Metrika (2004) 60: 261–277
DOI 10.1007/s001840300309

Metrika
© Springer-Verlag 2004

The association between two random elements: A complete characterization and odds ratio models

Gerhard Osius

Institut für Statistik, Fachbereich 3, Universität Bremen; Bibliotheksstrasse, 28359 Bremen, Germany (E-mail: osius@math.uni-bremen.de)

Abstract. For random elements X and Y (e.g. vectors) a complete characterization of their association is given in terms of an odds ratio function. The main result establishes for any odds ratio function and any pre-specified marginals the unique existence of a corresponding joint distribution (the joint density is obtained as a limit of an iterative procedure of marginal fittings). Restricting only the odds ratio function but not the marginals leads to semi-parmetric *association models* for which statistical inference is available for samples drawn *conditionally on either* X *or* Y. *Log-bilinear association models* for random vectors X and Y are introduced which generalize standard (regression) models by removing restrictions on the marginals. In particular, the logistic regression model is recognized as a log-bilinear association model. And the joint distribution of X and Y is shown to be *multivariate normal* if and only if both marginals are normal and the association is log-bilinear.

Key words: Association, Marginal fitting, Multivariate normal distribution, Odds ratio, Semi-parametric models

1 Introduction and outline

The question how a random output vector Y of a system (e.g. the health status of a human) is associated to a random input vector X (e.g. consumption of tobacco and alcohol, environmental pollution and other risk factors) is of major importance in statistical science. The standard approach is to consider the outcome Y for given values of the input X which amounts to specifying the conditional distribution $\mathscr{L}(Y|X)$ by means of an appropriate model - but leaving the marginal distribution $\mathscr{L}(X)$ of the input arbitrary. The corresponding sampling design is to collect a random sample $Y_i \sim \mathscr{L}(Y|X = x_i)$ for given values x_i with $i = 1, \ldots, I$.

Sometimes however, the reverse approach may be more appropriate which asks what input X has been responsible for given values of the output. This leads to a model for the conditional distribution of $\mathscr{L}(X|Y)$ and a sample $X_i \sim \mathscr{L}(X|Y = y_i)$ with fixed values $y_i\,(i = 1, \ldots, I)$. A remarkable example are case-control studies in epidemiology, where $Y \in \{0, 1\}$ is an indicator for a disease and the subpopulations $\{Y = 1\}$resp. $\{Y = 0\}$ are called cases resp. controls (cf. Breslow and Day 1980).

Comparing both approaches for random vectors X and Y the question arises, which amount of information concerning the joint distribution of (X, Y) is contained in both conditional distribution $\mathscr{L}(X|Y)$ and $\mathscr{L}(Y|X)$. Conditioning on X resp. Y removes the information on the marginal distribution $\mathscr{L}(X)$ resp. $\mathscr{L}(Y)$ from the joint distribution $\mathscr{L}(X, Y)$ and the question arises what *exactly* is left if we simultaneously remove the information on both marginal distributions from the joint distribution. Focussing on *densities* - on which likelihood analysis and much of Bayesian statistics are based -we are looking for an *association object Assoc(X,Y)* derivable from *any* of the two conditional densities. The joint distribution $\mathscr{L}(X, Y)$ should be uniquely determined by the tripel $(\mathscr{L}(X), \mathscr{L}(Y), Assoc(X, Y))$ and - more important, *any* such tripel should give rise to a (unique) joint distribution. This approach allows to define a model for the joint distribution by specifying separate models for the association structure *Assoc(X,Y)* and the marginals. Any parametric model for the association *Assoc(X,Y)* which leaves the marginals of X and Y completely arbitrary leads to an *association model,* which is *semi-parametric* with respect to the unspecified marginals. The advantage of an association model over a regression model (specifying the *conditional* density of Y given X) is that inference about the parameters of the association model can be based on samples from either $\mathscr{L}(Y|X)$ or $\mathscr{L}(X|Y)$, because the association structure *Assoc(X,Y)* is a common part of both conditional densities.

An association model may be obtained as a generalization from a common model (e.g. a regression model) by removing the restrictions imposed on the marginals. We briefly describe two examples to be studied in detail later. Consider first the *joint* multivariate normal distribution for random vectors X and Y, which serves as a basic model in multivariate analysis. Identifying its association structure (which is *not* given by the correlation matrix, cf. **4.4**) allows generalizations of multivariate techniques to non-normal *joint* distributions with the same association structure but non-normal *marginals* (cf. van der Linde 2002a). And second, for Y with finite range and a vector X the widely used logistic regrssion model is equivalent to a corresponding association model (cf. **4.2**) and thus semi-parametric with respect to the marginals.

The decomposition of a joint distribution into the association and the marginals is also useful to distinguish measures of *associations* from measures of *dependence*: the former should not depend on the marginals but the latter may. Furthermore, in Bayesian statistics a deeper understanding of the association between an observation Y and a random parameter Θ (taking the role of X) is fundamental and can be studied through our approach (cf. van der Linde 2002b).

The purpose of this paper is to define the association structure *Assoc(X,Y)*, which turns out to be an odds ratio *function*, establish its basic properties and provide parametrizations of association models as a formal framework for statistical analysis. More detailed statistical applications of our results are given in van der Linde (2002ab) or van der Linde and Osius

(2003). Before outlining the present work we look at two special cases which motivate the general approach.

For *simple* random variables $X \in \{0, \ldots, J\}$ and $Y \in \{0, \ldots, K\}$ (i.e. with finite range) the joint distribution is given by $\pi_{jk} = P(X = i, Y = k)$ for all j, k And the $J \times K$ odds ratio (or cross product ratio) matrix θ with entries $\theta_{jk} = (\pi_{jk}\pi_{00})(\pi_{j0}\pi_{0k})^{-1}$ for $j, k \geq 1$ is known to be the wanted object *Assoc(X Y)*, see e.g. Plackett (1974, Sec.3.4). The result, that for any matrix θ with positive entries and given marginal distribution there exists a unique joint distributions π may be obtained in two ways, both of which will be generalized here. First, the joint distribution π may be derived through a method used by Sinkhorn (1967) as a limit of iteratively rescaled distributions π_n with the given odds ratio matrix θ. Using a geometrical description of this iterative proportional fitting procedure, Fienberg (1970) gave a different proof of its convergence. Some applications of this procedure to data sets as well as extensions to multidimensional tables can be found e.g. in Bishop, Fienberg and Holland (1975). - Second, the joint distribution π may be obtained from result due to Haberman (1974, Theorem 2.6) as the unique argument maximizing a strictly concave function arising from the log-likelihood of Poission distributions.

