MATHEMATIK-ARBEITSPAPIERE

The Association Between Two Random Elements: A Complete Characterization In Terms of Odds Ratios

Gerhard Osius

Preprint No. 53 November 2000

UNIVERSITÄT BREMEN

MATHEMATIK-ARBEITSPAPIERE

A: MATHEMATISCHE FORSCHUNGSPAPIERE

The Association Between Two Random Elements: A Complete Characterization In Terms of Odds Ratios

Gerhard Osius

Preprint No. 53 November 2000

FACHBEREICH MATHEMATIK UND INFORMATIK
UNIVERSITÄT BREMEN

Bibliothekstraße
D-28359 Bremen
Germany

The Association Between Two Random Elements: A Complete Characterization In Terms of Odds Ratios

Gerhard Osius [1]

Summary

For random elements X and Y (e.g. vectors or functions) a complete characterization of their association is given by specifying an association function (in terms of odds ratios) which uniquely determines their *joint* distribution within a broad class of distributions with *fixed marginals.* The main result establishes for any association function (satisfying integrability conditions) the existence of a corresponding joint distribution with pre-specified marginals. The proof is constructive in the sense that the joint density is obtained as a limit of an iterative procedure of marginal fittings. Two applications are given, the first deals with *models* for the association. And the second specialises to multivariate *normal marginals* and provides a characterization of multivariate *joint normal* distributions via their association function.

Keywords: association, Kullback-Leibler information, logistic regression, marginal fitting, multivariate normal distributions, odds ratio, proportional fitting.

1 Institut für Statistik, Fachbereich 3, Universität Bremen
Postfach 330440, 28334 Bremen, Germany
e-mail: osius@math.uni-bremen.de

1. Introduction and Outline

The question how a random output Y of a system (e.g. the health status of a human) is associated to a random input 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 $\mathcal{L}(Y \mid X)$ by means of an appropriate model - but leaving the marginal distribution $\mathcal{L}(X)$ of the input arbitrary. The corresponding sampling design is to collect a random sample $Y_i \sim \mathcal{L}(Y \mid X = x_i)$ for given values x_i with $i = 1, ..., I$.

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 $\mathcal{L}(X \mid Y)$ and a sample $X_i \sim \mathcal{L}(X \mid Y = y_i)$ with fixed values y_i $(i = 1, ..., I)$. 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 the question arises, which amount of information concerning the joint distribution of (X,Y) is contained in both conditional distributions $\mathcal{L}(Y \mid X)$ and $\mathcal{L}(X \mid Y)$. Since conditioning on X resp. Y removes the information on the marginal distribution $\mathcal{L}(X)$ resp. $\mathcal{L}(Y)$ from the joint distribution $\mathcal{L}(X,Y)$, the question arises what *exactly* is left if we simultaneously remove the information on both marginal distributions from the joint distribution. More precisely we are looking for an unknown object $?(X,Y)$ - which will be called the *association* between X and Y - such that the joint distribution $\mathcal{L}(X,Y)$ is uniquely characterized by the tripel $\left(\mathcal{L}(X), \mathcal{L}(Y), ?(X,Y)\right)$ with independently varying components. Let us emphasize that we neither want a single *measure* nor specific *models* of associations but the *complete* association structure itself. The purpose of this paper is to define this structure, $?(X,Y)$, which turns out to be a *function* - and establish its basic properties. Before outlining the present work we look at some special classes of distributions, where the association object $?(X,Y)$ is already well known.

1.1 Normal Distributions

If X and Y are random vectors and their *joint* distribution is multivariate normal, then the correlation matrix $\varrho(X,Y)$ obviously is the desired association object ?(X,Y). However, this no longer holds if only the *marginal* distributions are normal, but the joint distribution is non-normal. In fact an example of two joint distributions (one normal but the other non-normal) will be given in **4.4** which have the same normal marginal distributions for X and Y as well as the same correlation matrix. Hence the correlation matrix is not the wanted association object ?(X,Y) unless we restrict on *joint normal* distribution.

1.2 Simple Random Variables

Looking at *simple* random variables (i.e. with finite range) we first consider the most elementary situation, i.e. $X \in \{0,1\}$ and $Y \in \{0,1\}$ are indicators. The joint distribution of X and Y is given by the probabilities

(1.1) $$\pi_{jk} = \mathrm{P}\{X = j, Y = k\} \qquad \text{for all } j, k.$$

It is well known (and easy to prove) that the odds ratio

$$\theta = (\pi_{11}\pi_{00})/(\pi_{10}\pi_{01})$$

may be taken as the desired association object ?(X,Y) - see e.g. Bishop et al (1975).

