Statistica Sinica 3(1993), 173-187

CONSTRUCTION OF CONTINUOUS BIVARIATE DENSITY FUNCTIONS

Yuchung J. Wang

Academia Sinica

Abstract: The information contained in a discrete two-way contingency table can be decomposed into three independent components: row marginals, column marginals and cross-product ratios. The log-linear models and association models for ordered contingency tables demonstrate the richness and flexibility of considering these components separately. Holland and Wang (1987a) introduced the local dependence function, which mimics the cross-product ratios, for continuous bivariate densities. We proved that the local dependence function and two marginal densities uniquely determine the joint density. An iterative procedure to approximate the joint density was presented. Wang (1987) demonstrated this approach for the bivariate normal density. Different systems of bivariate distributions are examined to reveal their strength and limitation from a discrete perspective. Possible extensions are suggested.

Key words and phrases: Local dependence function, marginal-free, iterative marginal replacement, systems of bivariate density functions.

1. Introduction

The construction of continuous bivariate density functions (bpf's) has been approached from different directions. An insight into the pattern of association between two random variables can influence our approaches to regression and correlation analysis. Yule (1897) and Pearson (1905, 1923) studied the relationship between some skewed bivariate surface and non-linear regression. Their observations reflected the need to study non-normal bpf's with prescribed dependence properties. According to Mardia (1970), some of the systematic methods for constructing continuous bpf's can be classified as:

(a) The solution of Pearson's system of differential equations (see Van Uven (1947-1948));
(b) series expansion using the normal distribution or marginal distributions as basis, such as Edgeworth expansion and canonical expansion; and
(c) the translation method of Johnson (1949).

No family of distribution functions can offer a spectrum broad enough to describe adequately all the dependence structure contained in the continuous bpf's used in application. Kotz (1975) stated "One of the central problems of modern multivariate distribution theory is, no doubt, the proper and meaningful assessment of dependence between components". His statement, unfortunately, still bears some truth today. This, to a certain extent, is rooted in our conventional wisdom to summarize dependence in a single measure of association. A discrete perspective, as we are about to unveil, may shed some new light on this old problem.

Plackett (1965) discretized a continuous bivariate frequency surface into 2×2 discrete contingency tables and considered the 2×2 odds ratios as measures of dependence between two continuous random variables (r.v.'s). He is among the first to approach the construction of bpf's with a discrete intuition. But how to generate a bivariate normal distribution from Plackett's system is still unknown. One reason his approach is not a complete success may be due to his scheme of discretization. Considering only a 2×2 partition oversimplifies the process of discretization. In this paper we will partition a bivariate surface into an $r \times c$ table and study the construction of a continuous bivariate density analogous to the construction of a two-way contingency table.

In the following, we first review some known results in contingency tables. To mimic the local cross-product ratios proposed by Yule & Kendall (1950, p.56) and Goodman (1969) a local dependence function is defined. The construction of a continuous bpf is discussed in Section 3. We prove that the local dependence function along with any two univariate marginal densities uniquely determines a bpf, which can be successively approximated by the iterative marginal replacement process. This establishes a systematic way to construct all bivariate bpf's. In Section 4, we examine the local dependence function of different systems of bivariate densities. It is demonstrated that Pearson's family and Johnson's family, among others, pose restrictions on their local dependence functions. Possible generalizations and extensions are also suggested.

2. Two-Dimensional Dependence Functions

In order to demonstrate the analogy between discrete and continuous distributions, some notions about dependence are briefly reviewed.

2.1. Dependence measures for two discrete random variables

An $r \times c$ contingency table with cell probabilities p_{ij} specifies the joint probability distribution of two discrete r.v.'s, S and T, via

$$p_{ij} = Pr(S = i, T = j), \quad \text{for} \quad 1 \leq i \leq r \quad \text{and} \quad 1 \leq j \leq c. \tag{2.1}$$

The two marginal distributions for S and T are $Pr(S=i)=\sum_{j=1}^{c}p_{ij}=p_{i+}$ and $Pr(T=j)=\sum_{i=1}^{r}p_{ij}=p_{+j}$, respectively. Yule & Kendall (1950) and Goodman (1969) suggested the following non-redundant set of local cross-product ratios

$$\alpha_{ij}=\frac{p_{ij}p_{i+1j+1}}{p_{ij+1}p_{i+1j}} \quad \text{for} \quad 1\le i<r \quad \text{and} \quad 1\le j<c. \tag{2.2}$$

Also, let $\gamma_{ij}=\ln\alpha_{ij}$. Both α_{ij} and γ_{ij} measure the association of the 2×2 subtables formed by adjacent rows and adjacent columns. These $(r-1)\times(c-1)$ 2×2 tables were called *tetrades* by Yule & Kendall. The local cross-product ratios are invariant under the operations of both *row multiplication* and *column multiplications*, which transform $\{p_{ij}\}$ into $\{p_{ij}a_i\}$ for positive a_i and into $\{p_{ij}b_j\}$ for positive b_j, respectively. It is known that the set $\{\alpha_{ij}\}$, or equivalently $\{\gamma_{ij}\}$, together with the marginal probabilities $\{p_{i+}\}$ and $\{p_{+j}\}$ uniquely determine the population contingency table $\{p_{ij}\}$, see Plackett (1981, p.36).

