Statistica Sinica **7**(1997), 713-738

NON-LINEAR INTEGRAL EQUATIONS TO APPROXIMATE BIVARIATE DENSITIES WITH GIVEN MARGINALS AND DEPENDENCE FUNCTION

Geert Molenberghs and Emmanuel Lesaffre

Limburgs Universitair Centrum and Katholieke Universiteit Leuven

Abstract: The local dependence function, introduced by Holland and Wang (1987) and studied by Wang (1993) as a continuous version of the local cross-ratio, describes the local relation between two random variables. Three explicit numerical algorithms are proposed to approximate bivariate densities given the marginal densities and the local dependence function. This approach is suited for simulation purposes, to provide illustrative examples of densities with given marginals, and for estimation of model parameters. The technique involves non-classical integral equation theory. The accuracy of the approximations is investigated.

Key words and phrases: Bivariate density, integral equation, local cross-ratio, local dependence function, numerical integration, Plackett family.

1. Introduction

The construction of bivariate distributions with specified marginals has been discussed by several authors, e.g. Morgenstern (1956), Plackett (1965), Johnson (1987), Johnson and Tenenbein (1981) and Marshall and Olkin (1988). These methods concentrate on fixing the marginal distributions. However, the association structure is either indirectly modelled (Morgenstern family, Edgeworth expansion) or difficult to interpret.

Recently, Holland and Wang (1987) and Wang (1993) introduced the *local dependence function* of a bivariate density. It is the rate of change of the local cross-ratio and is defined as the limit of the local cross-ratio defined for adjacent cell probabilities, formed by a two-dimensional rectangular grid, when length and width of the rectangles shrink to zero. They show that it equals the mixed derivative of the natural logarithm of the density function. It completely characterizes the local behaviour of the density function. Holland and Wang (1987) show that, under mild regularity conditions, specification of the bivariate density is equivalent to the specification of the marginal densities, together with the local dependence function. Wang (1993) suggested a theoretical method to determine a bivariate density function, given its marginal densities and its local dependence function: the *iterative marginal replacement algorithm*.

In this paper, the work of Holland and Wang (1987) and Wang (1993) is extended by proposing numerical algorithms to approximate the bivariate density,

given the marginals and the local dependence function. The algorithms are based on the solution of non-linear integral equations. Numerical techniques, used to solve a classical eigenvalue problem (Baker (1977)) are the basis for constructing iterative methods to solve the equations of interest. An important feature of our approach is that approximations are given by explicit functional expressions, enabling the calculation of the density over the whole domain. This feature makes the approach suited for e.g. simulation and estimation purposes, as will be illustrated using a real example.

A brief review of the approach of Holland and Wang (1987) and Wang (1993) is presented in Section 2. Section 3 is dedicated to integral equation theory. Section 3.1 presents a short overview of integral equation theory. Three numerical techniques are suggested to compute bivariate densities, all based on adapting classical integral equation methods for our purpose. Section 3.2 presents the degenerate kernel method, Section 3.3 the expansion method and Section 3.4 the numerical integration method. Both the expansion method and the numerical integration method can be viewed as specific implementations of the iterative marginal replacement algorithm. The advantage of the methods we propose is the availability of a functional form. In Section 4 error bounds for the approximations are derived. The method is applied to reconstruct some well-known densities in Section 5, and to generate some "new" densities. In Section 6 we discuss application of the technique, including a real data example.

2. The Local Dependence Function

Consider an $r \times c$ table with cell probabilities p_{ij}, $(1 \leq i \leq r, 1 \leq j \leq c)$. For any two pairs of indices (i, j) and (k, l) the cross-product ratio is defined as $\alpha_{ij,kl} = (p_{ij}.p_{kl})/(p_{il}.p_{kj})$. The local cross-product ratios are defined by

$$\alpha_{ij} = \frac{p_{ij}.p_{i+1,j+1}}{p_{i,j+1}.p_{i+1,j}}, \qquad 1 \leq i \leq r-1,\ 1 \leq j \leq c-1. \tag{2.1}$$

Often the log cross-ratios $\gamma_{ij} = \ln \alpha_{ij}$ are considered for the added advantage of linearity on the log scale. It is well-known that the set $(\alpha_{ij})_{ij}$ (or $(\gamma_{ij})_{ij}$) together with the marginal probabilities $(p_{i.})_i$ and $(p_{.j})_j$ completely determine the table of cell probabilities. They are a maximal invariant of (p_{ij}) under row and/or column multiplication. Holland and Wang (1987) and Wang (1993) extended this construction to the situation of continuous bivariate densities. They proceeded as follows.

Consider a bivariate density function $f(x, y)$ with support $D = \{(x, y) | f(x, y) > 0\}$. Partition the support by an infinitesimal rectangular grid. The probability of the rectangle $[x, x + dx] \times [y, y + dy]$ equals $f(x, y)\, dx\, dy$. The cross-ratio of a pair of points (x, y) and (u, v) is defined as

$$\alpha(x, y; u, v) = \frac{f(x, y).f(u, v)}{f(x, v).f(u, y)}.$$

The local cross-ratio at (x, y) is given by

$$\gamma_f(x,y) = \lim_{dx\to 0, dy\to 0} \frac{\ln \alpha(x, y; x + dx, y + dy)}{dx\, dy}.$$

The function $\gamma_f(x, y)$ is called the *local dependence function* (ldf). It is easily seen that

$$\gamma_f(x,y) = \frac{\partial^2}{\partial x \partial y} \ln f(x,y). \tag{2.2}$$

$\gamma_f(x, y)$ is defined whenever $\ln f(x, y)$ is a mixed differentiable function. The local dependence function has some attractive properties :

1. the random variables X and Y with joint density $f(x, y)$ are independent if and only if γ_f equals zero,
2. γ_f is margin free in the sense that $\gamma_f = \gamma_g$ if $g(x, y) = f(x, y)\psi_1(x)\psi_2(y)$,
3. if $f_{1.2}$ and $f_{2.1}$ are the obvious conditional densities, then $\gamma_f = \gamma_{f_{1.2}} = \gamma_{f_{2.1}}$, and
4. γ_f is maximal invariant under both x-marginal and y-marginal replacement, defined in the sense of Wang (1993), Def. 2.1: x-marginal replacement is given by $f(x, y)g_1(x)/f_1(x)$ where $f_1(x)$ is the x margin of $f(x, y)$ and $g_1(x)$ is an arbitrary density with the same support as $f_1(x)$.

For the bivariate normal density with correlation coefficient ρ the local dependence function is constant and equal to $\gamma_N = \rho/(1 - \rho^2)$. For the bivariate Cauchy family with density function $f(x, y) = c/[\pi(x^2 + y^2 + c^2)^{3/2}]$, $(c > 0)$, the local dependence function is equal to $\gamma_c = 6xy/[(x^2 + y^2 + c^2)^2]$ (see Figure 1).

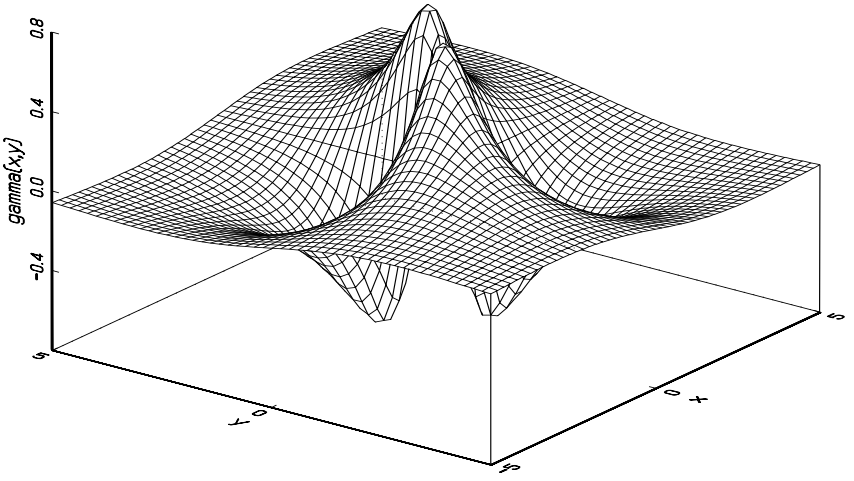

Figure 1. The local dependence function of the bivariate Cauchy density.

Observe that γ_c is positive in the first and third quadrants and negative in the others. This fact indicates that a single number association measure (Pearson's and Spearman's correlation, Kendall's τ) cannot describe the sign-varying association structure.