More generally, let $X \in \{0,...,J\}$ and $Y \in \{0,...,K\}$ be simple random variables with joint distribution given by (1.1). Then the $J \times K$ odds ratio matrix θ with entries

$$\theta_{jk} = (\pi_{jk}\pi_{00})/(\pi_{j0}\pi_{0k}) \qquad \text{for } j, k \geq 1$$

will serve as the wanted association ?(X,Y), see e.g. Plackett (1974, Sec. 3.4). The result, that for any matrix θ with positive entries and given marginal distributions there exists a unique joint distribution π may be obtained in different ways. First, the joint distribution π may be derived by a method - due to Sinkhorn (1967) - as a limit of iteratively rescaled distributions π_n with the given odds ratio matrix θ. Second, the joint distribution π may be obtained from results due to Haberman (1974, Theorem 2.6) as the unique argument maximizing a strictly concave function (arising from the log-likelihood of Poisson distributions).

1.3 Odds Ratios For Binary Output

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°

$$OR^\circ(x) \;=\; \frac{P\{Y=1 \mid X=x\}}{P\{Y=0 \mid X=x\}} \Bigg/ \frac{P\{Y=1 \mid X=x^\circ\}}{P\{Y=0 \mid X=x^\circ\}}$$

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

$$\text{(1.2)} \qquad \begin{aligned} P\{Y=1 \mid X=x\} \;&=\; \left(1 + a^{-1}\exp\{-\psi^\circ(x)\}\right)^{-1} && \text{resp.} \\ \operatorname{logit} P\{Y=1 \mid X=x\} \;&=\; \alpha + \psi^\circ(x) && \text{with} \\ \alpha \;=\; \log a \;&=\; \operatorname{logit} P\{Y=1 \mid X=x^\circ\}, \end{aligned}$$

where $\operatorname{logit} p = \log\left(p/[1-p]\right)$ denotes the logistic transformation. The marginal distribution of Y is given by

$$\text{(1.3)} \qquad P\{Y=1\} \;=\; E\left[\left(1 + a^{-1}\exp\{-\psi^\circ(X)\}\right)^{-1}\right].$$

This expectation is a strictly increasing function of a and approaches the limits 1 resp. 0 as $a \to \infty$ resp. $a \to 0$. Hence for $0 < P\{Y=1\} < 1$ there exists a unique $0 < a < \infty$ such that (1.3) holds. This shows that for fixed marginal distributions the joint distribution of (X,Y) 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 (1.2) with a obtained from (1.3).

1.4 General Case: An Outline

We now outline the basic concepts and results for the general case, where X and Y can be any type of random element (typically random *vectors* or *functions*). The purpose of this paper is to show that a straightforward generalization of the above odds ratio function provides the wanted association object ?(X,Y). 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 modified log odds ratio function ψ (which does not refer to reference values x° and y°) is defined as a projection of the log-density into an appropriate association space.

In section **3** we show as our main result that the odds ratio function characterizes the association and hence may be taken as the desired association object ?(X,Y). Using the Kullback-Leibler information we first prove (under mild integrability conditions which hold in particular for integrable log-densities) that the joint distribution is *uniquely determined* by the marginal distributions and the odds ratio function. To establish the *existence* of a joint distribution having fixed marginals and a given odds ratio function requires a greater effort and stronger additional assumptions. The wanted joint distribution will be obtained from a convergence theorem as a limit of a marginal fitting sequence of densities, which generalizes the sequence used by 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). Although the concepts here are straightforward generalizations derived from Sinkhorn's and Haberman's work, their approaches exploit unique features of finite distributions which are not available for non-simple random variables.

Two immediate applications of the characterization of association are given in section **4**. The first briefly deals with *association models* which only specify the association structure (i.e. the odds ratio function) but leave the marginal distributions completely arbitrary. The second application considers X and Y with multivariate *normal marginal* distributions and characterizes the *joint normal* distributions as those with a log-bilinear association. In particular a variety of non-normal joint distributions may be specified through an appropriate non-bilinear association function.

An important feature of the approach adopted here is a symmetry in presentation between X and Y. By interchanging X with Y any concept or argument entails a *dual* which only occasionally will be stated explicitly. - Although we attempt to a fairly self contained exposition the reader is referred to standard textbooks on probability and measure theory - e.g. Billingsley (1986) and Bauer (1991, 1992) - for definitions and results not given here.