Lemma 2.1. *The α_{ij} for $1\le i<r$ and $1\le j<c$ are a maximal invariant of $\{p_{ij}\}$ under row multiplication and column multiplication.*

Proof. If two $r\times c$ tables have the same set of α_{ij}, then there exists a row and column multiplication mapping one to another.

2.2. Local dependence function for continuous bivariate densities

In the sequel, let (X,Y) denote two continuous random variables with a mixed-differentiable density function $f(x,y)$ having the support

$$K=\{(x,y):f(x,y)>0\}. \tag{2.3}$$

In this paper we assume that the support of $f(x,y)$ is an open connected set in R^2. The x-marginal density and y-marginal density are defined as $f_1(x)=\int f(x,y)\,dy$ and $f_2(y)=\int f(x,y)\,dx$, respectively. Let the support of $f_1(x)$ be the interval (a,b) and the support of $f_2(y)$ be the interval (c,d), where a,b,c, and d can be infinite. Then, K is a subset of the Cartesian product set $(a,b)\times(c,d)$.

Suppose that K has been partitioned by a fine rectangular grid. The probability content of the rectangle containing the point (x,y) with sides dx and dy is approximately equal to $f(x,y)dxdy$. This probability may be viewed as one cell probability of a large two-way table. For this two way table the local cross-product ratio at (x,y) is defined to be

$$\alpha(x,y)=\frac{f(x,y)f(x+dx,y+dy)}{f(x+dx,y)f(x,y+dy)}, \tag{2.4}$$

which is the analogue of α_{ij} of (2.2). Holland and Wang (1987a) considered the limiting case of $\ln\alpha(x,y)$ as the partitioning grid becomes infinitely fine. The *local dependence function* (ldf), denoted by $\gamma_f(x,y)$, is defined as

$$\gamma_f(x,y) = \frac{\partial^2}{\partial x \partial y} \ln f(x,y). \tag{2.5}$$

Note, $\gamma_f(x,y)$ can be defined for any positive mixed-differentiable function which need not be a density function.

It is easy to see that $\gamma_f(x,y) = 0$ if and only if (X,Y) are independent provided K is a Cartesian product set. The local dependence function satisfies the "margin-free" property defined in the following definitions.

Definition 2.1. Let $f(x,y)$ be a continuous bpf with marginal densities $f_1(x)$ and $f_2(y)$. Let $g_1(x)$ and $g_2(y)$ be, respectively, continuous univariate density defined on the same supports of $f_1(x)$ and $f_2(y)$. The mapping $\Omega_1(f(x,y), g_1(x)) = f(x,y)g_1(x)/f_1(x)$ is called the x-marginal replacement by $g_1(x)$ and the mapping $\Omega_2(f(x,y), g_2(y)) = f(x,y)g_2(y)/f_2(y)$ is called the y-marginal replacement by $g_2(y)$.

Definition 2.2. A function computed from a bpf is called *margin-free* if it is invariant when both x-marginal replacement and y-marginal replacement are applied to the bpf. A statistic is called *margin-free* if it is invariant under both x-marginal replacement and y-marginal replacement.

The x-marginal replacement changes the x-marginal density of $f(x,y)$ from $f_1(x)$ to $g_1(x)$. The *ldf* is not only margin-free but also a maximal invariant under both x-marginal and y-marginal replacements. The proof is postponed to Section 3.

3. Construction of Bivariate Density Function and the Iterative Marginal Replacement Algorithm

For any two univariate density functions $g(x)$ over (a,b) and $h(y)$ over (c,d), let $\{g,h\}$ be the set of all bpf's over $(a,b)\times(c,d)$ which has $g(x)$ as its x-marginal density and $h(y)$ as its y-marginal density. That is

$$\{g,h\} = \{f(x,y) : f_1(x) = g(x) \text{ and } f_2(y) = h(y),\ \ (x,y) \text{ in } (a,b)\times(c,d)\}. \tag{3.1}$$

Also, for any subset S of R^2 and a bivariate function $\gamma(x,y)$ defined on S, let $\{\gamma, S\}$ be the collection of bivariate functions (not necessary bpf's) defined in S which have $\gamma(x,y)$ as their ldf, i.e.,

$$\{\gamma, S\} = \{f : f \text{ has } S \text{ as its support and } \frac{\partial^2}{\partial x \partial y} \ln f(x,y) = \gamma(x,y) \text{ for all } (x,y) \text{ in } S\}. \tag{3.2}$$

For the discrete r.v.'s the specification of α_{ij} is independent of the marginal total p_{i+} and p_{+j}. The margin-free property of ldf shows that the bpf can be decomposed into three disjoint components: the two marginal densities, $f_1(x)$ and $f_2(y)$, and the ldf, $\gamma_f(x,y)$. One way to construct a smooth bpf is to "paste" back the ldf with any two univariate densities as its marginals. We summarize this in the following theorem.