Holland and Wang (1987) prove the following result :

Theorem 1. (Holland and Wang) *For any given integrable function $\gamma(x, y)$, defined over $D =]a, b[\times]c, d[$ and any given continuous density functions $f_1(x)$ and $f_2(y)$, defined on $]a, b[$ and $]c, d[$ respectively, there exists a unique bivariate density function $f(x, y)$ defined over D such that*
(a) (2.2) *holds with $\gamma_f \equiv \gamma$,*
(b) *$f_1(x)$ and $f_2(y)$ are the respective marginal densities of $f(x, y)$.*

For a proof, see Holland and Wang (1987), Theorem 6.1 or Wang (1993), Theorem 3.1.

This result implies that the bivariate normal density is the only one with normal marginals and constant local dependence function.

In spite of this result, no explicit methods have been given to construct bivariate densities, given marginals and ldf. Wang (1987) suggested a contingency table approach to approximate the bivariate normal density. This method can be extended in a straightforward fashion to construct discretized bivariate densities with other marginals and local dependence function. It is a direct application of the iterative proportional fitting algorithm. The continuous version is the iterative marginal replacement algorithm, proposed by Wang (1993). Wang showed that his algorithm converges but did not provide a close-form solution of the density function. This will be the purpose of the next section.

3. An Analytical Construction of the Bivariate Density

Assume that $\gamma(x, y)$ and the marginal distributions fulfill the requirements of Theorem 1. As $\gamma(x, y)$ is margin-free it is sensible to introduce a *dependence kernel* as follows :

Definition 1. An integrable function $K(x, y)$, defined over D is called a *dependence kernel* for $\gamma(x, y)$ if

$$\gamma_f(x, y) = \frac{\partial^2}{\partial x \partial y} \ln K(x, y).$$

Note that a dependence kernel is the continuous analogue of the initial table used with the iterative proportional fitting algorithm (Bishop et al. (1975), Plackett (1981), p. 36). Due to the uniqueness of $f(x, y)$ one necessarily has $f(x, y) = K(x, y)\phi_1(x)\phi_2(y)$, where f is the bivariate density satisfying Theorem 1.

For a constant local dependence function the simplest dependence kernel is given by $K(x,y) = \exp(\gamma xy)$. Furthermore, for any given ldf, a straightforward integration

$$K(x,y) = \exp\Big(\iint \gamma(x,y)\,dx\,dy\Big)$$

yields a dependence kernel. Otherwise, if $f(x,y)$ is known, $K'(x,y) = f(x,y)/(f_1(x)f_2(y))$ also provides a dependence kernel. Note that $f(x,y)$, $f(x|y)$, and $f(y|x)$ are also dependence kernels. In general, the bivariate density function can be expressed as

$$f(x,y) = \phi_1(x)\phi_2(y)K(x,y) \tag{3.1}$$

and, hence, the two marginal conditions can be written as

$$\begin{aligned} f_1(x) &= \phi_1(x)\int_c^d \phi_2(s)K(x,s)ds, \\ f_2(y) &= \phi_2(y)\int_a^b \phi_1(s)K(s,y)ds. \end{aligned} \tag{3.2}$$

In the special but important case where both marginals are the same function $f(x)$ and K is a symmetric function, (3.2) reduce to

$$f(x) = \phi(x)\int_a^b \phi(s)K(x,s)ds. \tag{3.3}$$

Eqs. (3.2) and (3.3) are examples of integral equations. However, these are not classical cases due to the appearance of the unknown functions in a multiplicative way.

3.1. Integral equations

Some basic ideas of linear integral equations will be presented. For a thorough discussion we refer to Hochstadt (1973). The most studied integral equation is the Fredholm equation of the second kind

$$\phi(x) = f(x) + \lambda\int_a^b \phi(s)K(x,s)ds. \tag{3.4}$$

Here, $K(x,y)$ is a known kernel, $f(x)$ is a known function and $\phi(x)$ is an unknown function; x assumes all values in $[a,b]$ and λ is a paramater. The associated eigenvalue problem (found by setting $f(x) \equiv 0$) is written in operator form as $\phi = \lambda \boldsymbol{K}\phi$, with obvious notation. For a solution, ϕ is called an *eigenfunction* and λ is called an *eigenvalue*. A crucial role is played by the eigenvalues and eigenfunctions of the operator $\boldsymbol{K}$ (see Section 3.2).

Although (3.2) are non-linear integral equations, they are compared to the eigenvalue problem. We will proceed by adapting (numerical) techniques to solve the eigenvalue problem, described in Baker (1977), to our setting.

Theorem 1 guarantees a unique solution of the equations. Three approximation techniques for finding expressions for $\phi_1(x)$ and $\phi_2(y)$ are presented. The first method approximates the kernel K by degenerate kernels (the definition is given in Section 3.2). This method makes it possible to meet the marginal conditions exactly. The other two methods use the exact expression for the kernel (and thus of the local dependence function) and approximate the marginal conditions.

3.2. The degenerate kernel method

Let us first give some definitions.

Definition 2. A set of functions $h_i(x)$, $i = 1, \ldots, n$ is called *linearly independent* if $\sum_{i=1}^{n} a_i h_i(x) = 0$ almost everywhere implies that all a_i vanish.

Definition 3. A (dependence) kernel is called *degenerate* if it can be expressed as a finite sum

$$K(x, y) = \sum_{i=0}^{n} g_i(y) h_i(x), \tag{3.5}$$

where $g_i(y)$ and $h_i(x)$ are integrable functions.

In what follows, assume that $\{h_i(x), i = 1, \ldots, n\}$ is a linearly independent set of functions. This property will be required in both variables, so decompose the kernel alternatively as

$$K(x, y) = \sum_{j=0}^{m} q_j(x) p_j(y), \tag{3.6}$$

with $\{ p_j(y), j = 1, \ldots, m\}$ linearly independent.

Note that the word "degenerate" is used in accordance with its meaning in integral equation theory, and not in the usual statistical sense. For a detailed discussion on degenerate kernels we refer to Hochstadt (1973). A degenerate kernel may be considered as a truncated series expansion of a general kernel.

It is a commonly used technique to approximate kernels by infinite series. A critical issue is how to choose an accurate expansion which can be truncated to form a series of degenerate kernels. In the case of a square integrable, symmetric kernel, there is a very natural solution to this problem. In that case, $K(x, y)$ can be expressed as (see Hochstadt (1973), pp. 44-64)

$$K(x, y) = \sum_{i} \alpha_i \varphi_i(x) \varphi_i(y), \tag{3.7}$$

where the α_i and φ_i are eigenvalues and eigenfunctions of the operator $\boldsymbol{K}$ (see Section 3.1). It also holds true that $\sum_i \alpha_i^2$ is finite. Thus, it is possible to control the error of a degenerate approximation by considering $\sum_{i>n} \alpha_i^2$. However, for general kernels it is difficult to determine these eigenvalues and eigenfunctions.

In the case of a degenerate kernel the following result holds.

Theorem 2. *Let $K(x, y)$ be a degenerate kernel, decomposed as in both* (3.5) *as well as in* (3.6). *The solution to* (3.2) *then is of the form $\phi_k(x) = f_k(x)/P_k(x)$, $(k = 1, 2)$, where $P_1(x) = \sum_{i=0}^n \mu_i h_i(x)$ and $P_2(y) = \sum_{j=0}^m \nu_j p_j(y)$. The unknown coefficients μ_i and ν_j can be found by solving*

$$\mu_i = \int_c^d \frac{f_2(s)g_i(s)}{P_2(s)} ds, \quad (0 \le i \le n),$$

and

$$\nu_j = \int_a^b \frac{f_1(s)q_j(s)}{P_1(s)} ds, \quad (0 \le j \le m).$$

Proof. First, plug in the proposed forms for ϕ_k in (3.2), and use the kernel decomposition (3.5) in the first equation and (3.6) in the second equation. Equating coefficients, and using linear independence, yields the proper expressions for μ_i and ν_j, completing the proof.

The solution for the symmetrical situation (3.3) is found by dropping the indexes for ϕ, f and P in Theorem 2 and putting $\mu_i \equiv \nu_i (i = 1, \ldots, n \equiv m)$:

Corollary 1. *Let $K(x, y) = \sum_{i=0}^n h_i(x)h_i(y)$ be a symmetric degenerate kernel, with $(h_i)_i$ a linearly independent set. Then the solution to* (3.3) *is of the form $\phi(x) = f(x)/P(x)$, where $P(x) = \sum_{i=0}^n \mu_i h_i(x)$ and the unknown coefficients μ_i can be found by solving*

$$\mu_i = \int_a^b \frac{f(s)h_i(s)}{P(s)} ds, \quad (0 \le i \le n).$$

Example 1. Degenerate approximations to the constant dependence kernel $\exp(\gamma xy)$. Consider the following approximations

$$K_n(x, y) = \sum_{i=0}^n \frac{\gamma^i}{i!} x^i y^i. \tag{3.8}$$

Choosing both marginals equal to $f(x)$, the solution is of the form $\phi_n(x) = f(x)/(\sum_{i=0}^n a_i x^i)$, where

$$a_i = \frac{\gamma^i}{i!} \int_a^b \frac{f(s)s^i}{\sum_{j=0}^n a_j s^j} ds, \quad 0 \le i \le n. \tag{3.9}$$

Eq. (3.9) describes a system of $n+1$ non-linear equations in $n+1$ unknowns. It can be solved in various iterative ways. All methods produce a sequence $(\phi^{(k)})$, which, if the problem is well conditioned, converges to the solution for (3.8). The meaning of *convergence* is made precise by the following definition.