2. The Odds Ratio Function

To formalize the introductary discussion we now consider arbitrary non-empty probability spaces $(\Omega_X, \mathscr{B}_X, \pi_X)$ and $(\Omega_Y, \mathscr{B}_Y, \pi_Y)$ and denote their product by $(\Omega, \mathscr{B}, \pi)$, i.e. the product set $\Omega = \Omega_X \times \Omega_Y$ equipped with the σ-algebra $\mathscr{B}$ generated by $\{B_X \times B_Y \mid B_X \in \mathscr{B}_X, B_Y \in \mathscr{B}_Y\}$ and the product probability 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 π: $P \ll \pi \ll P$. Further let $\mathscr{F}$ be the class of corresponding densities, i.e. the Radon-Nikodym derivatives $f = dP/d\pi$ for any $P \in \mathscr{P}$, and let $\Phi = \{\log f \mid f \in \mathscr{F}\}$ be the corresponding class of log-densities.

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

$$P\{X=j\}, P\{Y=k\} > 0 \quad \Rightarrow \quad P\{X=j, Y=k\} > 0 \qquad \text{for all } j, k.$$

For a multivariate *normal* distribution $P = N_k(\mu, \Sigma)$, 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 (see **4.4** for details).

2.1 Odds and Log Odds Ratio Function

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

$$OR_f(x, y \mid x', y') \;=\; \frac{f(x,y)\cdot f(x',y')}{f(x,y')\cdot f(x',y)} \;=\; \frac{f(x,y)}{f(x,y')} \Bigg/ \frac{f(x',y)}{f(x',y')} \,.$$

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

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

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 $OR_f^\circ = OR_f(\text{-}, \text{-} \mid x^\circ, y^\circ) : \Omega \longrightarrow (0, \infty)$ resp. $\psi_\varphi^\circ = \log OR_f^\circ$ since

$$OR_f(x,y\,|\,x',y') \;=\; \frac{OR_f^\circ(x,y\,)\cdot OR_f^\circ(x',y')}{OR_f^\circ(x,y')\cdot OR_f^\circ(x',y\,)} \qquad \text{resp.}$$

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

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. $\varphi = \log dP/d\pi \in \Psi$. Strictly speaking, the odds ratio function of P is not uniquely determined but only unique modulo the equivalence relation $\underset{\pi}{\sim}$ of π-almost-sure (π-a.s.) equality, i.e. $g \underset{\pi}{\sim} h$ if and only if $g = h$ π-a.s. Generally, we will say that a property holds *mudulo* an equivalence relation, if the corresponding property holds for the equivalence classes.

It is notationally 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. $X : \Omega_0 \longrightarrow \Omega_X$, $Y : \Omega_0 \longrightarrow \Omega_Y$ are measurable and P is the image measure of P_0 under the mapping $(X,Y)\colon \Omega_0 \longrightarrow \Omega$, which is uniquely determined by

$$P(B_X \times B_Y) = P_0\{X \in B_X,\, Y \in B_Y\}.$$

This is, of course, always possible, e.g. take $\Omega_0 = \Omega$ and X, Y as the projections. 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 P, i.e. $OR(X,Y) = OR(P)$.

2.2 Marginal and Conditional Distributions

If $f = dP/d\pi \in \mathscr{F}$ denotes the *joint* density of (X,Y) then the *marginal* densities f^X, f^Y of X, Y are given by

$$f^X(x) = \int f(x,y)\, d\pi_Y(y) > 0\,, \qquad\qquad f^Y(y) = \int f(x,y)\, d\pi_X(x) > 0\,.$$

Since f^X resp. f^Y is finite π_X-a.s. resp. π_Y-a.s. we may choose f such that f^X and f^Y are both finite. Then 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. It is essential, though trivial, that this also holds for the *conditional* density $f^{|Y} \in \mathscr{F}$ of X given Y, so that we have

$$OR(f^{|X}) \;=\; OR(f) \;=\; OR(f^{|Y})\,.$$

2.3 Change of Dominating Measures

Another 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 ν_X resp. ν_Y with positive density $\delta_X = d\pi_X/d\nu_X > 0$ resp. $\delta_Y = d\pi_Y/d\nu_Y > 0$. Then $P \in \mathscr{P}$ is dominated by the product measure $\nu = \nu_X \times \nu_Y$ and $f_\delta(x,y) = f(x,y)\,\delta_X(x)\,\delta_Y(y)$ defines a ν-density of P. Hence the odds ratio function may also be expressed with f replaced by f_δ

$$OR_f(x,y\,|\,x',y') \;=\; \frac{f_\delta(x,y\,)\cdot f_\delta(x',y')}{f_\delta(x,y')\cdot f_\delta(x',y\,)}\;.$$