Theorem 3.1. *For any continuous densities $g(x)$ and $h(y)$ defined on (a,b) and (c,d), respectively, and an integrable function $\gamma(x,y)$ defined over $K = (a,b) \times (c,d)$, there exists a unique continuous bpf, $f(x,y)$, defined in K such that*

$$\int f(x,t)\,dt = g(x), \tag{3.3}$$

$$\int f(t,y)\,dt = h(y), \text{ and} \tag{3.4}$$

$$\frac{\partial^2}{\partial x \partial y} \ln f(x,y) = \gamma(x,y) \quad \text{for all } (x,y) \in K. \tag{3.5}$$

Proof. For any given bpf $\pi(x,y)$, defined on K, the existence and uniqueness of a bivariate density of the form $\omega(x,y) = A(x)B(y)\pi(x,y)$ with prescribed marginal $g(x)$ and $h(y)$ have been proved by Kullback (1968). Kullback proved that $\omega(x,y)$ is the bivariate density which minimizes the Kullback information distance between $\pi(x,y)$ and $\{g(x),h(y)\}$, or equivalently, $\omega(x,y)$ is the Kullback information projection of $\pi(x,y)$ to $\{g(x),h(y)\}$. For a given $\gamma(x,y)$, a bivariate density of the form $v(x,y) = C\exp(\Gamma(x,y))$ can be constructed, where $\Gamma(x,y)$ is an anti-derivative of $\gamma(x,y)$ such that $\exp(\Gamma(x,y))$ is integrable over K and C is the integral of $\exp(\Gamma(x,y))$ over K. It is easy to see that the solution bpf, $f(x,y)$, is the Kullback Information projection of $v(x,y)$ to $\{g(x),h(y)\}$.

Lemma 3.2. *The local dependence function $\gamma_f(x,y)$ is a maximal invariant under both x-marginal and y-marginal replacements.*

Proof. If two densities $f(x,y)$ and $g(x,y)$ have the same ldf, i.e., $\gamma_f(x,y) = \gamma_g(x,y)$, then there exists a sequence of x-marginal replacements and y-marginal replacements which transform $f(x,y)$ to $g(x,y)$, due to Theorem 3.1.

Theorem 3.1 is equivalent to the fact that $\{\gamma(x,y),K\} \cap \{g(x),h(y)\} = \{f(x,y)\}$. There is one and only one element in the intersection of $\{\gamma(x,y),K\}$ and $\{g(x),h(y)\}$ regardless of the functional forms of γ, g and h. When $K \neq (a,b) \times (c,d)$ the intersection may be empty. If it is not empty then the solution of (3.3), (3.4) and (3.5) must also be unique, see Holland and Wang (1987b).

For the discrete bivariate distribution the iterative proportional fitting process (IPF) is used to construct a two-way table with prescribed marginals and prescribed cross-product ratios, see Plackett (1981). Kullback (1968), Csiszar (1975) and Speed & Kiiveri (1986) used the x-marginal replacement, Ω_1, and y-marginal replacement, Ω_2, iteratively to achieve the prescribed marginal densities. Similarly, the solution $f(x,y)$ can be approximated successively from $v(x,y) = C\exp(\Gamma(x,y))$ by the iterative marginal replacement described below. Let $f^{(0)}(x,y) = C\exp(\Gamma(x,y))$, $f^{(2k+1)} = \Omega_1(f^{(2k)}, g(x))$ and $f^{(2k+2)} = \Omega_2(f^{(2k+1)}, h(y))$ for $k = 0, 1, 2, \ldots$. The sequence $\{f^{(n)}(x,y), \text{ for } n > 0\}$ is called the *iterative marginal replacement process* (IMR). The above authors proved that the sequence $\{f^{(n)}, \text{ for } n > 0\}$ converges to a unique density.

Numerical implementation of IMR can provide a discrete approximation for the continuous bpf of Theorem 3.1. For example, the numerical solution satisfying $\gamma(x,y) = \theta$ for $0 \le x, y \le 1$, $g(x) = 1$ for $0 \le x \le 1$ and $h(y) = 1$ for $0 \le y \le 1$ can be computed at discrete points $\{(x_i = \frac{2i-1}{2n}, y_j = \frac{2j-1}{2n})$ for $1 \le i, j \le n\}$. The iteration starts from $f^{(0)}(x,y) = \exp(\theta xy)$ and proceeds to compute $f^{(1)}(x_i, y) = f^{(0)}(x_i, y)/f^{(0)}(x_i)$, where $f^{(0)}(x_i) = \int_0^1 f^{(0)}(x_i, t)\,dt$. This is a row-marginal replacement. The column-marginal replacement scales $f^{(1)}(x_i, y)$ into $f^{(2)}(x_i, y_j) = f^{(1)}(x_i, y_j)/f^{(1)}(y_j)$, where $f^{(1)}(y_j) = \int_0^1 f^{(1)}(t, y_j)dt$. The iteration will converge in a few cycles. The result is a discrete representation of the desired bpf. For marginals "curved" more than the uniform distribution, the selection of $\{x_i\}$ and $\{y_j\}$ needs to properly reflect the curvatures of the marginal distributions. Reasonable choices are $\{x_i = F_1^{-1}((2i-1)/2n)$ for $1 \le i \le n\}$ and $\{y_j = F_2^{-1}((2j-1)/2n)$ for $1 \le j \le n\}$, where $F_1(x)$ and $F_2(y)$ are the df's of $f_1(x)$ and $f_2(y)$, respectively.