Definition 4. $(\phi_i^{(k)})_{k\geq 0}, (i=1,2)$ is called a couple of convergent sequences for (3.2) if and only if

1. $f_1(x) = \phi_1^{(k+1)}(x) \int_c^d \phi_2^{(k)}(s)K(x,s)ds, \quad (k \geq 0),$
2. $f_2(y) = \phi_2^{(k+1)}(y) \int_a^b \phi_1^{(k)}(s)K(s,y)ds, \quad (k \geq 0),$
3. $\phi_i^{(k)} \xrightarrow{L^2} \phi_i, \quad (i=1,2),$
4. (ϕ_1, ϕ_2) is the solution to (3.2).

Note that a pair $(\phi_1^{(k+1)}, \phi_2^{(k)})$ statisfies the first marginal condition and a pair $(\phi_1^{(k+1)}, \phi_2^{(k+2)})$ satisfies the second marginal condition. The dependence condition is satisfied throughout the process. Thus, this algorithm is similar to the iterative proportional fitting algorithm (IPF). It is known that the IPF always converges (Bishop et al. (1975), p. 85). A similar argument guarantees the convergence of the algorithm, described in Defintion 4. In fact, the iterative marginal replacement algorithm, defined in Wang (1993) produces instances of such a sequence.

Example 1. (continued) As a numerical illustration, consider the constant dependence kernels for $n = 12, 18, 24$ with $\gamma = 1$, or equivalently, $\rho = 0.618$, and normal margins, on the rectangle $[-5,5] \times [-5,5]$. Symmetry entails that odd coefficients a_i in $\phi_n(x)$, defined in (3.9) are equal to zero. The even coefficients are displayed in Table 1. Clearly, the convergence of the lower order coefficients is already acceptable for $n = 12$.

Table 1. Degenerate kernel method. Truncated normal density on $[-5,5]$ with $\gamma = 1$. Coefficients a_i of $\phi_n(x) = f(x)/(\sum_i a_i x^i)$ for $n = 12, 18, 24$. Coefficients of order 20 and higher are smaller than 1e-10 and have been omitted from the table.

	a_i		
i	$n=12$	$n=18$	$n=24$
0	0.88664	0.88665	0.88665
2	0.27400	0.27399	0.27399
4	0.04237	0.04233	0.04233
6	0.00437	0.00436	0.00436
8	0.00034	0.00034	0.00034
10	0.00002	0.00002	0.00002
12	1e-6	1e-6	1e-6
14	-	5e-8	4e-8
16	-	2e-9	2e-9
18	-	1e-10	1e-10

In the next sections methods are described to construct functions that satisfy the condition on the local dependence function exactly, and approximate the marginal conditions to any degree of accuracy.

3.3. The expansion method

In Section 3.2 the dependence function is replaced by an approximation and the two marginal conditions are solved iteratively for that approximation. ϕ_k, presented in Theorem 2, contains the marginal density in the numerator, the denominator is a linear combination of the same functions as occur in the kernel. Here, we shall express the solution as $\phi(x) = f(x)/\psi(x)$, with $\psi(x)$ a series of the form

$$\psi(x) = \sum_i \mu_i h_i(x). \tag{3.10}$$

For $(h_j(x))_j$ any convenient basis for $L^2[a,b]$ (the square integrable functions on the interval $[a,b]$) can be taken. If $h_j(x) = x^j, (j \geq 0)$, then the solution is expressed as a formal power series, where $(\mu_j)_j$ is a sequence of unknown coefficients. In general, substituting (3.10) in (3.3) gives

$$\sum_j \mu_j h_j(x) = \int_a^b \frac{1}{\sum_i \mu_i h_i(s)} f(s)K(x,s)ds. \tag{3.11}$$

To determine a numerical solution, $\psi_n(x)$ say, restrict the series expansion in (3.10) to $n+1$ terms. Whence (3.11) is approximated by

$$\sum_{j=0}^{n} \mu_j h_j(x) \simeq \int_a^b \frac{1}{\sum_{i=0}^{n} \mu_i h_i(s)} f(s)K(x,s)ds = P(x). \tag{3.12}$$

Note that this method leaves the dependence function exact but approximates the marginal densities via (3.10). An exact solution cannot be obtained anymore, but, as there are $n+1$ unknown parameters μ_i, an exact solution can be obtained for all x in the set of $n+1$ points $\{x_0, \ldots, x_n\}$.

Again, this system can be solved iteratively. Let $\boldsymbol{m}$ be the column vector with elements μ_i, A be the matrix with elements $a_{ij} = h_j(x_i)$ and $\boldsymbol{p}$ be the column vector with ith element $P(x_i)$. This system of equations is written in the form

$$A\boldsymbol{m} = \boldsymbol{p}. \tag{3.13}$$

Since the $h_j(x)$ are linearly independent and the x_i are different, A is nonsingular with probability 1. We may write $\boldsymbol{m} = A^{-1}\boldsymbol{p}$, which can be used in an iteration scheme. Eq. (3.13) differs from a classical eigenvalue problem in that $\lambda \boldsymbol{I} m$ (where $\boldsymbol{I}$ denotes the identity matrix), is replaced by $\boldsymbol{p}$, which is nonlinear in the μ_i.

When using power functions $h_j(x) = x^j$, we chose starting values to be $\mu_0 = 1$ and $\mu_i = 0$ for $i > 0$. This simple set yields very satisfactory results in the sense that we never observed divergence, except for uniform marginals with high n, a problem discussed in Section 4.2.

The method is presented for identical marginals, but generalization to different marginals is straightforward. In that case, there are $n + m + 2$ unknown parameters $(\mu_i)_i$ and $(\nu_j)_j$, arranged in two column vectors $\boldsymbol{m}$ and $\boldsymbol{n}$ respectively. Further, there are two matrices A and B, and two functional column vectors $\boldsymbol{p}$ and $\boldsymbol{q}$. The iteration scheme is given by

$$\begin{pmatrix} \boldsymbol{m} \\ \boldsymbol{n} \end{pmatrix} = \begin{pmatrix} A^{-1} & \mathbf{0} \\ \mathbf{0} & B^{-1} \end{pmatrix} \begin{pmatrix} \boldsymbol{p} \\ \boldsymbol{q} \end{pmatrix}.$$

Inevitably, the solutions will depend on the choice of the functions $h_j(x)$. It is natural to seek an optimal choice. Note that if $h_j(x) = x^j$, the matrix A is equal to a Vandermonde matrix, which is known to be ill-conditioned. Therefore, these functions should only be used with great care.

Example 1. (continued) Let us consider the bivariate normal example again for approximation orders $n = 12, 18, 24$. To enable comparison of this solution with the one obtained for the degenerate kernel method, we opt for power functions $h_j(x) = x^j$ and the nodes x_i are chosen to be equally spaced in $[-5, 5]$. The coefficients are displayed in Table 2. Alternatives to power functions, in the context of the degenerate kernel method, are discussed in Section 4.1. Observe that the coefficients of $\phi_{24}(x)$ in Table 1 and the coefficients of $\psi_{24}(x)$ in Table 2 are virtually identical. However, the lower order approximations are clearly better in Table 1. In Section 4.3 we study whether this property translates into better approximations for the true (bivariate normal) density.

Table 2. Expansion method. Truncated normal density on $[-5, 5]$ with $\gamma = 1$. Coefficients μ_i of $\psi(x) = \sum_i \mu_i x^i$ for $n = 12, 18, 24$. Coefficients of order 20 and higher are smaller than 1e-10 and have been omitted from the table.

	μ_i		
i	$n = 12$	$n = 18$	$n = 24$
0	0.88703	0.88665	0.88665
2	0.22454	0.27409	0.27399
4	0.14592	0.04185	0.04233
6	-0.05064	0.00501	0.00436
8	0.01132	-0.00003	0.00034
10	-0.00089	0.00012	0.00002
12	0.00003	-0.00001	1e-6
14	-	1e-6	3e-8
16	-	6e-8	1e-8
18	-	1e-9	-8e-10

Alternatively, an approximation method can be developed which is invariant to the choice of a linearly independent base. This is the subject of the next section.