Now let $Y \in \{0, 1\}$ be an indicator (e.g. for a disease) and X a random vector. The *odds ratio* of an input value x with respect to a reference value x°

$$OR^\circ(x) = \frac{P(Y = 1|X = x)}{P(Y = 0|X = x)} \bigg/ \frac{P(Y = 1|X = x^\circ)}{P(Y = 0|X = x^\circ)}$$

is a fundamental concept in epidemiology and we now briefly show that the corresponding *odds ratio function* OR° is the desired association object *Assoc(X,Y)*. Using the log odds ratio function $\psi^\circ = \log OR^\circ$, the conditional distribution $\mathscr{L}(Y|X = x)$ is determined by the conditional probability

$$P(Y = 1|X = x) = \{1 + a^{-1}\exp[-\psi^\circ(x)]\}^{-1} \quad \text{resp.} \tag{1}$$
$$\text{logit } P(Y = 1|X = x) = \alpha + \psi^\circ(x),$$

where $\text{logit} p = log[p/(1-p)]$ denotes the logistic transformation and $\alpha = \log a = \text{logit } P(Y = 1|X = x^\circ)$. The marginal distribution of Y is given by

$$P(Y = 1) = E(\{1 + a^{-1}\exp[-\psi^\circ(X)]\}^{-1}) \tag{2}$$

which is a strictly increasing function of a. Hence for $0 < P(Y = 1) < 1$ there exists a unique $0 < a < \infty$ such that (2) holds. This shows that for fixed marginal distributions the joint distribution is uniquely determined by the log odds ratio function ψ°. Furthermore, for fixed marginal distributions and a given log odds ratio function ψ° a joint distribution is defined by (1) with a obtained from (2).

Hence the logistic regression model (1) does not restrict the marginal distribution of Y. Thus any model for the odds ratio function, e.g. the $\psi^\circ(x) = (x - x^o)^T\theta$ with parameter vector θ, represents a semi-parametric distribution model in the sense above (the superscript "T" denotes the transpose). As a consequence - which is well known in epidemiology - statistical inference about the odds ratio parameter θ may be based either on cohort studies (where sampling is conditional on X) or on case-control studies (with sampling conditional on Y), whichever are easier to conduct in the particular application. These results will be extended to the logistic regression model for (non-binary) Y with finite range (cf. **4.4**) but the proofs are substantially more complicated. However the major statistical benefit of our approach lies in a further generalization to semi-parametric association models for random *vectors* – or even arbitrary *random elements* – X and Y.

Although our applications in section **4** deal only with random *vectors*, we will derive the main results for arbitrary *random elements* (including e.g. random *functions*) X and Y for two reasons. First, our arguments do not exploit specific properties of *finite-dimensional* euklidean space and second, a restriction to random *vectors* would not even simplify our proofs. However the reader may prefer to think of familiar random vectors instead of random elements.

We now outline the basic concepts and results for the general case, where X and Y are random elements. The wanted association object *Assoc(X,Y)* turns out to be a straightforward generalization of the above odds ratio function. Given a *positive* density p of the joint distribution with respect to some dominating product measure $\nu_X \times \mu_Y$ of σ-finite measures the odds ratio function with respect to a reference pair (x°, y°) is defined as

$$OR^\circ(x,y) = \frac{p(X=x, Y=y)}{p(X=x, Y=y^\circ)} \Big/ \frac{p(X=x^\circ, Y=y)}{p(X=x^\circ, Y=y^\circ)}$$

where the *joint* density p can equivalently be replaced by the *conditional* density of either Y given X or conversely. Furthermore the function OR° is invariant under a change of the dominating measure and a natural choice for the dominating measure is the product of the marginal distributions. The formal definition of the odds ratio function and its elementary properties are given in section **2**. Within the important class of joint distributions having *integrable* log-densities a *centered* log odds ratio function ψ is defined, which does not refer to reference values x° and y°.

In section **3** we show that the odds ratio function characterizes the association and hence may be taken as the desired association object *Assoc(X,Y)*. Using the Kullback-Leibler information we first prove (under mild integrability conditions) that the joint distribution is *uniquely determined* by the marginal distributions and the odds ratio function. To establish the *existence* of a joint distribution with given fixed marginals and an odds ratio function requires a greater effort and stronger additional assumptions. As our main result the wanted joint distribution will be obtained from a convergence theorem as the limit of a marginal fitting sequence of densities, which generalizes the approach of Sinkhorn (1967) for simple random variables. Furthermore the joint density may also be obtained by maximizing a strictly concave function which corresponds to the log-likelihood used by Haberman (1974) and is closely related to Kullback-Leibler information. Although the concepts here are straightforward generalizations derived from the work of Sinkhorn and Haberman, their approaches exploit unique features of *finite* distributions which are not available for non-simple random variables.

The benefit of the theoretical results for statistical modelling is adressed in section **4**. First *association models* are introduced which only specify the odds ratio function, but not the marginal distributions. Explicit representations of the joint and the conditional densities are given in terms of odds ratios and nuisance parameters. From the existence theorem we conclude that these association models are *semi-parametric*, because any set of marginal distributions for X and Y may be achieved by a proper choice for the nuisance parameters. Removing restrictions on the marginals from standard (regression) models yields an important class of *log-bilinear association models*. This is illustrated for joint and conditional *normal* distribution models used in multivariate analysis for random vectors X and Y. Furthermore we establish that joint multivariate normality holds if and only if both marginal distributions are normal and the association is *log-bilinear*. Hence log-bilinear association models are semi-parametric generalizations of multivariate normal distribution models. Statistical aspects of these results are pursued further in van der Linde (2002a) and van der Linde and Osius (2002).

An important feature of the approach here is a *symmetry* in presentation between X and Y: by interchanging X with Y any concept or argument entails its *dual*.

2 The Odds ratio function

To formalize the introductary discussion we consider arbitrary non-empty probability spaces $(\Omega_X, \mathscr{B}_X, \pi_X)$ and $(\Omega_Y, \mathscr{B}_Y, \pi_Y)$ as well as their product $(\Omega, \mathscr{B}, \pi)$, i.e. the set $\Omega = \Omega_X \times \Omega_Y$ equipped with the product measure $\pi = \pi_X \times \pi_Y$. Let $\mathscr{P}$ denote the class of probability measures P on $(\Omega, \mathscr{B})$ which have a *positive* density $f > 0$ with respect to π, i.e. P is dominated by π and dominates $\pi : P \ll \pi \ll P$. Further let $\mathscr{F}$ be the class of corresponding densities, i.e. the Radon-Nikodym derivatives $f = dP/d\pi$ for any $P \in \mathscr{P}$.