Theorem 3.1 can be used to characterize bivariate densities. It is well known that different bpf's can have the same marginals. For example, Gumbel (1960) and Marshall & Olkin (1967) each defined a bivariate exponential distribution. Their distributions have the same exponential marginal densities but different ldf's. Continuous bpf's can be characterized and constructed via Theorem 3.1. For example, the bivariate normal distributions can be characterized in the following corollary.

Corollary 3.1. *Let $f(x,y)$ be a bivariate density defined in R^2. If*

$$\begin{aligned} \frac{\partial^2}{\partial x \partial y} \ln f(x,y) &= \text{a constant, } \eta, \\ \int f(x,t)\,dt &= (2\pi)^{-1/2}\exp(-x^2/2), \\ \textit{and } \int f(s,y)\,ds &= (2\pi)^{-1/2}\exp(-y^2/2), \end{aligned}$$

then $f(x,y)$ is the bpf of a normal distribution with correlation coefficient $\{(1 + 4\eta^2)^{1/2} - 1\}/(2\eta)$.

Similarly, bivariate Cauchy densities have univariate Cauchy marginals and $\gamma(x,y) = 3xy/(x^2+y^2+c^2)^2$ for $c > 0$. The ldf indicates that the Cauchy distributed X and Y are positively dependent in the first and third orthant and negatively dependent in the second and fourth orthant. Bivariate Gamma distributions have univariate Gamma marginals and $\gamma(x,y) = (1-q)(y-x)^2$ for $q > 0$. Bivariate F distributions have F marginals and $\gamma(x,y) = (1/2)(vv_1v_2/v_0^2)/(1+(v_1x+v_2y)/v_0)^2$. The above density functions can be found in Mardia (1970).

4. Comparisons with Other Systems of Bivariate Densities

The attempts to construct continuous bpf's with prescribed marginals include Morgenstern (1956), Plackett (1965), Johnson (1949), Narumi (1923), Lancaster (1958), Fréchet (1951) and more recently Kimeldorf and Sampson (1975), Frank (1979), Johnson & Tenenbein (1981) and Marshall & Olkin (1988). Their differences are caused by different strategies to model and express the dependence structure. One method of constructing a bivariate bpf may not include distributions which can be generated by other methods. An examination of the ldf's for different families of constructing bivariate densities shall reveal the reason why no system is broad enough to include all the continuous bpf's. Barndorff-Nielsen (1978, p.27) defined the concept of variation independent for variables. Let $\omega_1,\ldots,\omega_m$ be a collection of variables and let $M_1,\ldots,M_m$ denote their domains of variation. If the domain of variation M say, of the combined variable $\omega = (\omega_1,\ldots,\omega_m)$ is equal to the product, $M_1 \times \cdots \times M_m$, of each and every domain then $\omega_1,\ldots,\omega_m$ are called *variation independent.* Theorem 3.1 implies that the three parameter functions of $f(x,y) : f_1(x), f_2(y)$ and $\gamma(x,y)$, are variation independent. In the following ,we will examine different systems of bpf's. It becomes clear that the ldf is not variation independent of the marginal densities. In other words, the constraint placed on the ldf by the marginals limits the scope of the bpf.

4.1. The Pearson family

Pearson proposed the differential equation approach to generate univariate densities. Van Uven (1947-1948) employed a pair of Pearson-type differential equations to generate bpf's. Consider the following two equations

$$\frac{\partial}{\partial x}\ln f(x,y) = \frac{L_1(x,y)}{Q_1(x,y)} \tag{4.1}$$

$$\frac{\partial}{\partial y}\ln f(x,y) = \frac{L_2(x,y)}{Q_2(x,y)} \tag{4.2}$$

where L_j and Q_j are linear functions and quadratic functions of (x,y), respectively. Since $(\partial^2/\partial x\partial y)\ln f = (\partial^2/\partial y\partial x)\ln f$, the L_j's and Q_j's can not be chosen arbitrarily. Also, some restrictions about the coefficients of L_j and Q_j are needed to ensure the positiveness of $f(x,y)$. The ldf for the Pearson family is of the form:

$$\begin{aligned}\gamma(x,y) &= \Big(\frac{1}{Q_2}\Big)^2 \Big(Q_2(x,y)\Big(\frac{\partial L_2(x,y)}{\partial x}\Big) - L_2(x,y)\Big(\frac{\partial Q_2(x,y)}{\partial x}\Big)\Big) \\ &= \Big(\frac{1}{Q_1}\Big)^2 \Big(Q_1(x,y)\Big(\frac{\partial L_1(x,y)}{\partial y}\Big) - L_1(x,y)\Big(\frac{\partial Q_1(x,y)}{\partial y}\Big)\Big). \qquad (4.3)\end{aligned}$$

Some constrains must be put on (L_1, Q_1) and (L_2, Q_2) to ensure the equality in Equation (4.3). Moreover, the ldf in (4.3) is restricted to a quadratic function divided by a squared quadratic function.