3.4. The numerical integration method

Without aproximating the dependence kernel, we could approximately solve the problem by replacing the exact integrations by numerical integrations. Consider numerical quadrature rules for an integrable function g and h on $[a, b]$ and $[c, d]$ respectively:

$$\int_a^b g(x)dx = \sum_{i=0}^{n} w_i g(x_i),$$
$$\int_c^d h(y)dy = \sum_{j=0}^{m} v_j h(y_j), \tag{3.14}$$

where w_i (v_j) are the weights and x_i (y_j) are the nodes of the quadrature rule. Using (3.14), (3.2) can be approximated as follows

$$\frac{f_1(x)}{\phi_1(x)} = \int_c^d \phi_2(y)K(x,y)dy \simeq \sum_{j=0}^{m} v_j \phi_2(y_j)K(x,y_j),$$
$$\frac{f_2(y)}{\phi_2(y)} = \int_a^b \phi_1(x)K(x,y)dx \simeq \sum_{i=0}^{n} w_i \phi_1(x_i)K(x_i,y). \tag{3.15}$$

An approximate solution can be derived by letting x (y) assume all values x_i (y_i). Putting $\mu_i = \phi_1(x_i)$ and $\nu_j = \phi_2(y_j)$ leads to a system of $n + m + 2$ non-linear equations in the same number of unknowns:

$$\frac{f_1(x_i)}{\mu_i} = \sum_{j=0}^{m} v_j \nu_j K(x_i, y_j), \quad (0 \le i \le n),$$

$$\frac{f_2(y_j)}{\nu_j} = \sum_{i=0}^{n} w_i \mu_i K(x_i, y_j), \quad (0 \le j \le m).$$

Solving this system and substituting the solution in (3.15) gives the approximate formulae

$$\phi_1(x) \simeq \frac{f_1(x)}{\sum_{j=0}^{m} v_j \nu_j K(x, y_j)},$$
$$\phi_2(y) \simeq \frac{f_2(y)}{\sum_{i=0}^{n} w_i \mu_i K(x_i, y)}. \tag{3.16}$$

Note that the weights w_i and v_j depend on the quadrature rule and the integration interval only, and are therefore independent of the integrands. Further,

they remain constant throughout the iterative process. We opted for a repeated Newton quadrature rule with $n = 3k$ and equally spaced nodes x_i. The weights are determined as follows. Let $I = (3/8)(b-a)/n$, then $w_0 = w_n = I$, $w_{3\ell} = 2I$, for $\ell = 1, \ldots, k-1$, and $w_{3\ell+1} = w_{3\ell+2} = 3I$, for $\ell = 0, \ldots, k-1$. The number of μ_i is odd with this scheme. A similar construction is made for nodes y_j, weights v_j, and unknowns ν_j. We chose as starting values a vector of zeros, with a single 1 for the middle entry. At the tth iteration, $\mu_i^{(t)}$ and $\nu_j^{(t)}$ have been determined. A simple updating scheme is:

$$\mu_i^{(t+1)} = \frac{f_1(x_i)}{\sum_{j=0}^{m} v_j \nu_j K(x_i, y_j)}, \quad (0 \le i \le n),$$

$$\nu_j^{(t+1)} = \frac{f_2(y_j)}{\sum_{i=0}^{n} w_i \mu_i K(x_i, y_j)}, \quad (0 \le j \le m).$$

This method always led to convergence.

Example 1. (continued) Using the quadrature rule and the nodes described previously, we revisited the bivariate normal example. The number of nodes was chosen to be $n = 12, 18, 24$, to enable comparison with the previous methods. Coefficients are presented in Table 3. Although coefficients for different degrees of approximation n were easy to compare with the previous methods (Tables 1 and 2), a similar comparison is less straightforward with the current method. Indeed, changing n also changes the grid at which the function $\phi(x)$ is evaluated. For instance, the grid for $n = 12$ is a subset of the grid for $n = 24$, but the grid for $n = 18$ overlaps only partially. However, it is seen that comparable function values are very close to each other. A more formal comparison of the three methods will be made in Section 4.3.

Table 3. Numerical integration method. Truncated normal density on $[-5, 5]$ with $\gamma = 1$. Coefficients μ_i of $\phi(x)$ for $n = 12, 18, 24$. The first column shows the nodes x_i of the integration method. Since $\phi(x) = \phi(-x)$, only positive nodes are displayed.

	μ_i		
x_i	$n = 12$	$n = 18$	$n = 24$
0.0000	0.46241	0.45121	0.44999
0.4167			0.39099
0.5556		0.35076	
0.8333	0.25774		0.25652
1.1111		0.16531	
1.2500			0.12710
1.6667	0.04632	0.04746	0.04756
2.0833			0.01343
2.2222		0.00830	
2.5000	0.00282		0.00287

	μ_i		
x_i	$n = 12$	$n = 18$	$n = 24$
2.7778		0.00088	
2.9167			0.00046
3.3333	0.00006	0.00006	0.00006
3.7500			5e-6
3.8889		2e-6	
4.1667	4e-7		4e-7
4.4444		5e-8	
4.5833			2e-8
5.0000	7e-10	7e-10	7e-10

4. Error Bounds

We now investigate how close the dependence condition (3.1) and the two marginal conditions (3.2) are met by the three approximations described in Section 3. Note that the error made consists of a

numerical part caused by numerically solving a (non-linear) system of equations: rounding errors, convergence criterion,

analytical part caused by replacing exact equations by approximate formulae. For this discussion, suppose that the first error is negligible when compared to the second one.

In Section 4.1 we consider the error of the degenerate kernel method. We also study some properties of the degenerate kernel approximations to the dependence kernel $K(x, y) = \exp(\gamma xy)$ with constant ldf. Section 4.2 is devoted to the other methods. A simple but useful error bound for the departures from the marginal conditions is given. Finally, an indication is given for the comparison of the three methods, by solving a problem for which the exact solution is known.

4.1. Error for the degenerate kernel method

In this case the error arises from the approximation of the kernel $K(x, y)$ by a degenerate one $K_n(x, y)$.

We first consider the degenerate kernel approximations (3.8) of the constant dependence case and then continue with eigenvalue and eigenfunction approximations to square integrable kernels.

Of major importance is the fact that (3.8) may only be called a dependence kernel if it corresponds to an integrable local dependence function $\gamma_n(x, y)$. For this, $K_n(x, y)$ should be positive on $D = [a, b] \times [c, d\,]$, as otherwise the logarithm is not defined. It follows from next lemma, that this condition is met for every rectangle D if and only if n is an even integer. In the sequel we suppose n is even. Consider a constant ldf and the finite terms approximation $K_n(x, y)$ as in (3.8).

Lemma 1. *Let $H_n(t) = \sum_{i=0}^{n} \frac{t^i}{i!}$. Then, for all n, H_{2n} has no real roots and H_{2n+1} has exactly one real root.*

Proof. The assertion clearly holds for $n = 0$. The result follows by induction.

It follows from Definition 1 that the dependence function $\gamma_n(x, y)$, associated with $K_n(x, y)$ equals

$$\gamma_n(x, y) = \gamma - \frac{\gamma\lambda^{(n)}}{K_n(x, y)} + \frac{n\gamma\lambda^{(n)}\left(K_{n-2}(x, y)\frac{\gamma xy}{n-1} - \lambda^{(n-1)}\right)}{K_n(x, y)^2}, \tag{4.1}$$

where $\lambda^{(k)} = \frac{\gamma^k}{k!}x^k y^k$. Convergence of $\gamma_n(x, y)$ to $\gamma(x, y)$ is guaranteed by the following lemma.

Lemma 2. *For $n = 2k$, $\gamma_n(x, y)$ converges pointwise to γ as $k \to \infty$. Furthermore, convergence is uniform on every bounded domain.*

Proof. Consider (4.1). Let (x, y) be an arbitrary but fixed point. As $K_n(x, y) \to \exp(\gamma xy)$, there exists N_1 such that for all $n > N_1$: $\frac{1}{|K_n(x,y)|} < A$ and $|K_n(x, y)| < B$, with A and B constants. Also, if $n > N_2$: $|\frac{\gamma xy}{n-1}| < C$ for constants N_2 and C. For every $\epsilon > 0$, there exists N_3 such that if $n > N_3$: $\max\{|\lambda^{(n)}|, |n\lambda^{(n)}|\} < \epsilon$. Choose $n > \max\{N_1 + 2, N_2, N_3 + 1\}$, then $|\gamma_n(x, y) - \gamma| < |\gamma|\{A + (BC + \epsilon)A^2\}\epsilon$, implying that $\gamma_n(x, y) \to \gamma$. The proof is complete.