For common dominating measures - e.g. Lebesgue's resp. the counting measure for continuous resp. simple random variables - this representation is commonly 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. multivariate normal distributions 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 mappings (with respect to some σ-algebras $\mathscr{B}_U$ resp. $\mathscr{B}_V$) with 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

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

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 arbitrary b, d. If the covariance matrices Σ_X and Σ_Y of X and Y are non-singular, then the choice of $A = \Sigma_X^{-1/2}$, $b = \mathrm{E}(X)$, $C = \Sigma_Y^{-1/2}$, $d = \mathrm{E}(Y)$ yield *standardized* random vectors U and V (i.e. with zero expectation and identity matrix as covariance matrix). Hence the odds ratio function of (X,Y) is easily obtained from those of the *standardized pair* (U, V).

2.5 Vector Spaces of Random Variables

Let $\mathscr{L}^0$ denote the vector space of all random variables from $(\Omega,\mathscr{B})$ into $\mathbb{R}$. For any $\varphi \in \Phi \subset \mathscr{L}^0$ the log odds ratio function $\psi^\circ_\varphi \in \mathscr{L}^0$ for the reference pair (x°, y°) will now be characterized as a projection of φ onto the linear subspace

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

First, let $\mathscr{L}^0_X$ denote the space of random variables defined on Ω_X. Since any $\xi \in \mathscr{L}^0_X$ naturally extends to a map $\xi^* \in \mathscr{L}^0$ defined by $\xi^*(x,y) = \xi(x)$, the space $\mathscr{L}^0_X$ is isomorphic to the corresponding subspace $\mathscr{L}^{0*}_X \subset \mathscr{L}^0$. For simplicity we will identify ξ with ξ^* and $\mathscr{L}^0_X$ with $\mathscr{L}^{0*}_X$ unless clarity demands their distinction (as in the next subsection). By duality, the space $\mathscr{L}^0_Y$ is isomorphic to the subspace $\mathscr{L}^{0*}_Y \subset \mathscr{L}^0$.

Next, the *marginal subspace* $\mathscr{M}^0$ is defined as the smallest subspace of $\mathscr{L}^0$ containing $\mathscr{L}^{0*}_X$ and $\mathscr{L}^{0*}_Y$, i.e.

$$\mathscr{M}^0 = \mathscr{L}^{0*}_X + \mathscr{L}^{0*}_Y = \{ \eta_X + \eta_Y \mid \eta_X \in \mathscr{L}^0_X,\, \eta_Y \in \mathscr{L}^0_Y \}.$$

The space $\mathscr{L}^0$ is the direct sum $\mathscr{M}^0 \oplus \mathscr{A}^0$, i.e. any $\xi \in \mathscr{L}^0$ may uniquely be written as $\xi = \eta + \psi$ with $\eta \in \mathscr{M}^0$ and $\psi \in \mathscr{A}^0$. The unique η resp. ψ - the *projection* of ξ onto $\mathscr{M}^0$ resp. $\mathscr{A}^0$ - is given by

$$\eta(x,y) = \xi(x, y^\circ) + \xi(x^\circ, y) - \xi(x^\circ, y^\circ)\,,$$

$$\psi(x,y) = \xi(x, y) + \xi(x^\circ, y^\circ) - \xi(x, y^\circ) - \xi(x^\circ, y)\,.$$

Hence the log odds ratio function may be written as a projection

$$\psi^\circ_\varphi = \Pi_{\mathscr{A}^0}(\varphi) = \Pi(\varphi \mid \mathscr{A}^0),$$

where $\Pi_{\mathscr{H}}$ or $\Pi(\text{-} \mid \mathscr{H})$ denotes the projection operator for a subspace $\mathscr{H}$.

2.6 The Space of Integrable Log-Densities

For a further representation of the log odds ratio function we consider the class $\mathscr{P}^1$ of probability measures P, for which the log-density $\varphi = \log dP/d\pi$ is *π-integrable.* Let $\mathscr{L}^1 \subset \mathscr{L}^0$ be the subspace of π-integrable random variables, $\mathscr{F}^1 = \{ f \in \mathscr{F} \mid \log f \in \mathscr{L}^1 \}$ the class of densities for $\mathscr{P}^1$, and $\Phi^1 = \Phi \cap \mathscr{L}^1$ the subspace of integrable log-densities. We now show that for $\varphi \in \Phi^1$ its log odds ratio function $\psi_\varphi \in \mathscr{L}^1$ is uniquely determined by the projection of φ onto a suitable linear subspace $\mathscr{A}^1 \subset \mathscr{L}^1$. Strictly speaking, this results holds only π-a.s., i.e. in the quotient space $\tilde{\mathscr{L}}^1 = \mathscr{L}^1/\underset{\pi}{\sim}$ of equivalence classes with respect to $\underset{\pi}{\sim}$ and the corresponding subspace $\tilde{\mathscr{A}}^1 \subset \tilde{\mathscr{L}}^1$. Using the *marginal functions* ξ^X and ξ^Y for $\xi \in \mathscr{L}^1$ given by

