r-2}(X)\}$. Only when $r = 2$ is $E\{X\, He_{r-1}(X)\}$ non-zero. It follows that $E(\mathbf{d}_n) = 0$ for $n \geq 2$.

Yet again using the independence and arguments about odd and even functions, $\text{cov}(\mathbf{d}_n, \mathbf{h}_{n-1}) = 0$. Turning to $\text{cov}(\mathbf{d}_n, \mathbf{h}_{n-2})$, the first row contains elements

$$\text{cov}(z_1 g_{n-1}(z_1), g_{n-2}(z_1)),\ \text{cov}(z_1 g_{n-1}(z_1), g_{n-3}(z_1)\, g_1(z_2)),\ \ldots$$

all of which, apart from the first element, are zero. Using a result from Rayner et al. (2008), this element is

$$z_1 g_{n-1}(z_1) - E\{Z_1 g_{n-1}(Z_1) g_{n-2}(Z_1)\} g_{n-2}(z_1) = z_1 g_{n-1}(z_1) - (n-1)^{0.5}\, g_{n-2}(z_1).$$

Again from Rayner et al. (2008), quoting Abramowitz & Stegun (1970, Chapter 22), the recurrence relation for the unnormed Hermite polynomials when normed gives

$$z g_{r-1}(z) - (r-1)^{0.5}\, g_{r-2}(z) = r^{0.5}\, g_r(z).$$

Using this result the first element is $n^{0.5}\, g_r(z_1)$.

The second row contains elements

$$\text{cov}\{z_1 g_{n-2}(z_1)\, g_1(z_2), g_{n-2}(z_1)\},\ \text{cov}\{z_1 g_{n-2}(z_1) g_1(z_2), g_{n-3}(z_1) g_1(z_2)\},\ \ldots$$

all of which, apart from the second element, are zero. Using the previous approach, this element is

$$z_1 g_{n-2}(z_1)\, g_1(z_2) - \mathrm{E}\{Z_1 g_{n-2}(Z_1) g_{n-3}(Z_1)\} g_{n-3}(z_1) g_1(z_2)$$

$$= \{z_1 g_{n-2}(z_1) - (n-2)^{0.5}\, g_{n-3}(z_1)\}\, g_1(z_2) = (n-2)^{0.5}\, g_{n-1}(z_1)\, g_1(z_2).$$

Subsequent terms yield the results anticipated in Example 2.

3. Application: smooth tests of goodness of fit for bivariate distributions

3.1 Smooth tests for bivariate distributions

Suppose that we have a random sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ from a bivariate distribution. We wish to test if the distribution of (X, Y) has pdf $f(x, y; \beta)$, in which β is a vector of p nuisance parameters, such as the means, variances and covariances when testing for bivariate normality. The construction of the smooth tests starts by imbedding $f(x, y; \beta)$ in a smooth alternative of order k ($> p$). Define Neyman alternatives of the form

$$g_k(x, y;\, \theta, \beta) = C(\theta, \beta) \exp\{\sum_{(i,j)\in I} \theta_{i,j} h_{i,j}(x, y; \beta)\} f(x, y; \beta)$$

in which $\{h_{i,j}(x, y; \beta)\}$ is a set of functions orthonormal on $f(x, y; \beta)$ and $C(\theta, \beta)$ is a normalising constant and I is an index set containing k unique pairs $(i, j) \neq (0, 0)$. Here θ is a vector containing the k parameters $\theta_{i,j}$. The $\theta_{i,j}$ are inserted into θ in some well-defined order, such as, if appropriate, lexicographic order. Similarly define Barton alternatives of the form

$$g_k(x, y;\, \theta, \beta) = \{1 + \sum_{(i,j)\in I} \theta_{i,j} h_{i,j}(x, y; \beta)\} f(x, y; \beta)\,.$$

The merits of both forms, and more, are discussed in Rayner et al. (2009). Here we simply note that the Barton form does not require the existence of the normalising constant, but may not be a proper pdf.