The above lemma is illustrated in Figure 2 (top row plots) where we have plotted in the case $\gamma(x, y) = 1$, $\gamma_n(x, y)$ for $n = 2, 10, 18$. Observe that for $n = 18$ the approximation is much better, but spikes are still found at the corners $(2, -2)$ and $(-2, 2)$. Indeed, the system $x^i y^i$ is not the most optimal choice.

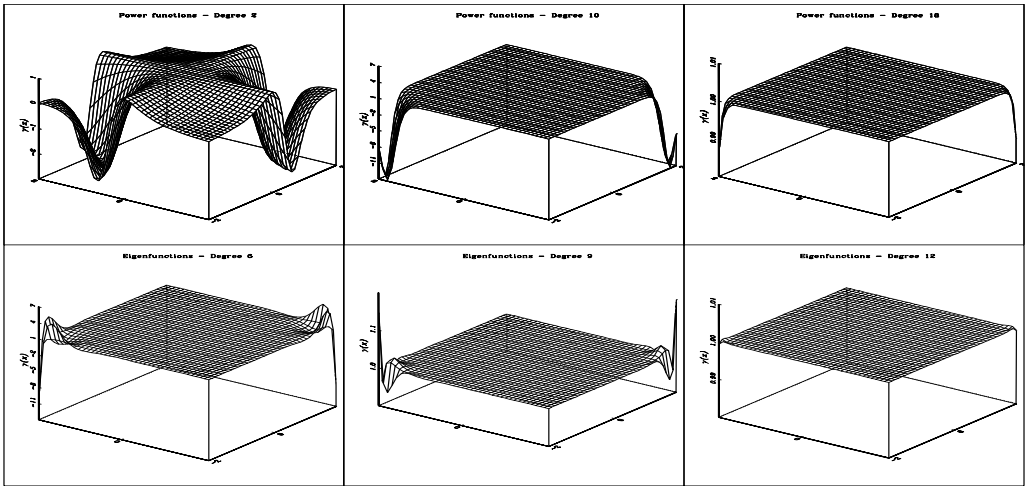

Figure 2. Degenerate kernel approximation $K_n(x, y)$ for a constant dependence function $\gamma(x, y) = 1$. Plot of $\gamma_n(x, y)$ on $[-2, 2] \times [-2, 2]$ for: power functions: (a) $n = 2$, (b) $n = 10$, and (c) $n = 18$; and eigenfunctions (c) $n = 6$, (d) $n = 9$, and (e) $n = 12$.

In the case of a square integrable, symmetric kernel, it is advisable to consider expansion (3.7) which is in terms of eigenvalues and eigenfunctions. Indeed, $\sum_{i=0}^{+\infty} \alpha_i$ is finite, whence for each positive value ϵ, there exists an integer n_ϵ such that $\|K - K_n\|_2^2 < \epsilon$ whenever n exceeds n_ϵ.

Let us consider such a decomposition for the constant dependence kernel. We approximate $K(x, y)$ by $\tilde{K}(x, y)$ such that

$$\tilde{K}(x, y) = \sum_{s=0}^{n} \alpha_s \varphi_s(x) \varphi_s(y)$$

with α_s eigenvalues and φ_s eigenfunctions of K. These functions can be approximated in turn, using the numerical integration method (Baker (1977), Chapter 3). Consider an integration rule such as (3.14), then $\sum_{j=0}^{n} w_j K(x, y_j)\varphi(y_j) = \alpha\varphi(x)$ and by choosing $x = y_i$ for all grid points a standard eigenvalue problem follows: $(KW)\varphi = \alpha\varphi$ with the matrices K and W defined by $(K)_{ij} = K(y_i, y_j)$ and $(W)_{jr} = \delta_{jr} w_j$ and the vector φ by $(\varphi)_r = \varphi(y_r)$. Solving this system yields the following approximate eigenfunctions:

$$\varphi_s(x) \simeq \frac{1}{\alpha_s} \sum_{j=0}^{n} w_j K(x, y_j)\varphi_s(y_j).$$

Figure 2 (bottom row plots) shows the approximations that result for the corresponding ldf, for degrees $n = 6, 9, 12$ (n has to be a multiple of 3 with the adopted quadrature rule). The superiority of the eigenfunctions is immediately clear. To support this claim, we computed the error over a grid of 50×50 points. The error for $n = 6$ with the eigenfunctions is of the same order as the one with $n = 10$ using power functions. For $n = 12$ the maximum error is 6.0123×10^{-4} and for $n = 18$ the maximum error is 1.79×10^{-8}. In the case of power functions, the maximum error for $n = 18$ is still 1.35×10^{-2}. Therefore, the use of eigenfunctions can be advocated, in particular when tail probabilities of the approximated densities are of direct interest.

4.2. Error for the expansion and numerical integration methods

Since the expansion and numerical integration methods use the exact dependence kernel, it suffices to consider the marginal quantities

$$\epsilon_1 = \max_{x \in [a,b]} \Big| f_1(x) - \phi_1(x) \int_c^d \phi_2(y) K(x, y) dy \Big|,$$
$$\epsilon_2 = \max_{y \in [c,d]} \Big| f_2(y) - \phi_2(y) \int_a^b \phi_1(x) K(x, y) dx \Big|.$$

Sharp error bounds can be derived for square integrable symmetric kernels in standard eigenvalue problems but are difficult for more general kernels (Baker (1977)). Even for this type of kernel our problem is non-standard, so we present only a simple error bound, which necessarily is too pessimistic.

First, consider the expansion method. For simplicity we assume a symmetric kernel and marginal densities that coincide. Let us choose an arbitrary point x in $[a, b]$ and let x_i be the closest grid point. We assume the grid points to be equally spaced. Let f denote the true marginal density and $f^{(n)}$ the approximation of order n. Then $f(x_i) = f^{(n)}(x_i)$ and $|f(x) - f^{(n)}(x)| \leq |f(x) - f(x_i)| + |f^{(n)}(x) -$

$f^{(n)}(x_i)|$. Let $H_g = \max_{x\in[a,b]} |g'(x)|$ for a function g defined over $[a,b]$, then

$$\begin{aligned}|f(x) - f^{(n)}(x)| &\leq (H_f + H_{f^{(n)}})|x - x_i| \\ &\leq (H_f + H_{f^{(n)}})\frac{b-a}{n}.\end{aligned}$$

As the order of n increases, it is sensible to approximate $H_{f^{(n)}}$ by H_f, yielding (for n sufficiently large)

$$\|f - f^{(n)}\|_\infty \leq 2H_f\frac{b-a}{n}. \tag{4.2}$$

Note that this inequality is only approximate and could be violated in the tail areas. Its primary use therefore lies in the areas where $H_{f^{(n)}}$ is close to H_f, most probably around the mode of f.

A similar result holds for the numerical integration method, but we also need to consider the term $|f_1(x_i) - f_1^{(n)}(x_i)|$ as it differs from zero in this case. Note that

$$\begin{aligned}f_1(x_i) &= \mu_i \sum_{j=0}^{m} v_j \nu_j K(x_i, y_j), \\ f_1^{(n)}(x_i) &= \mu_i \int_c^d \phi_2^{(n)}(y) K(x_i, y) dy.\end{aligned}$$

Hence, $|f_1(x_i) - f_1^{(n)}(x_i)| = |\mu_i| E_n(\phi_2(y)K(x_i,y))$ with $E_n(g)$ the error attached to the quadrature method adopted. For a trapezoidal rule, the error is $O(n^{-2})$, while for Simpson's rule we obtain $O(n^{-4})$. This term must be added to (4.2).

Sometimes it is of interest to study the error directly in terms of the bivariate density. We restrict attention to the symmetric case. Observe that the true joint density can be expressed as (3.1) where the function $\phi(x)$ is approximated by $\phi_n(x) = f(x)/\psi_n(x)$ in the expansion method and by (3.1) in the numerical integration method. Given that the dependence kernel is exact with these methods, the error in the joint density depends only on the error in $\phi(x)$. In particular, an approximate upper bound for the error is given by

$$|f(x,y) - f_n(x,y)| \leq 2|f_n(x,y)|.\|\phi - \phi_n\|_\infty.$$

The error in $\phi(x)$ will be discussed in Section 4.3.

The difference of accuracy between both methods, is clearly seen in an example. Table 4 compares the expansion and numerical integration methods directly on three bivariate densities with constant dependence function ($\gamma = 1$) and with marginal density functions: uniform on $[0,1]$, standard normal and Cauchy. The powers of x were used as expansion functions. The numerical integration method

used is the repeated Newton rule (based on a quadratic approximation of the integrand). It is clear from the table that the numerical integration method improves monotonically as the order of the polynomial increases. This is not necessarily true for the expansion method.