The restriction to *positive* densities is essential for the definition of the odds ratio function and only rules out less interesting joint distributions. For *simple* random variables X and Y the condition $P \in \mathscr{P}$ is equivalent to

$$P(X = i), P(Y = k) > 0 \quad \Rightarrow \quad P(X = i, Y = k) > 0 \quad \text{for all}\, i, k.$$

And if, for example, P is a multivariate *normal* distribution $N_k(\mu, \Sigma)$, then the condition $P \in \mathscr{P}$ holds if and only if the rank of the covariance matrix Σ equals the sum of the ranks for both marginal covariance matrices (cf. Osius 2000, **4.4**).

2.1 Odds and log odds ratio function

For any joint distribution $P \in \mathscr{P}$ with density $f \in \mathscr{F}$ the *odds ratio function*, denoted by OR_f resp. $OR(f)$ or $OR(P)$, is a map $\Omega \times \Omega \longrightarrow (0, \infty)$ defined by

$$OR_f(x, y|x', y') = [f(x, y) \cdot f(x', y')] \cdot [f(x, y') \cdot f(x', y)]^{-1}$$

The *log odds ratio function* $\log OR_f$ depends only on the log-density $\varphi = \log f$ and will also be denoted by $\psi_\varphi = \log OR_f$, i.e.

$$\psi_\varphi(x, y|x', y') = \varphi(x, y) + \varphi(x', y') - \varphi(x, y') - \varphi(x', y).$$

For any pair of *fixed reference values* $x^\circ \in \Omega_X$ and $y^\circ \in \Omega_y$ the odds resp. log odds ratio function is already determined by its partial function $OR_f^\circ = OR_f(\text{-}, \text{-}\,|\, x^\circ, y^\circ) : \Omega \to (0, \infty)$ resp. $\psi_\varphi^\circ = \log OR_f^\circ$ since

$$\psi_\varphi(x, y|x', y') = \psi_\varphi^\circ(x, y) + \psi_\varphi^\circ(x', y') - \psi_\varphi^\circ(x, y') - \psi_\varphi^\circ(x', y).$$

Now for any $P \in \mathscr{P}$ its odds resp, log odds ratio function is defined as OR_f resp. ψ_φ for $f = dP/d\pi \in \mathscr{F}$ resp. $\varphi = \log dP/d\pi$. Strictly speaking, the odds ratio function of P is only unique modulo the equivalence relation of π-almost-sure equality, i.e. g and h are equivalent if and only if $g = h$ π-almost surely.

It is convenient to view any $P \in \mathscr{P}$ as a *joint distribution* of a pair (X, Y) of random elements defined on a probability space $(\Omega_0, \mathscr{B}_0, P_0)$ with values in Ω, i.e. P is the image measure of P_0 under the mapping $(X, Y) : \Omega_0 \rightarrow \Omega$. Following usual practice, we extend concepts defined for *probability measures to random elements* via their distribution, e.g. the odds ratio function for (X, Y) is that of their joint distribution: $OR(X, Y) = OR(P)$.

2.2 Marginal and conditional distributions

If $f \in \mathscr{F}$ is the joint density of (X, Y) then the *marginal* densities of X and Y are denoted by $f^X(x) = \int f(x,y)\, d\pi_Y(y) > 0$ and $f^Y(y) = \int f(x,y)\, d\pi_X(x) > 0$. Since f^X and f^Y are finite almost surely we may choose f such that f^X and f^Y are both finite. The *conditional* density $f^{|X} \in \mathscr{F}$ of Y given X - defined by $f^{|X}(y|x) = f(x,y)/f^X(x)$ - is positive and evidently has the *same* odds ratio function as the unconditional density f. This also holds for the *conditional* density $f^{|Y} \in \mathscr{F}$ of X given Y, and hence $OR(f^{|X}) = OR(f) = OR(f^{|Y})$.

2.3 Change of dominating measures

An important property of the odds ratio function is its invariance with respect to dominating σ-finite measures. More precisely, suppose that the marginal distribution π_X resp. π_Y is dominated by a σ-finite measure v_X resp. v_Y with positive density $\delta_X = d\pi_X/dv_X$ resp. $\delta_Y = d\pi_Y/dv_Y$. Then $P \in \mathscr{P}$ is dominated by the product measure $v = v_X \times v_Y$ and $f_\delta(x,y) = f(x,y)\delta_X(x)\delta_Y(y)$ defines a v-density of P. Hence the odds ratio function may also be expressed with f replaced by f_δ

$$OR_f(x,y|x',y') = [f_\delta(x,y) \cdot f_\delta(x',y')] \cdot [f_\delta(x,y') \cdot f_\delta(x',y)]^{-1}$$

For common dominating measures - e.g. Lebesgue's resp. the counting measure for continuous resp. simple random variables - this representation is typically used to *define* the odds ratio. However the product measure π of the marginal distributions is a *canonical* choice for a dominating measure and moreover allows a definition of the odds ratio in situations where no densities with respect to the above standard measures are available (e.g. for a multivariate normal distribution with singular covariance matrix, cf. **4.4**).

2.4 One-to-one transformations

The odds ratio of one-to-one transformations $U = g(X)$ and $V = h(Y)$ are easily obtained as the odds ratio of X and Y evaluated at the corresponding inverse points. More precisely, let $g : \Omega_X \longrightarrow \Omega_U$ resp. $h : \Omega_Y \longrightarrow \Omega_V$ be measurable (with respect to some σ-algebras $\mathscr{B}_U$ resp. $\mathscr{B}_V$) having measurable inverses g^{-1} resp. h^{-1}. Consider the space $\Omega' = \Omega_U \times \Omega_V$ equipped with the product $\pi' = \pi_U \times \pi_V$ of the distributions π_U and π_V of U and V and the corresponding product σ-algebra $\mathscr{B}'$. Then a positive π'-density of the joint distribution P' of *(U,V)* is given by $f'(u,v) = f(g^{-1}(u), h^{-1}(v))$, and hence the odds ratio function of *(U,V)* is

$$OR_{f'}(u,v|u',v') = OR_f(g^{-1}(u),h^{-1}(v)|g^{-1}(u'),h^{-1}(v')).$$

Random vectors. A typical example for transformations of random vectors X and Y are affine mappings $g(x) = A(x-b)$ and $h(y) = C(y-d)$ with appropriate quadratic non-singular matrices A, C and vectors b, d. If the covariance matrices Σ_X and Σ_Y of X and Y are non-singular, then the choice of $A = \Sigma_X^{-1/2}$, $b = E(X)$, $C = \Sigma_y^{-1/2}$, $d = E(Y)$ yield *standardized* random vectors U and V (i.e. with zero expectation and identity covariance matrix). Hence the odds ratio function of (X, Y) is easily obtained from those of the *standardized pair* (U, V).