To test for $f(x, y; \beta)$ the score test statistic for testing H: $\theta = 0$ against K: $\theta \neq 0$ is derived. Put

$$V_{r,s} = V_{r,s}(\beta) = n^{-0.5} \sum_{j=1}^{n} h_{r,s}(X_j, Y_j; \beta)$$

and define $V = (V_{i,j})$ using the same pairs $(i, j) \in I$ and ordering as for θ. The choice of $\theta_{i,j}$ in $g_k(x, y;\, \theta, \beta)$ depends on the purposes of the analysis; see Example 4. Again using the same ordering, put

$$\Sigma(\beta) = I_k - \mathrm{cov}_0(V(\beta), \partial \log f / \partial\beta)\, \mathrm{var}_0^{-1}(\partial \log f / \partial\beta)\, \mathrm{cov}_0(\partial \log f / \partial\beta, V(\beta))\,.$$

Here the subscript 0 indicates the variance or covariance is evaluated under the null hypothesis $\theta = 0$ and I_k is the $k \times k$ identity matrix. The same test statistic,

$$S_k = V^T(\hat{\beta})\, \Sigma^{-1}(\hat{\beta})\, V^T(\hat{\beta})$$

results when using either the Neyman or the Barton form. Here $\hat{\beta}$ is the maximum likelihood estimator of β under the null hypothesis; that is, when the true pdf is $f(x, y; \beta)$. The asymptotic null distribution of S_k is χ^2_{k-p}, where the degrees of freedom are the number of components or (r, s) pairs minus the number of parameters estimated.

In Rayner et al. (2009) the smooth tests have been used to test for a variety of distributions, such as the normal, Poisson, logistic and generalised Pareto distributions. Most distributions examined are univariate; only two multivariate distributions are assessed. One is the multivariate normal where the orthonormal polynomials are constructed by transforming to independent standard normals and using the product set of polynomials, as in Examples 1 and 3. For the bivariate Poisson smooth tests were used, but not orthonormal polynomials. This approach leads to a better understanding and improvement of certain known tests in the literature.

The orthonormal functions given in this paper mean we may now construct smooth tests for a large class of bivariate distributions not previously accessible. A particular class of such distributions is copulas, discussed in the next sub section.

3.2 Copulas

Copulas are a general class of multivariate distributions that are characterised by having uniform marginal distributions. All parameters of the copula are thus related to the bivariate moments. Because a uniform distribution can be transformed to any continuous distribution by applying the quantile function transformation, copulas may be used to model a very wide class of multivariate distributions in a very flexible manner. Among the many applications of copulas, we mention financial risk assessment, hydrology and survival analysis; see, for example, Embrechts (2009), Genest & Favre (2007) and Salvadori & De Michele (2007). For an introduction to copulas see Nelson (2006).

For a review of other types of goodness of fit tests for copulas see Berg (2009) and Genest (2009).

Example 4. The Farlie-Gumbel-Mortgenstern (FGM) copula. Copulas are typically defined in terms of their cdf. The FGM copula has cdf

$$F(x, y, \beta) = xy + \beta xy(1 - x)(1 - y) \text{ for } (x, y) \in [0, 1]^2 \text{ and } \beta \in [-1, 1].$$

The score function is thus

$$\partial \log f(x, y, \beta)/\partial\beta = (1 - 2x)(1 - 2y)/\{1 + \beta(1 - 2x)(1 - 2y)\}.$$

The test is performed as follows.

1) The marginal empirical distribution function is used to transform the marginals to be uniformly distributed. The orthonormal polynomials are then calculated as described in section 2.
2) Maximum likelihood is used to estimate the nuisance parameter β from the sample data, for example by using the copula package in R. The estimate is denoted by $\hat{\beta}$.
3) The component statistics $V_{r,s}(\hat{\beta})$ are computed for all $(r, s) \in I$.
4) The variance estimate $\text{var}_0(\partial \log f / \partial\beta)\big|_{\beta=\hat{\beta}}$ is obtained from the copula package in R.

5) The covariance estimate $\text{cov}_0(V(\beta), \partial \log f / \partial \beta)\big|_{\beta=\hat{\beta}}$ is numerically approximated based on a random sample of 1,000,000 data points from a FGM copula with nuisance parameter $\hat{\beta}$.

6) The test statistic S is calculated, and the p-value obtained from a χ^2 distribution.

Now consider the data of Stone (1978) in Table 1. The data comes from an experiment with items subjected to a 55-voltage stress; two correlated failure times, (T_1, T_2) are recorded. Craiu & Bercia (2007) analysed the data and concluded that of the several copula distributions considered the FGM copula gave the best fit to the data.