Table 4. Comparison of error bounds of the expansion and numerical integration methods applied to three densities with constant dependence ($\gamma = 1$) and marginals: Uniform on $[0, 1]$, standard Normal and standard Cauchy.

n	Expansion			Numerical Integration		
	Uniform	Normal	Cauchy	Uniform	Normal	Cauchy
3	1.13e-4	0.399	0.364	5.59e-6	2.064	1.617
6	1.00e-8	0.401	0.366	3.12e-7	0.0823	0.0406
18	3.75e-6	4.01e-6	2.28e-4	3.81e-9	1.32e-3	5.25e-3
24	-	3.06e-9	3.87e-6	1.21e-9	4.22e-5	1.45e-3
30	-	4.91e-8	5.02e-8	4.99e-10	5.13e-7	4.07e-4

The expansion method did not converge in the case of uniform marginals, for $n = 24, 30$. On the other hand, its fit was already excellent for $n = 6$, judged from the maximal error of 1.00e-8. The numerical problem with the expansion method can be described as follows. With both normal and Cauchy margins the coefficient with the largest absolute value is the intercept μ_0. It converges to 0.887 and 0.810 respectively. Higher order coefficients are relatively small. With uniform margins, a similar pattern is observed up to order 18. The intercept converges to 0.890. For higher order approximations, the coefficients increase rapidly, e.g. for $n = 21$, the coefficient of order 13 lies around 240. The problem is ill-conditioned, as can be deduced from the fact that consecutive iterates do not converge but show an irregular oscillating pattern. It is worthwile to note that no such problems are encountered with the numerical integration method. The absolute value of the largest coefficient converges to 0.450 for normal and Cauchy margins, while it tends to 1.123 for uniform margins. Using Legendre or Chebychev polynomials as expansion functions did not remove this ill-conditioning problem. Alternatively, an orthonormal system in terms of eigenvalues and eingenfunctions of the kernel could be used.

These results are very plausible in the light of the derived error structure. Observe that the numerical integration method error is greater for smaller values of n, due to the presence of the numerical integration term in the error formula. However, this term approaches zero much faster than the other terms, explaining why finally this method is able to perform better. Also, note that for the uniform density the numerical integration method is better from the start. Indeed, in this case the integration is exact.

4.3. Comparison of the three methods

If the exact solution is known, the three approximate solutions can be compared to it. For example, the bivariate normal density is specified by a constant local dependence function and normal marginals. The exact solution clearly is

$$\phi(x) = \Big[\frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{x^2}{2}\Big)\Big]\Big[\sqrt[-4]{1-\rho^2} \exp\Big(-\frac{\rho^2}{1-\rho^2}\frac{x^2}{2}\Big)\Big]$$

and $\gamma(x,y) = \gamma = \rho/(1-\rho^2)$. Suppose $f(x,y)$ is the exact density and we have an approximation $f_*(x,y) = \phi_*(x)\phi_*(y)K_*(x,y)$. Let us compare f with f_*. Let f_d, f_e and f_i denote the degenerate kernel, expansion and numerical integration solutions respectively. In all three cases, let $n = 30$ and truncate the density on $D = [-5,5] \times [-5,5]$. Take $\gamma = 1$, corresponding to a correlation coefficient $\rho = 0.618$. Then for all $(x,y) \in D$ the absolute errors are

$$\begin{aligned} |f_d(x,y) - f(x,y)| &< 1.09e-7, \\ |f_e(x,y) - f(x,y)| &< 1.09e-7, \\ |f_i(x,y) - f(x,y)| &< 4.34e-7. \end{aligned}$$

For the relative error the following figures are obtained :

$$\begin{aligned} |\left(f_d(x,y) - f(x,y)\right)/f(x,y)| &< 1.08e+20, \\ |\left(f_e(x,y) - f(x,y)\right)/f(x,y)| &< 2.35e+2, \\ |\left(f_i(x,y) - f(x,y)\right)/f(x,y)| &< 1.47e-2. \end{aligned}$$

All three methods give an excellent fit, but based on the relative error the numerical integration method is far better than the others. Table 4 indicates that these conclusions may not be generalized. Several remarks ought to be made. First, the interval $[-5,5]$ is fairly wide for a standard normal density. If the region D is reduced to $D = [-3,3] \times [-3,3]$ then all three absolute errors for the bivariate density are equal to $4.7e-4$, and the relative errors vary between 0.14 (numerical integration) and 0.47 (degenerate kernel). Secondly, although Table 1 suggests that the approximation is better for the degenerate kernel method, one should keep in mind that the kernel itself is approximated. It was seen in Section 4.1 that eigenfunctions are able to improve the latter approximation. Thirdly, the maximum absolute (relative) errors for the corresponding univariate normal marginals are $3e-5$ $(2e+1)$ for the expansion method and $5e-7$ $(7e-5)$ for the numerical integration method. Although in both cases graphs of the true and the approximated marginal densities cannot be distinguished visually, there is a small anomaly with the expansion method, where the true density in $x = 5$

equals $1e-6$, and the approximation equals $3e-5$. This is due to the odd terms in the expansion: while theoretically zero, some have very small nonzero values because of numerical inaccuracies. Restricting them to zero removes this anomaly and reduces the relative error for the expansion method to 2.9.

In Section 4.2 it was stated that the error in the bivariate density is directly related to the error in the functions $\phi(x)$ for the expansion and numerical integration methods. The maximum absolute (relative) errors for $\phi(x)$ are $5e-7$ $(2e+1)$ for the expansion method and $7e-6$ $(7e-3)$ for the numerical integration method. Restricting the odd coefficients in the expansion method to zero reduces the relative error to 1.0.

In conclusion, the numerical integration method is best suited for this particular problem. In addition, a more optimal choice for the expansion functions in both the degenerate kernel method and the expansion method could improve the performance. More work on optimal choices of expansion functions would be welcome. There is a trade off between the simplicity of e.g. power expansions and the performance of expansion functions that could turn out to be less tractible.

The errors in this section are computed on an equally spaced grid of 40×40 subintervals of the region D.

5. Examples

It has long been a tradition in statistics to construct bivariate densities with normal marginals; two famous examples are the Morgenstern and Plackett families of densities. Our technique adds to these an infinite number of bivariate densities with normal marginals but with a variety of association structures. As an illustration, we have constructed two such densities, with local dependence function: (1) the Cauchy dependence function and (2) $\gamma(x,y) = -4xy/(x^2+y^2)^2$. Both densities have negative dependence in two quadrants and postive in the others and the dependence vanishes as one moves away from the origin. In Figure 3 these densities are compared with the standard normal density ($\gamma = 1$). All three densities are truncated on the rectangle $= [-5,5] \times [-5,5]$, and computed using the numerical integration method. Observe that the third density has 4 modes !

It is also illustrative to vary the constant dependence and again keep the marginals fixed. Figure 4 shows three bivariate densities with uniform marginals and constant dependencies (1, 3 and 5). The densities are constructed with the degenerate kernel method of degree 12.

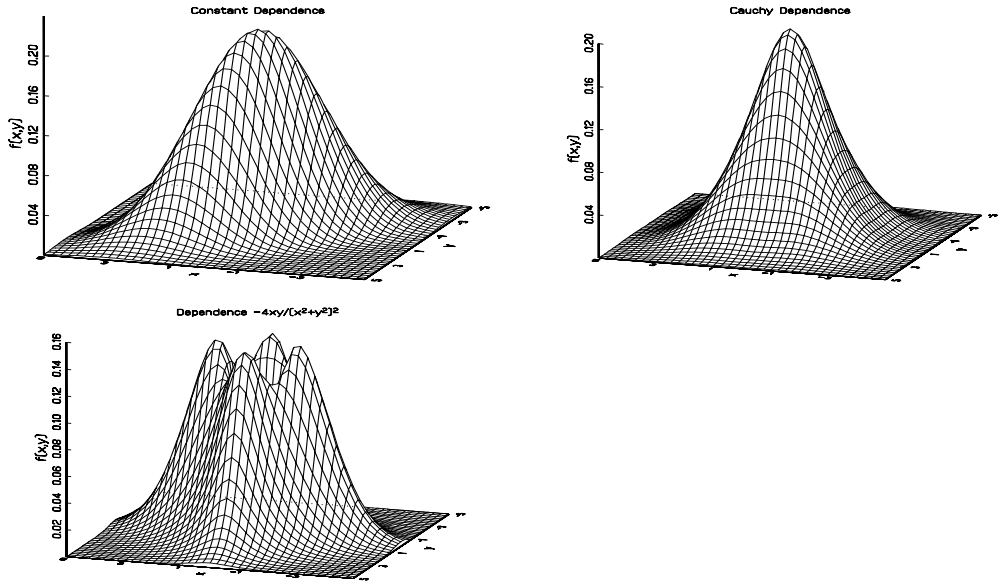

Figure 3. Three densities with standard normal marginals and local dependence function (a) $\gamma(x, y) = 1$, (b) Cauchy dependence $\gamma(x, y) = 6xy/(x^2 + y^2 + c^2)^2$ and (c) $\gamma(x, y) = -4xy/(x^2 + y^2)^2$.