2.5 Decomposition of densities

For a joint density $f \in \mathscr{F}$ its logarithm $\varphi = \log f$ may be written as

$$\begin{aligned}
&\varphi(x,y) = \alpha^\circ + \beta^\circ(x) + \gamma^\circ(y) + \psi^\circ(x,y) \qquad \text{with}\\
&\alpha^\circ = \varphi(x^\circ,y^\circ), \quad \beta^\circ(x) = \varphi(x,y^\circ) - \varphi(x^\circ,y^\circ),\\
&\gamma^\circ(y) = \varphi(x^\circ,y) - \varphi(x^\circ,y^\circ).
\end{aligned} \tag{3}$$

and $\psi^\circ = \log OR_f^\circ$ as the (partial) log odds ratio function with respect to (x°, y°). Let $\mathscr{L}^0$ denote the vector space of all random variables from Ω into $\mathbb{R}$. Then any log odds ratio function ψ° belongs to the linear subspace

$$\mathscr{A}^0 = \{\xi \in \mathscr{L}^0 | \xi(x^\circ,y) = 0, \xi(x,y^\circ) = 0 \quad \text{for all}\, x,y\}$$

Furthermore let $\mathscr{L}^1$ resp. $\mathscr{L}_X^1$, $\mathscr{L}_Y^1$ denotes the vector space of all random variables from Ω resp. Ω_X, Ω_Y into $\mathbb{R}$ which are integrable with respect to π resp. π_X, π_Y. For any $\varphi \in \mathscr{L}^1$ the *marginal functions* φ^X resp. φ^Y on Ω_X resp. Ω_Y defined by

$$\varphi^X(x) = \int \varphi(x,y)d\pi_y(y), \quad \varphi^Y(y) = \int \varphi(x,y)d\pi_X(x),$$

give rise to another decomposition

$$\begin{aligned}
&\varphi(x,y) = \alpha + \beta(x) + \gamma(y) + \psi(x,y) \qquad \text{with}\\
&\alpha = \int \varphi\, d\pi, \quad \beta(x) = \varphi^X(x) - \int \varphi\, d\pi, \qquad \gamma(y) = \varphi^Y(y) - \int \varphi d\pi,\\
&\psi(x,y) = \varphi(x,y) - \varphi^X(x) - \varphi^Y(y) + \int \varphi\, d\pi.
\end{aligned} \tag{4}$$

Note that ψ does not depend upon the point (x°, y°) and belongs to the subspace

$$\mathscr{A}^1 = \{\xi \in \mathscr{L}^1 | \xi^X = 0, \xi^Y = 0\}.$$

The function ψ is already determined by ψ°- and vice versa

$$\begin{aligned}
&\psi(x,y) = \psi^\circ(x,y) - \psi^{\circ X}(x) - \psi^{\circ Y}(y) + \int \psi^\circ d\pi\\
&\psi^\circ(x,y) = \psi(x,y) - \psi(x,y^\circ) - \psi(x^\circ,y) + \psi(x^\circ,y^\circ).
\end{aligned} \tag{5}$$

ψ is called the *centered log odds ratio function of f resp.* φ.

3 A characterization of association

This section contains our main result, the characterization of association in terms of the odds ratio function under mild conditions. First we will argue that any joint distribution $P \in \mathscr{P}$ of (X, Y) with given marginal distributions π_X and π_Y is uniquely determined by its odds ratio function *OR(P)*. Conversely, for any π-integrable $\psi^\circ \in \mathscr{A}^\circ$ we provide sufficient conditions for the existence of a π-integrable log-density $\varphi = \log(dP/d\pi)$ such that the joint distribution P has the pre-specified marginal distributions π_X and π_Y and ψ° as its log odds ratio function. The results of uniqueness and existence imply that the odds ratio function completely characterizes the association, i.e. the information in the joint distribution which is not contained in the marginal distributions. For this reason the odds ratio function will also be referred to as the *association function*

3.1 Uniqueness

The results on uniqueness are summarized as follows.

Uniqueness-Theorem: *suppose $P_1, P_2 \in \mathscr{P}$ are joint distributions both with common marginal distributions π_X resp. π_Y and with a common odds ratio function $OR^\circ(P_1) = OR^\circ(P_2)$ with respect to an arbitrary reference pair (x°, y°). Let $\varphi_i = log(dP_i/d\pi)$ denote the log-densities for i= 1,2. Then any of the two integrability conditions implies $P_1 = P_2$:*

$$(i) \quad \int |\varphi_1(x, y^\circ) - \varphi_2(x, y^\circ)| d\pi_X(x) < \infty,$$

$$(ii) \quad \int |\varphi_1(x^\circ, y) - \varphi_2(x^\circ, y)| d\pi_Y(y) < \infty.$$

Note that (i) resp. (ii) always hold if Ω_X resp. Ω_Y *is finite.*

Corollary: *If the log-densities φ_1, φ_2 are π-integrable, then there exists (x°, y°) such that (i) and (ii) hold.*

We only sketch the proof and refer to Osius (2000) for details. Using the decomposition (3) for φ_1 and φ_2, the *Kullback-Leibler information* (Kullback 1959)

$$I(f_1, f_2) := \int f_1 \log(f_1/f_2) d\pi = \int (\varphi_1 - \varphi_2) dP_1$$

can shown to zero, which implies $P_1 = P_2$. ■

3.2 Existence

For a given function $\psi^\circ \in \mathscr{A}^\circ$ we now wish to construct a joint distribution P with given marginal distribution π_X and π_Y such that ψ° is the log odds ratio function of P. The construction of P requires certain restrictions on ψ°, the first being that ψ° is π-*integrable*, which in turn will yield a π-integrable log-density φ. Passing from ψ° to the centered version ψ given by (5), we may alternatively start with a given function $\psi \in \mathscr{A}^1$ and look for a log-density φ with *centered* log odds ratio function ψ. We will construct the wanted distribution P under the following *existence conditions*.

(EC1) *There exists a* π_X*–integrable* $\bar{\beta} : \Omega_X \longrightarrow \mathbb{R}$ *such that* $\exp(\bar{\beta} + \psi)$ *is* π*–integrable.*

(EC2) *There exists a* π_Y*–integrable* $\bar{\gamma} : \Omega_X \longrightarrow \mathbb{R}$ *such that* $\exp(\bar{\gamma} + \psi)$ *is* π*–integrable.*

Equivalent formulations in terms of the marginal functions of $q = \exp \psi$ are as follows (cp. Osius 2000):

(EC1)′ $\log q^X = \log\left[(\exp\,\psi)^X\right]$ *is* π_X*–integrable.*

(EC2)′ $\log q^Y = \log\left[(\exp\,\psi)^Y\right]$ *is* π_Y*–integrable.*

A stronger version of both (EC1) and (EC2) is

(EC3) $\exp(\psi)$ *is* π*–integrable.*

This condition trivially holds for *bounded* ψ, which covers an important range of applications in practice where the sample space Ω typically is a *compact* subset of $\mathbb{R}^k$ and ψ is a *continuous* function. Note that for *finite* Ω_X resp. Ω_Y the condition (EC3) is even equivalent to (EC1) resp. (EC2). However, for $\Omega = \mathbb{R}^k$ we will provide an example for π (cf. **4.4**), where (EC) fails but (EC1) and (EC2) hold.