$$\xi^X(x) = \int \xi(x,y)\, d\pi_Y(y)\ , \qquad \xi^Y(y) = \int \xi(x,y)\, d\pi_X(x)\ ,$$

we define the subspace

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

Considering the corresponding subspaces $\mathscr{L}_X^1 \subset \mathscr{L}_X^0$ and $\mathscr{L}_Y^1 \subset \mathscr{L}_Y^0$ of integrable functions - which are isomorphic to their respective images $\mathscr{L}_X^{1*}$ and $\mathscr{L}_Y^{1*}$ in $\mathscr{L}^1$ - we now define the *marginal subspace* of integrable functions as

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

Since this is not a *direct sum,* we look for another decomposition of $\mathscr{M}^1$. Observe first that any $\alpha \in \mathbb{R}$ corresponds to a constant map $\alpha^* \in \mathscr{L}^1$ via $\alpha^*(x,y) = \alpha$ which makes $\mathbb{R}^*$ a subspace of $\mathscr{L}^1$. Defining

$$\mathscr{M}_X^1 \ = \ \{\, \beta \in \mathscr{L}_X^1 \mid \int \beta\, d\pi_X = 0\,\}\ , \qquad \mathscr{M}_Y^1 \ = \ \{\, \gamma \in \mathscr{M}_Y^1 \mid \int \gamma\, d\pi_Y = 0\,\}$$

we get the desired decomposition as a direct sum

$$\mathscr{M}^1 = \ \mathbb{R}^* \oplus \mathscr{M}_X^{1*} \oplus \mathscr{M}_Y^{1*}$$

i.e. any $\eta \in \mathscr{M}^1$ has a unique representation $\eta = \alpha^* + \beta^* + \gamma^*$ with $\alpha \in \mathbb{R}$, $\beta \in \mathscr{M}_X^1$ and $\gamma \in \mathscr{M}_Y^1$, namely

$$\alpha = \int \eta\, d\pi\ , \qquad \beta = \eta^X - \int \eta\, d\pi\ , \qquad \gamma = \eta^Y - \int \eta\, d\pi\ .$$

The unique α^* resp. β^*, γ^* is the *projection* of η onto the respective subspace.

Now any $\xi \in \mathscr{L}^1$ with *finite* ξ^X and ξ^Y may uniquely be written as $\xi = \eta + \psi$ with $\eta \in \mathscr{M}^1$ and $\psi \in \mathscr{A}^1$ given by

(2.1)
$$\begin{aligned} \eta(x,y) \ &= \ \xi^X(x) + \xi^Y(y) - \int \xi\, d\pi \\ \psi(x,y) \ &= \ \xi(x,y) - \xi^X(x) - \xi^Y(y) + \int \xi\, d\pi\ . \end{aligned}$$

The uniqueness of the representation is obtained as follows. From $\xi = \eta + \psi$ and $\psi^X = 0$ we get $\xi^X = \eta^X + \psi^X = \eta^X$, and similiar $\xi^Y = \eta^Y$ and $\int \xi\, d\pi = \int \eta\, d\pi$. Hence, by the above decomposition $\eta = \alpha^* + \beta^* + \gamma^* = \eta^X + \eta^Y - \int \eta\, d\pi$ we get the first equation of (2.1) which in turn implies the second in view of $\xi = \eta + \psi$.

The unique η resp. ψ in the decomposition $\xi = \eta + \psi$ is the *projection* of ξ onto the respective subspace. The *finiteness* of ξ^X and ξ^Y is not crucial, since both are finite π-a.s. Hence we get a corresponding decomposition $\tilde{\mathscr{L}}^1 = \tilde{\mathscr{M}}^1 \oplus \tilde{\mathscr{A}}^1$ for the quotient spaces.