TABLE 1 ABOUT HERE

Applying our algorithm, the parameter β is estimated by maximum likelihood to be 0.8512. Using the nine (r, s) pairs with r and s = 1, 2, 3 gives a test statistic of 7.06. Using the asymptotic χ_8^2 distribution gives a p-value of 0.530. Thus at the 5% level the FGM copula is consistent with the data.

Another smooth test is of interest. In many copula applications the fit to the marginal distributions is separated from the assessment of the dependence structure. Thus, for example, the dependence structure may be assessed by considering only those polynomials that have only terms in $x^i y^j$ with $i, j \neq 0$. This may be achieved by means of an appropriate orthonormal transformation of the original set of orthonormal polynomials. In the R-code available on our web site more details can be found about the QR-decomposition that we have used for this purpose. After the transformation only three polynomials are of the desired form. The resulting test statistic takes the value 0.728 with asymptotic χ_3^2 p-value of 0.695. Again, at the 5% level, in regard to the dependence structure as assessed by these three components, the FGM copula is consistent with the data.

These smooth tests confirm the descriptive findings of Craiu & Bercia (2007).

4. Conclusion

The polynomials generated from the recurrence formulae above expand the scope for further work in many areas; here we comment briefly on categorical data analysis. Previous work has largely focused on analysing the association between categorical variables, typically univariate orthonormal polynomials derived using the recurrence formulae of Emerson (1968). For example, the partition of the Pearson chi-squared statistic using univariate polynomials has been considered by Lancaster (1953), Rayner & Best (1996), Best & Rayner (1996), Beh & Davy (1998, 1999), Beh (2001) and Rayner & Beh (2009b). Partitions of other measures of association using the univariate polynomials derived from Emerson's formulae include those of Lombardo, Beh & D'Ambra (2007), Beh et al. (2007). The graphical analysis of association between categorical variables has extensively used univariate polynomials. One may consider, for example, Beh (1997, 1998, 2008) and Lombardo et al. (2007), who focused on correspondence analysis.

There is great potential for the bivariate polynomials constructed as in this paper to be applied in this same context. The advantage of considering their use in a categorical data analysis framework is that they can allow the researcher to delve deeper into the association structure of the variables of interest by identifying generalised correlations; see Rayner & Beh (2009a). By determining bivariate polynomials of order (r, s) one can identify association structures that reflect the (r, s)th component. For example, if X and Y are defined as two categorical variables, the bivariate polynomial $h_{12}(X, Y)$, can be used to reflect the linear-by-quadratic association structure between X and Y. This certainly has potential when performing ordinal correspondence analysis (Beh, 1997) which will allow for a graphical perspective of such a structure.

The benefit of implementing the bivariate polynomials in a categorical data analysis framework is that they will allow the researcher to investigate a variety of complex association structures that exist between multiple categorical variables. For example, the univariate polynomials have been shown to be of value in the partition of the Goodman-Kruskal tau (GK-tau) index (Goodman & Kruskal, 1954) and the Marcotorchino index, a three-way analogue of GK-tau; see Marcotorchino (1985). The components of such partitions are akin to the generalised correlation already alluded to. However, where the univariate polynomials are unable to take into consideration the association structure required for calculating the Gray-Williams index (Gray & Williams, 1975), say, the bivariate polynomials will allow for such calculations to be made.

References

ABRAMOWITZ, M. & STEGUN, I.A. (1970). *Handbook of Statistical Functions*. New York: Dover.

BEH, E. J. (1997). Simple Correspondence analysis of ordinal cross-classifications using orthogonal polynomials. *Biom. J.*, 39, 589-613.

BEH, E. J. (1998). A comparative study of scores for correspondence analysis with ordered categories. *Biom. J.*, 40, 413-429.

BEH, E. J. (2001). Partitioning Pearson's chi-squared statistic for singly ordered two-way contingency tables. *Aust. N.Z. J. Stat.*, 43, 327-333.

BEH, E. J. (2008). Simple correspondence analysis of nominal-ordinal contingency tables. *J. App. Math. Decis. Sci.*, Volume 2008, Article ID 218140, 17 pages.

BEH, E. J. & DAVY, P. J. (1998). Partitioning Pearson's chi-squared statistic for a completely ordered three-way contingency table. *Aust. N.Z. J. Stat.*, 40, 465-477.