The conditions (EC1) and (EC2) are *sufficient* for our existence-result.

Existence-theorem: *For any* $\psi \in \mathscr{A}^1$ *the conditions (EC1) and (EC2) imply the existence of a joint distribution* $P_\psi \in \mathscr{P}$ *with given marginal distributions* π_X *and* π_Y *and a* π*-integrable log-density* $\varphi_\psi = \log(dP_\psi/d\pi)$ *having* ψ *as its centered log odds ratio function.*

Although the existence conditions will turn out (cf.**4.3**) to be weak enough for important applications, they are not *necessary* for the existence of P_ψ - at least not for *binary* Y (cf.**1**).

The desired density of P_ψ will be constructed as a limit of an iterative procedure of so-called *marginal fittings* to be explained first. For any π-density f we can adjust its marginal distribution of X to the wanted marginal π_X by passing to the density $f^{|X}(x,y) = f(x,y)/f^X(x)$. The marginal density of $f^{|X}$ is the constant 1 and hence the corresponding distribution has π_X as its marginal of X. Note, that $f^{|X}$ has the same odds ratio function as f. similiarly we can adjust f to the wanted marginal π_Y by passing to $f^{|Y}(x,y) = f(x,y)/f^Y(y)$. Unfortunately we can not adjust *both* marginals in a single step. An iterative procedure however, will do the job.

To construct an iterative sequence of densities we start with any density $f_0 \in \mathscr{F}$ with π-integrable $\varphi_0 = \log f_0$ and ψ as its centered log odds ratio function, e.g. a normalization of $\exp(\bar{\beta} + \psi)$ provided by (EC1). Given f_n, we successively adjust both marginals and obtain $f_{n+1} = (f_n^{|X})^{|Y}$. From the convergence theorem below we conclude that he sequence (f_n) converges pointwise to a limit f_ψ which is the density of the desired distribution P_ψ. As we will see later, the density f_ψ of the wanted joint distribution P_ψ may also be obtained by minimizing the Kullback-Leibler information $I(1,f) = -\int \log(f) d\pi$ with respect to all densities f having a π-integrable log-density $\varphi = \log f$ and ψ as its centered log odds ratio function. Note that the constant density 1 represents the joint distribution π under which X and Y are independent. For any such density f its logarithm φ is uniquely determined by

$$\eta = \varphi - \psi - \bar{\beta} - \bar{\gamma} \qquad \text{via} \qquad \varphi = \eta + \bar{\beta} + \bar{\gamma} + \psi.$$

with (fixed) functions $\bar{\beta}$, $\bar{\gamma}$ given by (EC1), (EC2) and chosen such that additionally

$$\int \bar{\beta}\, d\pi_X = 0, \qquad \int \bar{\gamma}\, d\pi_Y = 0$$

hold. By (4) the function η can be decomposed as a sum

$$\eta(x,y) = \eta_X(x) + \eta_Y(y),$$

and hence η belongs so-called the *marginal subspace*

$$\mathscr{M}^1 = \{\eta_X + \eta_Y | \eta_X \in \mathscr{L}_X^1, \eta_Y \in \mathscr{L}_Y^1\} \subset \mathscr{L}^1.$$

Now, minimizing $I(1,f)$ is equivalent to maximizing the functional $\ell : \mathscr{M}^1 \longrightarrow [-\infty, \infty)$ defined by

$$\begin{aligned} \ell(\eta) &= \int \left[\eta - \exp(\eta + \bar{\beta} + \bar{\gamma} + \psi)\right] d\pi \\ &= \int [\eta - f] d\pi = -I(1,f) - \int \psi d\pi - 1. \end{aligned}$$

ℓ is bounded from above by $\ell(-\bar{\beta} - \bar{\gamma} - \psi)$ and strictly convave modulo π-almost-sure equality on the convex set $\{\eta | \ell(\eta) \text{ is finite}\}$. The functional ℓ generalizes the function considered by Haberman (1974, Theorem 2.6) for (conditional) Poisson distributed Y.

Now we provide a convergence result for the iterated marginal fitting procedure from which the existence theorem will be derived.

Convergence-Theorem: *Let $(f_n \in \mathscr{F})_{n \geq 0}$ be any marginal fitting sequence of densities, i.e. $f_{n+1} = (f_n^{|X})^{|Y}$, such that the log-densities $\varphi_n = \log f_n$ are π-integrable. Furthermore, put $\eta_n = \varphi_n - \psi - \bar{\beta} - \bar{\gamma} \in \mathscr{M}^1$ for all n.*

If the existence conditions (EC1) and (EC2) hold, then any subssequence $(\eta_{m(n)})$ of (η_n) contains a further subsequence $(\eta_{m'(n)})$ which converges pointwise as well as in the mean to an element $\eta \in \mathscr{M}^1$ such that $f = exp(\eta + \bar{\beta} + \bar{\gamma} + \psi)$ is the π-density of a joint distribution $P \in \mathscr{P}$ with marginals π_X and π_Y.

The limit is π-almost surely independent of the chosen subsequence $(\eta_{m(n)})$ and independent of f_0, i.e. for two starting values f_{10}, f_{20} and any subse-

quences $(\eta_{1m(n)}), (\eta_{2m(n)})$ *the corresponding limits* η_1 *and* η_2 *coincide* π*-almost surely.*

Corollary: *For any starting value* $f_0 \in \mathscr{F}$ *with* $\eta_o \in \mathscr{M}^1$ *the sequence* $(\eta_n)_n \geq 0$ *converges in the mean to an element* $\eta \in \mathscr{M}^1$ *and* $(\ell(\eta_n))_{n\geq 0}$ *is a non-decreasing sequence with limit* $\ell(\eta)$*. Furthermore* η *is the* π*-almost surely unique argument maximizing the functional* ℓ*.*

For a proof we refer to Osius (2000) and only sketch the main steps here. First, *integrable bounds* for (η_η) and related sequences are derived, which imply that these sequences belong to *compact* sets of functions (equipped with the topology of pointwise convergence). Hence a subsequence $(\eta_{m'(n)})$ can be found which converges pointwise and (since the bounds are integrable) also in the mean to an element $\eta \in \mathscr{M}^1$ with the wanted properties. The independence of the limit with respect to the starting value f_0 and the subsequence $(\eta_{m(n)})$ uses the uniqueness-theorem.