The original purpose of Pearson's family is to generate a rich family with conditional distributions belonging to univariate Pearson's family. It does not place any model on the interdependence between X and Y, explicitly. Equations (4.1) and (4.2) determine the conditional distributions $f(x|y)$ and $f(y|x)$, respectively, whereas the marginal distributions are to be forged. "Sagrista (1952) pointed out a direct extension of Pearson's original first order equations into a system of partial differential equations is impractical, since the equations are difficult to solve and in most cases the solutions are too general in nature.", quoted from Kotz (1974). Kotz also stated that "these surfaces, as Pearson pointed out again, were thus of little practical importance" since they "did not fit the theoretical frequencies". Equations (4.1) and (4.2) contain redundant information about $\gamma(x,y)$, hence, are not variation independent. Our single differential equation (3.5) coupled with two integral equations, (3.3) and (3.4), is more flexible than Pearson's family. It eliminates the complicated restrictions on L_1, L_2, Q_1 and Q_2 represented in (4.3). It also offers to specify marginal densities explicitly.

4.2. Johnson's translation family

Johnson (1949) considered the distribution of (X, Y) such that $Z_1 = a_I((X - \zeta_1)/\lambda_1)$ and $Z_2 = a_J((Y - \zeta_2)/\lambda_2)$ have a joint bivariate normal distribution, where a_I and a_J are the marginal transformations which are both differentiable and one-to-one. From Corollary 3.1, if Johnson's translation family is put in the context of Theorem 3.1, Equation (3.5) takes the following form:

$$\frac{\partial^2}{\partial x \partial y} \ln f(x,y) = c\Big(\frac{d\,a_I(x)}{d\,x}\Big)^{-1} \Big(\frac{d\,a_J(y)}{d\,y}\Big)^{-1}. \qquad (4.4)$$

And equations (3.3) and (3.4) become the following boundary conditions:

$$\int h(x,y)\,dy = (2\pi)^{-1/2} \exp(-a_I^2(x)/2)(d\,a_I(x)/d\,x), \qquad (4.5)$$

and

$$\int h(x,y)\,dx = (2\pi)^{-1/2}\exp(-a_J^2(y)/2)(d\,a_J(y)/d\,y). \tag{4.6}$$

According to equations (4.4), (4.5) and (4.6), it can be concluded that Johnson's translation family is limited in two respects: First, the ldf is a product of a function of x only and a function of y only. Second, the ldf is completely specified by the marginal densities. In short, the joint density is *completely* determined by the two marginal transformations. The bivariate Cauchy, bivariate Gamma and bivariate F distributions, mentioned following Corollary 3.1, obviously can not be members of the Johnson translation family.

An straightforward generalization of Johnson's translation family is to relieve the confinement imposed on ldf by the marginal densities. We can replace the partial differential equation (4.4) via

$$\frac{\partial^2}{\partial x \partial y} \ln f(x,y) = \alpha(x)\beta(y), \tag{4.7}$$

which was proposed by Goodman (1985, p.39). The distribution satisfying (4.5), (4.6) and (4.7), though not completely general, is more flexible than Johnson's family. It was used to model a translated bivariate normal distribution by Goodman (1985).

4.3. Canonical representation

For any given marginal $f_1(x)$ and $f_2(y)$, Lancaster (1963) defined a set of complete orthonormal function $\{\zeta_i(x)\}$ and $\{\eta_j(y)\}$ such that the bpf, $f(x,y)$, can be decomposed as follows:

$$f(x,y) = \left\{1 + \sum_{i=1}^{\infty} \varphi_i \zeta_i(x)\eta_i(y)\right\} f_1(x) f_2(y), \tag{4.8}$$

where $\varphi_i = \iint \zeta_i(x)\eta_i(y)f(x,y)\,dxdy$ is called the ith canonical correlation. The right-hand side of (4.8), in theory, can represent all continuous bpf's. But in practice, it offers limited usefulness because it is hard to determine a bpf from the right-hand-side of (4.8). The ldf gives a clear indication about positive or negative association but Equation (4.8) reveals little about the local dependence. Lancaster pointed out that the above infinite expansion is a continuous version of the following canonical expansion in Fisher (1940) for a discrete $m \times n$ $(m \le n)$ contingency table:

$$p_{ij} = \left\{1 + \sum_{k=1}^{m-1} \psi_k \zeta_{ki} \eta_{kj}\right\} p_{i+} p_{+j}. \tag{4.9}$$

One motivation for (4.9) is the Pearson's chi-square statistic, $\chi^2 = \sum_i \sum_j \{(p_{ij} - p_{i+}p_{+j})^2/p_{i+}p_{+j}\}$, which can be decomposed as the sum of squared canonical correlations, $\sum \psi_i^2$. But the likelihood ratio statistic for a contingency table, $G^2 = \sum\sum(p_{i+}p_{+j})\ln(p_{i+}p_{+j}/p_{ij})$, does not have a simple expression.