BEH, E. J. & DAVY, P. J. (1999). Partitioning Pearson's chi-squared statistic for a partially ordered three-way contingency table. *Aust. N.Z. J. Stat.*, 41, 233-246.

BEH, E. J., SIMONETTI, B. & D'AMBRA, L. (2007). Partitioning a non-symmetric measure of association for three-way contingency tables. *J. Multivariate Anal.*, 98, 1391 – 1411.

BERG, D. (2009). Copula goodness-of-fit testing: an overview and power comparison. *Eur. J. Finance*, 15, 675-701.

BEST, D. J. & RAYNER, J. C. W. (1996). Nonparametric analysis for doubly ordered two-way contingency tables. *Biometrics*, 52, 1153 – 1156.

CRAIU, Mariana & BERCIA, Romeo (2007). A bivariate copula model for the failure of electrical cable insulation. *Proceedings of the 4th international collogium on mathematics in engineering and numerical physics (MENP-4)*, Bucharest, Romania, October 6-8, 2006. Geometry Balkan Press. BSG Proceedings 14, 52-57.

EMBRECHTS, P. (2009). Copulas: a personal view. *J. Risk & Insurance*, 76, 639-650.

EMERSON, P.L. (1968). Numerical construction of orthogonal polynomials from a general recurrence formula. *Biometrics*, 24, 695–701.

GENEST, C. (2009). Insurance. *J. Math. Econom*., 44, 199-213.

GENEST, C. & FAVRE, A.C. (2007). Everything you always wanted to know about copula modelling but were afraid to ask. *J. Hydrologic Engineering*, 12, 347-368.

GOODMAN, L. & KRUSKAL, W. (1954). Measures of association for cross-classifications. *J. Amer. Statist. Assoc.*, 49, 732–764.

GRAY, L.N. & WILLIAMS, J.S. (1975). Goodman and Kruskal's tau b: multiple and partial analogs, in *Proceedings of the Social Statistics Section, Amer. Statist. Assoc*, 444 – 448.

LANCASTER, H. O. (1953), A reconstitution of X^2 considered from metrical and enumerative aspects. *Sankhya*, 13, 1 – 10.

LOMBARDO, R., BEH, E. J. & D'AMBRA, L. (2007). Non-symmetric correspondence analysis with ordinal variables. *Comput. Statist. Data Anal*. 52, 566-577.

MARCOTORCHINO, F. (1985). Utilisation des Comparaisons par Paires en Statistique des Contingencies: Partie III, Report # F 081, Etude du Centre Scientifique, IBM, France.

NELSON, R.B. (2006). *An Introduction to Copulas*. New York: Springer.

RAYNER, J.C.W. & BEH, E. J. (2009a). Towards a Better Understanding of Correlation. *Stat. Neerl.*. 63, 324-333.

RAYNER, J. C. W. & BEH, E. J. (2009b). Components of Pearson's statistic for at least partially ordered *m*-way contingency tables. *J. App. Math. Decis. Sci.*, Volume 2009, Article ID 980706, 9 pages.

RAYNER, J. C. W. & BEST, D. J. (1996). Smooth extensions of Pearson's product moment correlation and Spearman's Rho. *Stat. Probab. Lett.*, 30, 171 – 177.

RAYNER, J.C.W., THAS, O. & BEST, D.J. (2009). *Smooth Tests of Goodness of Fit: Using R* (2^{nd} ed.). Singapore: Wiley.

RAYNER, J.C.W., THAS, O. & De BOECK, B. (2008). A generalised Emerson recurrence relation. *Aust. N.Z. J. Stat.*, 50, 235-240.

SALVADORI, G. & De MICHELE, C. (2007). On the use of copulas in hydrology: theory and practice. *J. Hydrologic Engineering*, 12, 369-380.

STONE, G.C. (1978). *Statistical Analysis of Accelereated Aging Tests on Solid Electrical Insulation*. Unpublished M.A.Sc. Thesis, University of Waterloo, Canada.

WITHERS, C.S. (2000). A simple expression for the multivariate Hermite polynomials. *Stat. Probab Lett.*, 47, 165-169.

TABLE 1

Stone's correlated failure time data

T_1	228	106	246	700	473	155	414	1374	128
T_2	30	8	66	72	25	7	30	90	4
T_1	1227	254	435	1155	195	117	724	300	
T_2	39	46	85	85	27	27	21	96	