To complete the proof of the existence-theorem, we apply the above corollary. It remains to show all that φ_n are integrable provided φ_0 is. This follows since $\log f^X$ and $\log f^Y$ are integrable if $\log f$ is (by Jensen's inequality).

4 Applications

The separation of the association from the marginals has an impact on statistical modelling which will now be illustrated. First, semi-parametric *association models* are introduced, which specify a model for the association function of X and Y but leave the marginal distributions completely arbitrary. Parametrizations of the joint and conditional densities in terms of odds ratios are given which serve as a framework for further statistical analysis. Motivated by standard (regression) models we specialize to *log-bilinear associations* for which the existence conditions (EC1) and (EC2) are restated in terms of cumulant-generating functions. Second, we look at random *vectors* X and Y and characterize the *joint* normal distributions for (X,Y) as those with normal marginals and a *log-bilinear association.* This emphasizes the importance of log-bilinear association models as a semi-parametric generalization of joint multivariate normality.

4.1 Association Models

From the uniqueness theorem we conclude that the joint distribution P of (X, Y) is determined by the marginal distributions π_X and π_Y and their association, i.e. the odds ratio function OR. If the focus of an investigation is on the association between X and Y rather than on the marginal distributions, then the appropriate models are semi-parametric *accociation models* or *odds ratio models,* which only specify the odds ratio function and leave the marginals completely arbitrary. The corresponding model for the density f with respect to the product $\pi = \pi_X \times \pi_Y$ of the marginals (or any other product measure, cf. **2.3**) may be written in terms of the log-density $\varphi = \log f$ and the log odds ratio $\psi^\circ = \log OR$ using (3) as

$$\varphi(x,y) = \alpha^\circ + \beta^\circ(x) + \gamma^\circ(y) + \psi^\circ(x,y).$$

Here ψ° is restricted to a subspace $\Psi^\circ \subset \mathscr{A}^\circ$ specifying the model and $\alpha^\circ \in \mathbb{R}$ as well as the functions β° and γ° are completely arbitrary. Identifiability may be achieved through the constraints $\beta^\circ(x^\circ) = 0$ and $\gamma^\circ(y^\circ) = 0$, which will be assumed here. Note that the definition of $\mathscr{A}^\circ$ already imposes the constraints $\psi^\circ(x, y^\circ) = 0 = \psi^\circ(x^\circ, y)$ for all x, y. The model space Ψ° is typically parametrized by means of a parameter $\theta \in \Theta$, i.e. $\Psi^\circ = \{\psi^\circ_\theta | \theta \in \Theta\}$. Assuming the log-density ψ to be π-integrable, the model can be rewritten using (4) as

$$\varphi(x,y) = \alpha + \beta(x) + \gamma(y) + \psi(x,y)$$

with ψ restricted to a subspace $\Psi \subset \mathscr{A}^1$, arbitrary $\alpha \in \mathbb{R}$ and functions β and γ which are integrable with respect to π_X resp. π_Y. An important point however is, that allowing arbitrary α, β and γ is no guarantee that the model is semi-parametric, i.e. does not restrict the marginal distributions π_X and π_Y. This, in fact, requires the existence theorem which explicitly states (under the existence conditions) that any given marginal distributions may be obtained for suitable values of the parameters α, β and γ. Note that X and Y are *independent* under P, i.e. $P = \pi_X \times \mu_Y$, if and only if $\psi^\circ = 0$ resp. $\psi = 0$.

In statistical applications the model is often equivalently specified using the *conditional* density $f^{|X}(y|x)$ of Y given X by

$$\log f^{|X}(y|x) = \beta^\circ_X(x) + \gamma^\circ(y) + \psi^\circ(x,y)$$

with $\beta^\circ_X(x) = -\log \int \exp[\gamma^\circ(y) + \psi^\circ(x,y)] d\pi_Y(y)$, or by the log-density *ratio*

$$\log\left(f^{|X}(y|x)/f^{|X}(y^\circ|x)\right) = \gamma^\circ(y) + \psi^\circ(x,y). \tag{6}$$

The dual formulation in terms of the conditional density of X given Y is

$$\log\left(f^{|Y}(x|y)/f^{|Y}(x^\circ|y)\right) = \beta^\circ(x) + \psi^\circ(x,y).$$

The major advantage of association models is that statistical inference concerning the odds ratio function (or its parameter θ) may be drawn from a sample of independent observations $(x_1, y_1), ..., (x_n, y_n)$ where each (x_i, y_i) may be taken from any of the two conditional distributions $\mathscr{L}(Y|X = x_i)$ and $\mathscr{L}(X|Y = y_i)$ or from the joint distribution $\mathscr{L}(X, Y)$.

Returning to the discussion in the introduction, we first specialize to a *finite* sample space Ω_Y and then briefly look at *log-bilinear association models*. However, statistical inference for these models is outside the scope of this paper and the reader is referred to van der Linde (2002ab) or van der Linde and Osius (2003) for detailed applications.

4.2 Output with Finite Range

Let Ω_Y be finite, say $\Omega_Y = \{0, 1, ..., K\}$. Then $\mathscr{L}(Y|X = x)$ has a multinomial distribution $M_{K+1}(1, \pi(x))$ with $K + 1$ classes and probabilities $\pi_k(x) = P(Y = k|X = x) > 0$. Using the multivariate logistic transformation $\operatorname{logit} \pi_k(x) = \log(\pi_k(x)/\pi_0(x))$ of $\pi(x)$, the association model (6) with $y^\circ = 0$ is equivalent to a *logistic regression model,*

$$\operatorname{logit} \pi_k(x) = \gamma^\circ_k + \psi^\circ_k(x), \quad k = 1, \ldots, K,$$

where the argument y is replaced by an index k. Although this regression model is widely used, it has not yet been emphasized (or proved) that the model does *not* restrict the marginal distribution of Y and thus represents an *association* model.

In a *linear* logistic regression model the log odds ratio functions ψ_k° are taken as linear functions of the form $\psi_k^\circ(x) = g(x)^T\theta_k$, where $g(x)$ is an S-dimensional vector of so called covariables and $\theta_k \in \mathbb{R}^S$ is an unknown parameter. Although typically the observation $x = (x_1, \ldots, x_t)$ is itself a finite-dimensional vector, the use of a transformation $g(x)$ instead of x provides more flexible models. For example, $g(x)$ may contain powers $x_1, x_1^2, \ldots$ of a continuous component x_1 as well as indicator variables $I\{x_2 = l\}$ for levels $l = 1, \ldots, L$ of a discrete component x_2. Introducing indicator variables $h_k(y) = I\{y = k\}$ for all values k of Y, the model $\psi_k^\circ(x) = g(x)^T\theta_k$ may equivalently be written as

$$\psi^\circ(x,y) = \sum_k g(x)^T\theta_k h_k(y) = g(x)^T\theta h(y)$$

where θ is the corresponding $S \times K$ matrix with columns $\theta_1, \ldots, \theta_K$. This representation serves as a motivation for the association models considered next.