Fisher and Lancaster, in fact, took a linear decomposition for cell probability directly. Their approaches are in contrast with the prevailing log-linear models for contingency table analysis. More specifically, the likelihood-based log-linear model offers some practical advantages. Its terms give a direct description of the local dependence, which is not clear from (4.9).

Goodman (1987) proposed canonical correlation models and association models to study ordered contingency table. His canonical models can be represented by (4.9) with ψ_i to be estimated from data. As an alternative to (4.9), he suggested the following log-linear model for the cell probability:

$$p_{ij} = \alpha_i\beta_j \exp\Big(\sum_{k=1}^{m-1} \lambda_k \nu_{ik}\mu_{jk}\Big). \tag{4.10}$$

This representation can be viewed as a discretized version of a continuous bpf $\alpha(x)\beta(y)\exp(\sum_1^{m-1}\lambda_i U(x)V(y))$, which is the solution of equations (4.7), (3.3) and (3.4). Goodman thought (4.10) could offer more flexibility in modeling cell probability than (4.9). Comparison between canonical correlation model (4.9) and association model (4.10) for discrete distribution can shed light on the comparison between the canonical expansion (4.8) and the system of densities proposed in Theorem 3.1 for a continuous bpf.

4.4. Edgeworth expansion

For any bivariate bpf $f(x,y)$, its Edgeworth expansion $q(x,y)$, can be written as

$$q(x,y) = \exp\Big(\Big\{k_{11}\Big(\frac{\partial}{\partial x}\Big)\Big(\frac{\partial}{\partial y}\Big) + \sum_{r+s\geq 3}(-1)^{r+s}k_{rs}\Big(\frac{1}{r!}\Big)\Big(\frac{1}{s!}\Big)\Big(\frac{\partial}{\partial x}\Big)^r\Big(\frac{\partial}{\partial y}\Big)^s\Big\}\phi(x)\phi(y)\Big), \tag{4.11}$$

where k_{rs} is the (r,s) cumulant of $f(x,y)$ and $\phi(x) = (1/2\pi)^{1/2}\exp(-x^2/2)$. The independent normal density $\phi(x)\phi(y)$ is used as a convenient base density (starting point). The partial derivatives are used to find orthogonal coordinates, such as the Hermite polynomials, in the expansion. Most Edgeworth expansion will be limited to order of $r + s \leq 4$. One major purpose of the Edgeworth expansion is to match the cumulants of $q(x,y)$ with that of $f(x,y)$. But the marginal densities of $q(x,y)$ are usually different from those of $f(x,y)$.

Besides, $q(x,y)$ can sometime become negative in the support of K, and it is well-known that the cumulants k_{rs} are not sufficient to specify the dependence structure of $f(x, y)$. In other words, the dependence structure contained in $f(x, y)$ can not be reproduced by $q(x, y)$. An alternative is to first find a finite series approximating the ldf and then to perform IMR for the matching of the marginals. Let $\{\zeta_i(x)\}$ and $\{\eta_j(y)\}$ be orthogonal sets of functions. The ldf can be modeled as

$$\gamma_f(x, y) = \sum_{i=1}^{r} \sum_{j=1}^{s} c_{ij} \zeta_i(x) \eta_j(y). \tag{4.12}$$

Or consider the set $\{\delta_k(x, y)\}$ of bivariate orthogonal functions. The ldf can be expanded as

$$\gamma_f(x, y) = \sum_{k=1}^{n} d_k \delta_k(x, y), \tag{4.13}$$

where the choice of r and s or n may depend on the complexity of $\gamma(x, y)$.

Let the primitive function (anti-derivative) of $\delta_k(x, y)$ or $\zeta_i(x)\eta_j(y)$ be expressed as $\Delta_k(x, y)$. The iterative marginal replacement algorithm can be used to find the expansion which takes the form

$$q(x, y) = A(x)B(y) \exp\Big(\sum_{k=1}^{n} d_k \Delta_k(x, y)\Big), \tag{4.14}$$

such that density function $q(x, y)$ will satisfy the boundary equation (3.3) and (3.4), i.e. $q_1 = f_1$ and $q_2 = f_2$. If Equation (4.12) or (4.13) is a good approximation of $\gamma_f(x, y)$, then $q(x, y)$ offers a good approximation of $f(x, y)$.

Arnold and Strauss (1991) considered a bpf whose conditional distributions belong to any specified exponential families. Their densities take the following forms:

$$q(x, y) = A(x)B(y) \exp\Big(\sum m_{ij} \xi_i(x) \eta_j(y) + \sum a_i \xi_i(x) + \sum b_j \eta_j(y)\Big), \tag{4.15}$$

which can be considered as a special case of (4.14), and the discrete version of (4.15) is Goodman's (4.10).