Finally, for any $P \in \mathscr{P}^1$ we can choose a log-density $\varphi \in \Phi^1$ with finite φ^X and φ^Y. From the easily verified identities

$$\Pi(\varphi \mid \mathscr{A}^1) = \Pi\big(\Pi(\varphi \mid \mathscr{A}^0) \mid \mathscr{A}^1\big)$$
$$\Pi(\varphi \mid \mathscr{A}^0) = \Pi\big(\Pi(\varphi \mid \mathscr{A}^1) \mid \mathscr{A}^0\big)$$

we conclude that the log odds ratio function $\psi^\circ_\varphi = \Pi(\varphi \mid \mathscr{A}^0)$ determines and is determined by the projection $\Pi(\varphi \mid \mathscr{A}^1)$ of the log-density φ onto the space $\mathscr{A}^1$. Note that this projection - in contrast to ψ°_φ - does not refer to a reference pair (x°, y°).

We now show that for any $f \in \mathscr{F}^1$ the *marginal* log-densities $\log f^X$ and $\log f^Y$ are integrable, too, which in turn implies the integrability of the *conditional* log-densities $\log f^{|X}$ and $\log f^{|Y}$. Indeed - and slightly more general - for any $\varphi \in \mathscr{L}^1$ and $\xi = \exp(\varphi)$ Jensen's inequality gives

$$\log \xi^X(x) = \log\big[\int \xi(x,y)\, d\pi(y)\big] \geq \int \log \xi(x,y)\, d\pi(y) = \varphi^X(x)$$

and from $\log(a) \leq a - 1$ we get the fundamental inequality

$$\text{(2.2)} \qquad \varphi^X \leq \log \xi^X \leq \xi^X - 1\,.$$

Hence $\log \xi^X$ is π_X-integrable, provided ξ is π-integrable (e.g. for $\xi \in \mathscr{F}^1$), which by duality also implies that $\log \xi^Y$ is π_Y-integrable.

It is tempting to pass to the subspace $\mathscr{L}^2 \subset \mathscr{L}^1$ of *square-integrable* random elements where the spaces $\mathscr{M}^2 = \mathscr{M}^1 \cap \mathscr{L}^2$ and $\mathscr{A}^2 = \mathscr{A}^1 \cap \mathscr{L}^2$ are in fact *orthogonal* with respect to the inner product $\langle \xi_1, \xi_2 \rangle = \int \xi_1 \xi_2 \, d\pi$. Furthermore $\mathbb{R}^* \subset \mathscr{L}^2$ and the intersections of $\mathscr{M}^{1*}_X$ and $\mathscr{M}^{1*}_Y$ with $\mathscr{L}^2$ are pairwise orthogonal too. However, we do not want to impose unneeded asssumptions and therefore restrict our attention to integrable log-densities and hence to the space $\mathscr{L}^1$.

As already indicated, from now on we notationally identify the various spaces like $\mathscr{M}^1_X$ with their isomorphic counterparts $\mathscr{M}^{1*}_X$.

3. A Characterization of Association

This section contains our main result, the characterization of association in terms of the odds ratio function. First we will show 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 $OR(P)$. In order to evaluate Kullback-Leibler information in the proof requires mild integrability conditions, which in particular hold if the log-density of P is π-integrable (i.e. $P \in \mathscr{P}^1$) or if either Ω_X or Ω_Y is *finite*.

Conversely, for a given $\psi \in \mathscr{A}^1$ we provide a necessary and sufficient condition for the existence of a joint distribution $P \in \mathscr{P}^1$ with the given marginal distributions π_X and π_Y and π-integrable log-density $\varphi = \log(dP/d\pi)$ such that the odds ratio function of P corresponds to ψ, i.e. ψ is the projection of φ onto $\mathscr{A}^1$. In this case the density φ may be found as a limit of an iterative procedure.

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

Suppose that $P_1, P_2 \in \mathscr{P}$ have the same marginal distributions π_X, π_Y and a common odds ratio function with respect to an arbitrary (x°, y°):

$$OR^\circ(P_1) = OR^\circ(P_2).$$

We want to show that P_1 equals P_2. For the densities $f_i = dP_i/d\pi \in \mathscr{F}$ we have $f_i^X = dP_i^X/d\pi_X$ and $f_i^Y = dP_i^Y/d\pi_Y$ so that P_i has the marginal distributions π_X and π_Y if and only if

$$\text{(3.1)} \qquad \begin{aligned} f_i^X &= 1 \qquad P_i^X\text{-a.s.,} \\ f_i^Y &= 1 \qquad P_i^Y\text{-a.s.,} \qquad \text{for } i = 1, 2. \end{aligned}$$

The projections of the log-densities $\varphi_i = \log f_i \in \Phi$ onto $\mathscr{A}^0$ are the log odds ratio functions and thus coincide for $i = 1,2$. Hence $\varphi_1 - \varphi_2$ equals its projection onto the marginal space $\mathscr{M}^0$ and may be written as

$$(\varphi_1-\varphi_2)(x,y) \;=\; \beta_1(x) + \gamma_1(y) \qquad \text{with}$$
$$\beta_1(x) \;=\; (\varphi_1-\varphi_2)(x,y^\circ)$$
$$\gamma_1(y) \;=\; (\varphi_1-\varphi_2)(x^\circ,y) - (\varphi_1-\varphi_2)(x^\circ,y^\circ).$$