4.3 Log-bilinear association models

Returning from finite to arbitrary Ω_Y, let $U = g(X)$ and $V = h(Y)$ be random vectors given by measurable maps $g : \Omega_X \to \mathbb{R}^{k_x}$ and $h : \Omega_Y \to \mathbb{R}^{k_y}$. In a semi-parametric *log-bilinear association model* for (X, Y) with respect to (g, h) the log odds ratio function for the joint distribution of (X, Y) is of the form

$$\text{(LBA)} \qquad \psi^\circ(x,y) = [g(x) - g(x^\circ)]^T A[h(y) - h(y^\circ)] \qquad \text{for all } x, y,$$

with a matrix A of parameters, which may additionally be restricted to a suitable subspace of $k_x \times k_y$ matrices. The reference values and the functions g, h will typically be chosen such that $g(x^\circ) = 0$ and $h(x^\circ) = 0$ thus making ψ° a bilinear function in the transformed variables $g(x)$ and $h(y)$.

The semi-parametric model (LBA) includes important regressions models, namely *Generalized Linear Models* with canonical links, i.e. linear resp. logistic or log-linear regression where the *conditional* distribution $\mathscr{L}(Y|X)$ is univariate normal resp. binomial or Poisson. Furthermore, we will show in 4.4 that (LBA) always holds if the *joint* distribution of U and V is multivariate *normal*, which serves as a basic model in multivariate analysis. Hence (LBA) generalizes several important statistical models – by removing restrictions of the marginal distributions – and may be considered a standard model for association. Further exploration of these ideas are given in van der Linde (2002ab) or van der Linde and Osius (2003).

The existence condition for ψ° can be restated in terms of the cumulant generating function of V resp. U. First, the centered log odds ratio function can be expressed in terms of the expectations $\mu_U = E(U)$ and $\mu_V = E(V)$ (which are assumed to exist) using (5) as

$$\psi(x,y) = \psi^\circ(x,y) - \psi^{\circ X}(x) - \psi^{\circ Y}(y) + \int \psi^\circ d\pi$$
$$= [g(x) - \mu_U]^T A[h(y) - \mu_V].$$

Next the marginal function q^X of $q = \exp(\psi)$ can be computed via the moment generating function m_V of V by

$$q^X(x) = m_V(A^T[g(x) - \mu_U]) \times \exp\{-[g(x) - \mu_U]^T A\mu_V\}.$$

Finally, we get $\log(q^X)$ in terms of the cumulant generating function $k_V = \log m_V$ of V as

$$\log q^X(x) = k_V(A^T[g(x) - \mu_U]) - [g(x) - \mu_U]^T A\mu_V.$$

Hence the existence condition (EC1)′ for ψ° in (LBA) is equivalent to

$(\mathrm{EC1})_A$ *The expectation of* $\kappa_V(A^T[U - \mu_U])$ *exists (i.e. is finite).*

By duality, (EC2)′ can be stated in terms of the cubulant generating function κ_U of U. These conditions are easily checked for standard marginal distributions of U and V. As a first example, $(\mathrm{EC1})_A$ holds if V has a multivariate normal distribution $N(\mu_v, \Sigma_V)$ with $\kappa_V(t) = t^T\mu_V + 1/2\ t^T\Sigma_V t$, provided the covariance matrix of U exists. And second, suppose V has a (univariate) Poisson distribution with $\kappa_V(t) = \mu_V[exp(t) - 1]$. Then $(\mathrm{EC1})_A$ holds, if the value $m_U(A)$ of the moment generating function of U is finite.

4.4 Multivariate Normal Distributions

In this section we show that the joint distribution P of *(X,Y)* is multivariate normal if and only if the marginals are normal and the log odds ratio function is bilinear. First we consider the *conditional Gaussian model*

$$\text{(CGM)} \qquad \mathscr{L}(Y|X = x) = N_{k_y}(a + Bx, C) \qquad \text{for any } x,$$

with a vector a, a matrix B and a non-singular covariance matrix C (not depending upon x). From the (log-) conditional density with respect to Lebesgue's measure

$$\log f_\lambda^{|X}(y|x) = -\frac{1}{2}(y - a - Bx)^T C^{-1}(y - a - Bx) + const$$

we get the log odds ratio function of the joint distribution P as

$$\psi_P^\circ(x,y) = (x - x^\circ)^T A(y - y^\circ) \qquad \text{with} \qquad A = B^T C^{-1}$$

which is *bilinear* for the canonical choices $x^\circ = 0$ and $y^\circ = 0$.

Now suppose, that the joint distribution $P = N_k(\mu, \Sigma)$ is normal and assume for simplicity first, that Σ is *non-singular*. Then both marginals $\pi_X = N_{k_x}(\mu_x, \Sigma_x)$ and $\pi_Y = N_{k_y}(\mu_y, \Sigma_y)$ are normal and the conditional Gaussian model (CGM) holds with

$$B = \Sigma_{xy}^T \Sigma_x^{-1}, \qquad C = \Sigma_y - \Sigma_{xy}^T \Sigma_x^{-1} \Sigma_{xy}, \qquad \Sigma_{xy} = \mathrm{Cov}(X, Y).$$

Hence the log odds ratio function of a normal distribution P is bilinear (assuming $x^\circ = 0, y^\circ = 0$). - Conversely any joint distribution $P \in \mathscr{P}$ with a π-integrable log-density and normal marginals is *normal* if and only if its log odds ratio function is bilinear:

Characterization Theorem: *Let $P \in \mathscr{P}$ be a joint distribution with a π-integrable log-density $\varphi = log(dP/d\pi)$ and marginal normal distributions $\pi_X = N_{k_x}(\mu_x, \Sigma_x)$ and $\pi_Y = N_{k_y}(\mu_y, \Sigma_y)$ Then P is a multivariate normal distribution if and only if its log odds ratio function ψ° with respect to the reference values $x^\circ = 0$ and $y^\circ = 0$ is bilinear, i.e. there is a $k_x \times k_y$ matrix A such that $\psi^\circ(x, y) = x^T\ Ay$ for all x and y.*

Note: The existence conditions $(EC1)_A$ and $(EC2)_A$ hold if both π_x and π_y are normal.