5. Conclusion

The idea used to construct contingency tables has been extended to the cases of continuous bivariate densities. The iterative marginal replacement process is the continuous analogue of iterative proportional fitting (Bishop, Fienberg & Holland (1975, p.83)). The separation of the dependence structure $\gamma(x, y)$ from the marginals, $f_1(x)$ and $f_2(y)$, offers flexibility to specify bivariate densities. Conceivably all the bivariate densities can be generated from the proposed system. The bivariate normal density takes a simple form by having $\gamma(x, y) = \rho/(1 - \rho^2)$ and $f_1 = f_2 = (1/2\pi)^{1/2} \exp(-x^2/2)$.

In the modeling of a two-way contingency table, the log-linear model represents the cell probability by

$$\ln p_{ij} = \mu + \mu_{1(i)} + \mu_{2(j)} + \mu_{12(ij)}. \tag{5.1}$$

An advantage of model (5.1) is the clear separation of marginal information from dependence. Consider $\mu_{1(i)}$, $\mu_{2(j)}$ and $\mu_{12(ij)}$ as parameters for the discrete distribution $\{p_{ij}\}$. These three sets of parameters are variation independent and L-independent, according to Barndorff-Nielsen (1978). They can be modeled separately without creating a consistency problem. For a continuous bpf, the $\gamma(x,y)$, $f_1(x)$ and $f_2(y)$ are parameters (parameter function) for $f(x,y)$. It is in light of this L-independence, that Theorem 3.1 provides a parameterization which factors the information of $f(x,y)$. It also offers flexibility in modeling the ldf independent of the marginals. In Section 4, we demonstrated some of the constraints, especially variation dependence, placed on the ldf by other systems of bpf. This helps the statistician to understand its limitations and to envision possible generalizations.

The purpose of the paper is to propose a general scheme for constructing all smooth bivariate densities. The scheme also suggests an approach to model data. We suggested modeling the two marginal distributions and ldf separately and then "paste" them back. Theorem 3.1 implies this procedure has a definite solution. It is interesting to note that either the loglinear models or association models are models for the local cross-product ratios (lcpr), α_{ij}, not for the marginal distributions. For example, the independence model assumes $\alpha_{ij} = 1$, the uniform association model assumes $\alpha_{ij} = \theta$ and the row-effect model assumes $\alpha_{ij} = a_i$. It is under these hypothetical models, the cell probabilities are estimated. An acceptable value for the goodness-of-fit statistics implies the model is adequate. The same approach can be applied to continuous data by assuming models for $\gamma_f(x,y)$ and models for marginal densities and finding the minimum-distance estimator for $f(x,y)$. Specifically, one could start with the most parsimonious model, $\gamma_f(x,y) =$ constant, and gradually increase the complexity of models for $\gamma_f(x,y)$ until an acceptable goodness-of-fit is achieved.

As Castillo and Galambos (1987) pointed out, and reiterated by Arnold and Strauss (1988), a researcher frequently has better insight into the forms of the conditional distributions of an experimental variable rather than the joint distribution. Usually data provide direct information about conditional distributions, not joint distributions. Since the ldf derived from the conditional bpf , $f(x|y)$, or $f(y|x)$, is the same ldf derived from the joint bpf, $f(x,y)$, models for (3.5) may come from prior knowledge about the conditional distribution or from slicing data.

Generalization of Theorem 3.1 to multivariate cases is possible, but far more complicated. For example, a trivariate density $f(x, y, z)$ can be identified by the three univariate marginal densities $f_1(x)$, $f_2(y)$, $f_3(z)$, two bivariate marginal ldf's $\gamma_{13}(x, z)$ and $\gamma_{23}(x, z)$ and the three-dimensional conditional ldf $\gamma_{12|3}(x, y|z) = \frac{\partial^2}{\partial x \partial y} \ln f(x, y, z)$ which measures the dependence between x and y conditioned on z. When the three-dimensional ldf $\gamma_{123}(x, y, z) = \frac{\partial^3}{\partial x \partial y \partial z} \ln f(x, y, z)$ is a null function, then $f_1(x)$, $f_2(y)$, $f_3(z)$, $\gamma_{12}(x, y)$, $\gamma_{13}(x, z)$ and $\gamma_{12}(x, y)$ are sufficient for $f(x, y, z)$. Wang (1991) used the above characterization to compute probability integrals for the trivariate normal distribution, because multivariate normal distributions exhibit no three-way dependence. For three-way contingency tables, there are five levels of complexity, ranging from independence to full saturated models. Each model has its corresponding continuous analogue.

References

Arnold, B. C. and Strauss, D. (1988). Bivariate distribution with Exponential conditionals. *J. Amer. Statist. Assoc.* **83**, 522–527.

Arnold, B. C. and Strauss, D. J. (1991). Bivariate distributions with conditionals in prescribed exponential families. *J. Roy. Statist. Soc. Ser.B* **53**, 365–375.

Barndorff-Nielsen, O. (1978). *Information and Exponential Families in Statistical Theory.* John Wiley, New York.

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

Castillo, E. and Galambos, J. (1987). Bivariate distributions with normal conditionals. In *Proc. Internat. Symp. Simulation Modeling and Development Cario*, 59–62. Acta, Anaheim.