To establish $P_1 = P_2$ - under integrability assumptions given later - we will show that the *Kullback-Leibler information*

$$I(f_1,f_2) \;=\; \int f_1 \log(f_1/f_2)\, d\pi \;=\; \int (\varphi_1-\varphi_2)\, dP_1 \;=\; \int (\beta_1+\gamma_1)\, dP_1$$

is zero, which in turn implies $f_1 = f_2$ π-a.s. and hence $P_1 = P_2$. We first note that

(3.2) $\qquad 0 \;\le\; I(f_1,f_2) \;\le\; \infty$

always holds, even if $\varphi_1 - \varphi_2$ is only quasi-integrable (i.e. the positive or the negative part is integrable) with respect to P_1. In fact, from $-\log a \le a^{-1} - 1$ we conclude for the negative part of $\varphi_1 - \varphi_2$

$$(\varphi_1-\varphi_2)^- \;=\; \big(\log(f_1/f_2)\big)^- \;\le\; f_2/f_1\,.$$

Since $\int (f_2/f_1)\, dP_1 = \int f_2\, d\pi = 1$ we get

(3.3) $\qquad (\varphi_1-\varphi_2)^-$ is P_1-integrable.

Hence $(\varphi_1-\varphi_2)^-$ is P_1-quasi-integrable and $I(f_1,f_2) > -\infty$. Now the first part of (3.2) follows from Jensen's inequality,

$$I(f_1,f_2) \;=\; \int -\log(f_2/f_1)\, dP_1 \;\ge\; -\log \int (f_2/f_1)\, dP_1 \;=\; 0\ ,$$

where equality holds if and only if $f_1 = f_2$ P_1-a.s. - which in view of $P_1 \ll \pi \ll P_1$ is equivalent to π-a.s. - and hence if $P_1 = P_2$.

Now, in order to evaluate the Kullback-Leibler information via

(3.4) $\qquad I(f_1,f_2) \;=\; \int (\beta_1+\gamma_1)\, dP_1 \;=\; \int \beta_1\, dP_1 + \int \gamma_1\, dP_1\,,$

we have to justify the decomposition under suitable integrability conditions. For any *non-negative* $\beta \in \mathscr{L}_X^0$ we have by Fubini's theorem and (3.1)

$$\int \beta\, dP_i \;=\; \int \beta \cdot f_i\, dP \;=\; \int \beta \cdot f_i^X\, d\pi_X \;=\; \int \beta\, d\pi_X \qquad \text{for } i = 1, 2.$$

Separate application to the positive and negative part of β_1 leads to

(3.5) $\qquad \beta_1$ is P_i-integrable $\quad\Leftrightarrow\quad \beta_1$ is π_X-integrable ,

$\qquad\qquad\qquad\qquad\qquad\qquad\quad \Rightarrow \quad \int \beta_1\, dP_i \;=\; \int \beta_1\, d\pi_X\,.$

The dual argument provides the corresponding result for γ_1, which even holds if *integrability* is replaced by *quasi-integrability*, (3.6) $\quad \gamma_1$ is P_i-quasi-integrable $\Leftrightarrow$ γ_1 is π_Y-quasi-integrable ,

$$\Rightarrow \quad \int \gamma_1 \, dP_i = \int \gamma_1 \, d\pi_Y .$$

Now we can state and prove the uniqueness results.

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°). Further 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) $\quad (\varphi_1 - \varphi_2)(\text{-}, y^\circ)$ *is π_X-integrable,*

(ii) $\quad (\varphi_1 - \varphi_2)(x^\circ, \text{-})$ *is π_Y-integrable.*

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.

Proof: By duality it suffices to prove the theorem under the assumption (i) that $\beta_1 = (\varphi_1 - \varphi_2)(\text{-}, y^\circ)$ is π_X-integrable and thus P_1-integrable by (3.5). Hence the negative part of $\gamma_1 = (\varphi_1 - \varphi_2) - \beta_1$ is P_1-integrable by (3.3) and π_Y-integrable by (3.6). This establishes (3.4) which, in view of (3.2), implies

(3.7) $$-\int \beta_1 \, d\pi_X \leq \int \gamma_1 \, d\pi_Y .$$

Switching the indices 1 and 2 we get

$$\varphi_2 - \varphi_1 = \beta_2 + \gamma_2 \qquad \text{with} \qquad \beta_2 = -\beta_1, \qquad \gamma_2 = -\gamma_1.$$

Since $\beta_2 = -\beta_1$ is π_X-integrable by (i), we obtain a result corresponding to (3.7)

$$-\int \beta_2 \, d\pi_X \leq \int \gamma_2 \, d\pi_Y .$$

Combined with (3.7) we get

$$-\int \beta_1 \, d\pi_X = \int \gamma_1 \, d\pi_Y$$

and hence $I(f_1, f_2) = 0$, which completes the proof of the theorem.