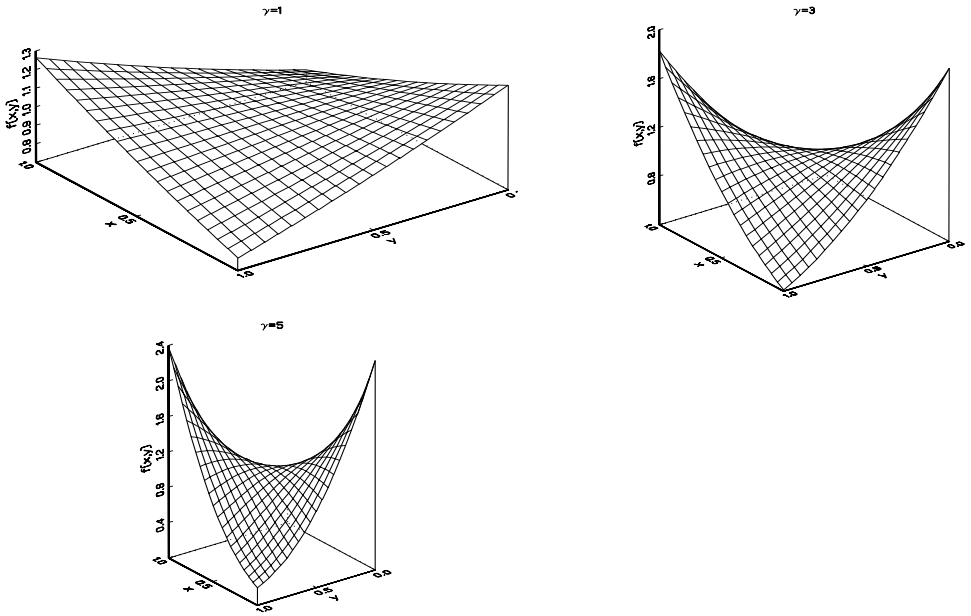

Figure 4. Three densities with Uniform marginals and local dependence function equal to (a) 1, (b) 3 and (c) 5.

One could also show how the shape of the density changes if the dependence function is held constant but the marginals vary.

6. Applications

Let us first discuss some categorical data applications. There is an analogy with loglinear modelling of $r \times c$ contingency tables, where the marginal probabilities are modelled via a logit function and the dependence is modelled via the local cross-ratios (see e.g. Agresti (1990)). This analogy can be used to model contingency tables, arising from a discretized underlying distribution: one can model the marginals in a very general way, together with certain prescribed local dependence structure. The simplest case is to assume a constant local cross-ratio. This establishes a connection with the bivariate probit model, where bivariate categorical data are assumed to arise from discretizing an underlying bivariate normal density.

Simulation studies form another area of application. Dale (1986) introduced the Dale model for bivariate ordered categorical data, and her method was generalized to arbitrary dimension by Molenberghs and Lesaffre (1994). Both the probit and Dale models assume an underlying (multivariate) density: the bivariate normal density for the probit model and the Plackett density for the Dale model. A simulation study was performed by Lesaffre, Verbeke and Molenberghs (1994) to investigate the sensitivity of these models to departures from their densities. Some ad hoc departures were chosen. Using the techniques proposed in this paper, a more systematic study could be undertaken. Indeed, as the Plackett distribution is characterized via a constant global cross-ratio it is natural to investigate the performance of the bivariate Dale model using alternatives with the same marginals but non-constant global cross-ratio structure. Similarly, for the probit model, one could construct alternatives for the bivariate normal underlying density with normal marginals but non-constant local dependence functions.

The method is potentially useful with bivariate continuous outcomes. Most bivariate methods assume bivariate normally distributed data, and even though marginal normality is relatively straightforward to assess, the testing of bivariate normality can be a complex task. Therefore, it can be useful to investigate the performance of a particular method under departures from normality, e.g. by varying the local dependence function while maintaining marginal normality or keeping the constant local dependence but elongating the normal marginals (Gleason (1993)). Again, simulation studies are useful in this context. Maximum likelihood estimation within wide classes of densitities becomes feasible with the proposed approach. For instance, if bivariate data are assumed to be marginally normal, but the association structure is expected to be more complex (e.g. a

Cauchy type dependence), then such a family can be proposed. The use of the proposed method within this framework will be illustrated with an example.

The Tibet Study

In this study, demographic variables together with several blood related characteristics on 320 subjects are collected. The age of the subjects ranges from 20 to 59 years (median of 39.5 years). We will study the relationship between age and the outcome variables high density lipoprotein (HDL) and low density lipoprotein (LDL). HDL ranges from 29 to 115, with a median of 53, while LDL ranges from 11.9 to 234, with a median value of 89. The Pearson correlation between HDL and LDL is -0.179. Age is positively correlated with HDL (0.139) and LDL (0.232). All three correlations are statistically significant at the 5 % level.

The (HDL,LDL) pair forms the bivariate outcome of interest. We seek a joint model that describes the dependence of this outcome on age. A classical "marginal" approach is standard bivariate regression, where the functional dependence of HDL and LDL on age is modelled, together with the correlation between HDL and LDL. Such a model can be fitted with standard statistical software (e.g. SAS procedures GLM and MIXED). The methodology outlined in this paper allows extending standard bivariate regression in two aspects: (1) the normal marginal structure can be replaced by any other marginal density; in theory both outcomes can be modeled using different densities, (2) the dependence structure can be generalized from a constant (such as the correlation coefficient) to arbitrary functions. It is wortwhile to note that each outcome separately, as well as the association between them, is allowed to depend on covariates such as age.

Alternatively, both conditional distributions (HDL given LDL and vice versa) could be specified, as in Arnold, Castillo, and Sarabia (1992). This procedure has severe drawbacks, such as the presence of an often intractible normalizing constant, making standard maximum likelihood difficult to implement (their page 38). Further, conditional normal densities lead not only to the bivariate normal model but also to an unexpected density with anomalous properties, such as a biquadratic regression function (their page 87).

We considered a set of models for the Tibet data. Parameters were estimated by maximum likelihood. Let us first outline the algorithm. The approximation techniques, suggested in this research, allow an efficient computational scheme. Indeed, within each iteration of a numerical likelihood maximization algorithm, the numerical values for the coefficients approximating the assumed density need to be computed only once. Let β_1 be the parameter vector for the regression of HDL and β_2 be the parameter vector for the regression of LDL. Both vectors contain an intercept, an age effect, and the square root of the variance. Further,

let β_3 parametrize the local dependence function between HDL and LDL. Assume that for all parameters starting values have been chosen and that t iterations have been carried out. It suffices to explain how the next iterate is obtained. At the current value of the parameter vector, compute an approximation to the density (for which only β_3 is needed). In case β_3 is independent of the covariates, we need to approximate the density only once within an iterative cycle, and then we proceed by calculating the density in all observed points. (When β_3 depends on covariates, then this evaluation needs to be performed for each covariate level.) For subject i, with outcomes $(\mathrm{HDL}_i, \mathrm{LDL}_i)$ and covariate AGE_i, this amounts to evaluating the density in

$$\left(\frac{\mathrm{HDL}_i - \beta_{11} - \beta_{12}\mathrm{AGE}_i}{\beta_{13}}, \frac{\mathrm{LDL}_i - \beta_{21} - \beta_{22}\mathrm{AGE}_i}{\beta_{23}}\right)$$

with $\beta_{j3} \equiv \sigma_j$, $(j = 1, 2)$, the square root of the variance for the jth variable. Denote the contribution of subject i to the log-likelihood by $\ln f_i$ with $f_i = f(\mathrm{HDL}_i, \mathrm{LDL}_i | \mathrm{AGE}_i, \beta_1, \beta_2, \beta_3)$. The log-likelihood immediately follows as the sum of these contributions and is easy to handle, in contrast to the conditional approach of Arnold, Castillo, and Sarabia (1992). We implemented the log-likelihood function in GAUSS, and maximized it numerically using a Newton-Raphson scheme by means of the built-in GAUSS routine OPTMUM. The data set, as well as the GAUSS routines, are available from the authors upon request.