Csiszar, I. (1975). I-divergence geometry of Probability Distributions and minimization Problems. *Ann. Probab.* **3**, 146–158.

Fisher, R. A. (1940). The precision of discriminate functions. *Ann. Eugenics* **10**, 422–429.

Frank, M. J. (1979). On the simultaneous associativity of $F(x, y)$ and $x + y - F(x, y)$. *Aequationes Math.* **19**, 194–226.

Fréchet, M. (1951). Sur les tableaux de correlation dont les marges sont donnees. *Ann. Univ. Lyon Sec.A, Ser.3*, **14**, 55–77.

Genest, C. and Mackay, R. J. (1986). Copules archimédiennes et familles de lois bidimensionnelles dont les marges sont données. *Canad. J. Statist.* **14**, 145–159.

Goodman, L. A. (1969). How to ransack social mobility tables and other kinds of cross-classification tables. *Amer. J. Soc.* **75**, 1–39.

Goodman, L. A. (1987). The analysis of cross-classified data having ordered and/or unordered categories: Association models, correlation models, and asymmetry models for contingency tables with or without missing entries. *Ann. Statist.* **13**, 10–69.

Gumbel, E. J. (1960). Bivariate Exponential Distributions. *J. Amer. Statist. Assoc.* **55**, 698–707.

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

Holland, P. W. and Wang, Y. J. (1987b). Regional dependence for continuous bivariate densities. *Comm. Statist. Theory Methods* **16**, 193–206.

Johnson, M. E. and Tenenbein, A. (1981). A Bivariate Distribution Family With Specified Marginals. *J. Amer. Statist. Assoc.* **76**, 198–201.

Johnson, N. L. (1949). Bivariate distributions based on simple translation systems. *Biometrika* **36**, 297–304.

Kimeldorf, G. and Sampson, A. R. (1975). Uniform representations of bivariate distributions. *Comm. Satist.* **4**, 617–627.

Kimeldorf, G. and Sampson, A. R. (1978). Monotone dependence. *Ann. Statist.* **6**, 895–903.

Kotz, S. (1974). Multivariate distributions at a cross road. *Statistical Distributions in Scientific Work*, Vol. 1 (Edited by G. P. Patil et al.), 247–270. D. Riedal Publishing Company, Dordrecht, Holland.

Kullback, S. (1968). Probability densities with given marginals. *Ann. Math. Statist.* **39**, 1236–1243.

Lancaster, H. O. (1963). Correlations and canonical forms of bivariate distribution. *Ann. Math. Statist.* **34**, 532–538.

Mardia, K. V. (1970). *Family of Bivariate Distributions.* Hafner Publishing Co., Darien, Conn.

Marshall, A. W. and Olkin, I. (1967). A multivariate exponential distribution. *J. Amer. Statist. Assoc.* **62**, 30–44.

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

Morgenstern, D. (1956). Einfache Beispiele Zweidimensionaler Verteilungen. *Mitteilingsblatt für Mathematische Statistik* **8**, 234–235.

Narumi, S. (1923). On the general forms of bivariate frequency distributions which are mathematically possible when regression and variation are subjected to limiting conditions. *Biometrika* **15**, 209–221.

Nelsen, R. B. (1986). Properties of a one-parameter family of bivariate distributions with specified marginals. *Comm. Statist.* **15**, 3277–3285.

Pearson, K. (1905). On general theory of skew correlation and non-linear regression. Drupper's Company Research Memoirs, *Biometrics Series*, 2.

Pearson, K. (1923). Notes on skew frequency surfaces. *Biometrika* **15**, 222–230.

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.* Charles Griffin & Company Limited, London.

Sagrista, S. N. (1952). On a generalization of Pearson's curves to the two dimensional case. *Trabajos de Estadistica* **3**, 273–314.

Schweizer, B. and Sklar, A. (1983). *Probabilistic Metric Spaces.* North-Holland, Amsterdam.

Speed, T. P. and Kiiveri, H. T. (1986). Gaussian Markov distributions over finite graphs. *Ann. Statist.* **14**, 138–150.

Van Uven, M. J. (1947-48). Extension of Pearson's probability distributions to two variables, I-III. *Proc. Roy. Academy Sci. Amsterdam* **32**, 793–807, 995–1007, 1085–1103.

Wang, Y. J. (1987). Probability integrals of bivariate normal distributions. A contingency table approach. *Biometrika* **74**, 185–190.

Wang, Y. J. (1991). Multivariate normal integrals and contingency tables with ordered categories. Technical Report B-91-2, Institute of Statistical Science, Academia Sinica, Taiwan, R.O.C.

Yule, G. U. (1897). On the significance of Bravais formula for regression in the case of skew correlation. *Proc. Roy. Soc. London* **60**, 477–489.

Yule, G. U. and Kendall, M. G. (1950). *An Introduction to the Theory of Statistics*, 14th edition. Charles Griffin & Company Limited, London.

Institute of Statistical Science, Academia Sinica, Taipei 11529, Taiwan.

(Received December 1990; accepted June 1992)