We only sketch the proof for non-singular Σ_x, Σ_y and refer to Osius (2000) for details. If $P = N_k(\mu, \Sigma)$ holds, then Σ is non-singular (since $P \in \mathscr{P}$) and ψ° is bilinear by the above argument. - Conversely, let ψ° be bilinear. Starting with a normal distribution P_0 having ψ° as its log odds ratio function (but not yet the wanted marginals) the corresponding marginal fitting sequence converges to a normal distribution, which (by the uniqueness theorem) coincides with P. ■

The odds ratio function $\exp[\psi^\circ(x, y)] = \exp(x^T A y)$ is in general not π-integrable, i.e. condition (EC3) need not hold when (EC1) and (EC2) are satisfied (cp. Osius 2000).

5 Discussion

The joint distribution $\mathscr{L}(X, Y)$ of two random elements X and Y has been shown to be completely determined by their marginal distributions $\mathscr{L}(X), \mathscr{L}(Y)$ and an odds ratio function *OR(X,Y)* derivable from the joint or any of the two conditional densities. Specifying each of these three parts separately yields (under mild integrability conditions) a unique joint distribution, which can be obtained as a limit of iterative marginal fittings. Hence the odds ratio function characterizes the association between X and Y in the sense that it carries all information on the joint distribution which is not contained in the marginal distributions. Thus the odds ratio function may be taken as a formal definition of association for arbitrary random elements. Since association is a *symmetric* concept, any result within this framework entails its *dual* by interchanging X with Y.

The decomposition of a joint distribution into the association and the marginals is of importance whenever the association but not the marginals are of primary interest. *Association models* and corresponding parametrizations of the (conditional) densities given here provide a unifying formal framework for statistical studies on the relationship between X and Y based on *densities*. These models are semi-parametric in the sense that they do not restrict the marginal distributions. Hence statistical inference may be based on samples from the conditional distribution of either variable given the other.

We introduced the class of *log-bilinear* association models which generalizes important standard models (by removing restrictions on the marginals), e.g. *Generalized Linear Models* with canonical link. In particular the logistic regression model (without restriction on the intercept parameters) represents a log-bilinear association model and hence does not restrict the marginal distribution of Y. And for random vectors X and Y with normal *marginals* the joint distribution is multivariate normal if and only if their association is log-bilinear. Statistical applications of log-bilinear association models to dimension reduction and linear discriminant functions in multivariate analysis are given in van der Linde (2002a).

Finally we briefly indicate how the separation of association from the marginals may provide new perspectives in related areas not covered here. For instance, *measures* of *dependence* or *association* may be defined by taking marginal or joint expectations of the odds ratio $OR(X,Y)$ - or some function of it. Van der Linde (2002b) provides such applications in Bayesian statistics (where a *random* parameter Θ replaces X). An extensive coverage of multivariate dependencies based on *distribution functions* and related concepts (rather than odds ratios) is given by Joe (1997).

Furthermore our existence theorem allows *classifications* as well as *constructions* of joint distributions with given marginals by specifying or modelling the odds ratio function. Distributions with *fixed* marginals have been studied extensively (with different methods and perspectives) and the reader is referred to Rüschendorf et al (1996), Joe (1997) and the literature cited therein for details. In particular joint distributions $\mathscr{L}(X,Y)$ with specified marginals for *real-valued* X and Y have been constructed using the joint and marginal *distribution functions* F_{XY} and F_X, F_Y (rather than *densities* used here). Plackett (1965) introduced a family of bivariate distributions depending on a single parameter and the marginals. A more general approach is given by Sklar's theorem which provides a relation $F_{XY}(x,y) = C(F_X(x), F_Y(y))$ by means of a *copula* C, which is a function with certain properties. For details and further references on copulas we refer to Nelson (1999) and Joe (1997). Sklar's theorem extends to random vectors $X = (X_1, X_2 \ldots)$ and $Y = (Y_1, Y_2 \ldots)$ and gives their joint distribution function $F_{XY}(x_1, x_2 \ldots, y_1, y_2 \ldots) = C(F_{X_1}(x_1), F_{X_2}(x_2) \ldots, \quad F_{Y_1}(y_1), F_{Y_2}(y_2) \ldots)$ in terms of a multivariate copula C and the marginal distribution functions F_{X_i} resp. F_{Y_i} of the components X_i resp. Y_i. Now the copula C does not only capture the association between the vectors X and Y but also dependencies among the *components* $X_1, X_2 \ldots$ of X resp. $Y_1, Y_2 \ldots$ of Y. Hence in the multivariate case the copula C contains additional information about the (marginal) distribution of X resp. Y and does not represent the association between X and Y in the sense discussed here.

Acknowlegdment. The author thanks both referees for their helpful comments which improved the first draft of the paper.

References

Bishop YMM, Fienberg SE, Holland PW (1975) Discrete Multivariate Analysis Theory and Practice. MIT Press, Cambridge (Massachusetts)

Breslow NE, Day NE (1980) Statistical Methods in Cancer Research, Volume I: The Analysis of Case-Control Studies. International Agency for Research on Cancer, Lyon

Fienberg SE (1970) An iterative procedure for estimation in contingency tables. Annals of Statistics 41:907–917

Haberman SJ (1974) The Analysis of Frequency Data. The University of Chicago Press, Chicago

Joe H (1997) Multivariate Models and Dependence concepts. Chapman & Hall, London

Kullback S (1959) Information Theory and Statistics. Mineola, New York: Dover Publications (1968) van der Linde A (2002a) Dimension reduction and linear discriminant functions based on an odds ratio parametrization. Mathematik-Arbeitspapiere 56, Universität Bremen (http://www.math.uni-bremen.de/~avdl)

van der Linde A (2002b) On the association between a random parameter and an observable. Mathematik-Arbeitspapiere 57, Universität Bremen (http://www.math.uni-bremen.de/~avdl/). Final version: Test 13 (2004), 85–111

van der Linde A, Osius G (2003) Discrimination based on an odds ratio parametrization. Seventh Valencia International Meeting on Bayesian Statistics, June 2002. Bernando, JM et al (Eds.) Bayesian Statistics 7, 233–248, Oxford University Press

Nelson RB (1999) An Introduction to Copulas. Lecture Notes in Statistics 139, Springer, New York

Osius G (2000) The association between two random elements: A complete characterization in terms of odds ratios. Mathematik-Arbeitspapiere 53, Universität Bremen (http://www.math.uni-bremen.de/~osius)

Plackett RL (1965) A class of bivariate distributions. Journal of the American Statistical Association 60:516–522

Plackett RL (1974) The Analysis of Categorical Data. Griffin, London

Rüschendorf L, Schweizer B, Taylor MD (eds.) (1996) Distributions with Fixed Marginals and Related Topics. IMS Lecture Notes-Monograph Series 28,. Institute of Mathematical Statistics, Hayward, California

Sinkhorn R (1967) Diagonal Equivalence to Matrices With Prescribed Row and Column Sums. American Mathematical Monthly 74: 402–405