Next, we present some examples of bivariate models, which are all based on univariate normal marginals, but have varying local dependence functions. Assuming a model with normally distributed outcomes and a constant ldf is equivalent to assuming a bivariate normally distributed outcome, and hence the estimates for the regression parameters should equal the values found by the ordinary least squares estimator. This was observed up to 7 decimal places. The values are presented in Table 5, Model 1. Departure of the constant local dependence (constant correlation) was investigated by two models:

$$\begin{aligned}
&\text{Model 2} \ : \ K(x, y) = \exp(\beta_{30}xy + \beta_{31}x^2y + \beta_{32}xy^2),\\
&\text{Model 3} \ : \ (x, y) = \exp(\beta_{30}xy + \beta_{33}x^2y^2),
\end{aligned}$$

corresponding to ldf's

$$\begin{aligned}
&\text{Model 2} \ : \ \gamma(x, y) = \beta_{30} + 2\beta_{31}x + 2\beta_{32}y,\\
&\text{Model 3} \ : \ \gamma(x, y) = \beta_{30} + 4\beta_{33}xy.
\end{aligned}$$

Model 2 includes linear deviations of the constant ldf, while Model 3 includes an interaction term between both variables. From Table 5 it is seen that these models do not provide a significant improvement over Model 1. These models

allow for detection of well-defined departures of the constant local dependence function. The raw correlation coefficient between HDL and LDL of -0.179 is reflected in the negative value for the dependence parameter β_{30}. In Model 4, the ldf is replaced by the Cauchy dependence function. Although this model clearly misspecifies the association (given the decrease in the log-likelihood), the marginal parameters are not too much affected. However, the marginals are affected with Model 5, where the local dependence function is of the form $-4xy/(x^2+y^2)^2$. Clearly, misspecifying the association could have an impact on the marginal regression parmaeters.

Table 5. Parameter estimates (standard errors) for five models fitted to the Tibet study. All models assume normal marginals. The local dependence functions are: (1) constant, (2) constant and linear term, (3) constant and interaction term, (4) Cauchy dependence, (5) $-4xy/(x^2+y^2)^2$.

Parameter	Model 1	Model 2	Model 3	Model 4	Model 5
HDL					
Intercept	48.62(2.79)	49.04(2.84)	48.57(2.80)	48.74(2.64)	47.49(2.25)
Age	0.17(0.07)	0.16(0.07)	0.17(0.07)	0.17(0.07)	0.23(0.05)
σ	13.77(0.54)	13.84(0.56)	13.79(0.55)	13.84(0.55)	13.78(0.56)
LDL					
Intercept	67.35(6.48)	67.31(6.45)	67.46(6.40)	66.63(6.21)	79.74(4.27)
Age	0.67(0.16)	0.67(0.16)	0.66(0.16)	0.70(0.15)	0.35(0.11)
σ	32.02(1.27)	32.05(1.29)	32.06(1.27)	32.28(1.28)	31.83(1.26)
Association					
β_{30}	$-0.23(0.06)$	$-0.26(0.08)$	$-0.26(0.07)$		
β_{31}		0.02(0.04)			
β_{32}		0.04(0.04)			
β_{33}			$-0.02(0.02)$		
loglik	-2221.03	-2220.47	-2220.55	-2226.25	-2279.68

All models show that both HDL and LDL significantly increase with age. From Models 1–3 it is clear that the negative association between HDL and LDL remains, even after adjusting for age. Further, it can be deduced that the difference between HDL and LDL increases with age. Indeed, a simple univariate regression of HDL minus LDL on age reveals a significant effect ($P = 0.007$). The corresponding model is $\text{HDL} - \text{LDL} = 18.7 + 0.5\text{AGE}$. It could be expected that this relationship has an effect on the association between HDL and LDL. In particular, one could foresee that this association changes with age. However, extending Model 1 to include an age effect in the association, e.g. $\beta_{30} = \beta_{300} + \beta_{301}\text{AGECLASS}$, where AGECLASS is 1 if age exceeds the median and 0 otherwise, yields parameter estimates $\beta_{300} = -0.23$ and $\beta_{301} = -0.0046$. The

log-likelihood changes only in the third decimal place. Our example illustrates the flexibility with which the dependence can be modeled as a function of covariates.

7. Concluding Remarks

We suggested a general method to construct bivariate densities, controlling the marginal densities and the local association structure separately. All techniques depend on numerical methods, involving approximations. Nevertheless it was shown that an excellent fit can be obtained by taking the order of the approximation sufficiently high and by choosing adequate expansion functions. An advantage of the proposed method over the contingency table approach of Wang (1987) is that the density can be computed in any point, without having to store a complete table of cell probabilities. The method improves the technique of Wang (1993) by producing functional expressions that can be evaluated over the whole domain. The performance of the numerical techniques can be assessed by checking the dependence and marginal conditions. A further area of research is the generalization of the local dependence function to a multivariate version. Wang (1993) suggested a trivariate version. The generalization of the existence and uniqueness theorem of Holland and Wang will certainly not have a straightforward generalization, as compatibility conditions of the marginals will be involved. Furthermore, numerical techniques will involve systems of several non-linear integral equations.

Finally, the method makes it feasible to compute extensions of bivariate normal regression to other margins and other dependence structures. It was shown using the Tibet study that computations are feasible. As a referee pointed out, this is potentially useful in economic data, where the dependence pattern is often more complicated than in medical or biological data.

All programs were written in GAUSS. They can be obtained from the corresponding author upon request.

Acknowledgements

The authors wish to thank an anonymous referee for helpful comments on earlier drafts of this manuscript. This research was made possible by Nationaal Fonds voor Wetenschappelijk Onderzoek (Belgium).

References

Agresti, A. (1990). *Categorical Data Analysis*. John Wiley, New York.

Arnold, B. C., Castillo, E. and Sarabio, J.-M. (1992). *Conditionally Specified Distributions*. Lecture Notes in Statistics 73. Springer, Berlin.

Baker, C. T. H. (1977). *The Numerical Treatment of Integral Equations*. Clarendon Press, Oxford.

Bishop, Y. M. M., Fienberg, S. E. and Holland, P. W. (1975). *Discrete Multivariate Analysis: Theory and Practice*. MIT Press, Cambridge, Mass.

Dale, J. R. (1986). Global cross-ratio models for bivariate, discrete, ordered responses. *Biometrics* **42**, 909-917.

Edlefsen, L. E. and Jones, S. D. GAUSS. Aptech Systems Inc., Kent, WA.

Gleason, J. R. (1993). Understanding elongation: the scale contaminated normal family. *J. Amer. Statist. Assoc.* **88**, 327-337.

Hochstadt, H. (1973). *Integral Equations*. John Wiley, New York.

Holland, P. W. and Wang, Y. J. (1987). Dependence function for continuous bivariate densities. *Comm. Statist. Theory Methods* **16**, 863-876.

Johnson, M. E. (1987). *Multivariate Statistical Simulation*. John Wiley, New York.

Johnson, M. E. and Tenenbein, A. (1981). A bivariate distribution family with specified marginals. *J. Amer. Statist. Assoc.* **76**, 198-201.

Marshall, A. W. and Olkin, I. (1988). Families of multivariate distributions. *J. Amer. Statist. Assoc.* **83**, 834-841.

Molenberghs, G. and Lesaffre, E. (1994). Marginal modeling of correlated ordinal data using a multivariate plackett distribution. *J. Amer. Statist. Assoc.* **89**, 633-644.

Lesaffre, E., Verbeke, G. and Molenberghs, G. (1994). A sensitivity analysis of two multivariate response models. *Comput. Statist. and Data Anal.* **17**, 363-391.

Morgenstern, D. (1956). Einfache beispiele zweidimensionaler verteilungen. *Mitteilungsblatt für Mathematische Statistik* **8**, 234-235.

Plackett, R. L. (1965). A class of bivariate distributions. *J. Amer. Statist. Assoc.* **60**, 516-522.

Plackett, R. L. (1981). *The Analysis of Categorical Data* (2*nd ed*). Griffin, London.

Wang, Y. J. (1987). The probability integrals of bivariate normal distributions: a contingency table approach. *Biometrika* **74**, 185-190.

Wang, Y. J. (1993). Construction of continuous bivariate density functions. *Statist. Sinica* **3**, 173-187.

Biostatistics, Limburgs Universitair Centrum, Universitaire Campus, B-3590 Diepenbeek, Belgium.

E-mail: gmolenb@luc.ac.be

Biostatistical Centre for Clinical Trials, School of Public Health, Katholieke Universiteit Leuven, Capucijnenvoer 35, B-3000 Leuven.

E-mail: emmanuel.lesaffre@biostat.be

(Received May 1993; accepted August 1996)