The corollary follows since integrability of φ_i implies that $\int |\varphi_i(\text{-}, y)| \, d\pi_X$ is finite for all y except on a set of π_Y-measure zero. Hence there exist a y° with $\varphi_i(\text{-}, y^\circ) \in \mathscr{L}^1_X$ and (i) holds. The existence of x° with (ii) follows by duality. □

3.2 Existence

For a given function $\psi^\circ \in \mathscr{A}^0$ we now wish to construct a joint distribution $P \in \mathscr{P}^1$ with given marginal distributions π_X and π_Y such that the odds ratio function of P corresponds to ψ°, i.e. ψ° is the projection of the log-density $\varphi = \log(dP/d\pi)$ onto $\mathscr{A}^0$. As one might expect the construction of P requires certain restrictions on ψ°, the first being that ψ° is *π-integrable*, this will in turn produce a π-integrable log-density φ. Since $\psi^\circ \in \mathscr{L}^1$ is uniquely determined by its projection $\psi = \Pi(\psi^\circ \,|\, \mathscr{A}^1)$ via $\psi^\circ = \Pi(\psi \,|\, \mathscr{A}^\circ)$ we may alternatively start with a given π-integrable „association function“ $\psi \in \mathscr{A}^1$ (which does not refer to the reference values x°, y°) and look for a log-density φ with $\psi = \Pi(\varphi \,|\, \mathscr{A}^1)$, i.e. φ is of the form $\varphi = \eta + \psi$ with $\eta \in \mathscr{M}^1$. We prove that such an $\eta \in \mathscr{M}^1$ exists if the following *existence conditions* hold:

(EC1) *There exists* $\bar\beta \in \mathscr{L}^1_X$ *such that* $\exp(\bar\beta + \psi)$ *is π-integrable.*

(EC2) *There exists* $\bar\gamma \in \mathscr{L}^1_Y$ *such that* $\exp(\bar\gamma + \psi)$ *is π-integrable.*

Passing from $\bar\beta$ to $\bar\beta - \int \bar\beta \, d\pi_X \in \mathscr{M}^1_X$ we may equivalently assume $\bar\beta \in \mathscr{M}^1_X$ in (EC1) and, by duality, $\bar\gamma \in \mathscr{M}^1_Y$ in (EC2). Equivalent formulations in terms of the marginal functions of $q = \exp\psi$ are the following:

(EC1)$'$ $\log q^X = \log\big((\exp\psi)^X\big)$ *is π_X-integrable.*

(EC2)$'$ $\log q^Y = \log\big((\exp\psi)^Y\big)$ *is π_Y-integrable.*

Suppose first that (EC1) holds and set $\xi = \exp\{\bar\beta + \psi\}$. Then

$$\xi^X(x) \;=\; \int \exp\{\bar\beta(x) + \psi(x,y)\}\, d\pi_Y(y) \;=\; \exp\{\bar\beta(x)\} \cdot q^X(x)$$

and ξ^X is finite π_X-a.s.. Hence $\log q^X = \log \xi^X - \bar\beta$ is π_X-integrable by (2.2) which yields (EC1)$'$. Conversely, if (EC1)$'$ holds, then q^X is finite π_X-a.s and there exists a version $\bar\beta \in \mathscr{L}^1_X$ such that $\bar\beta = -\log q^X$ π_X-a.s.. Hence

$$\int \exp\{\bar\beta(x) + \psi(x,y)\}\, d\pi(x,y) \;=\; \int \exp\{-\log q^X(x)\} \cdot q^X(x)\, d\pi_X(x) \;=\; 1$$

which proves (EC1). By duality we get the equivalence of (EC2) and (EC2)$'$.

A stronger version of both (EC1) and (EC2) is obtained by replacing $\bar\beta$ or $\bar\gamma$ with the constant zero function , i.e.

(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 existence conditions (EC1) and (EC2) are *sufficient* for the existence of the wanted joint distribution:

Existence-Theorem: *For any $\psi \in \mathscr{A}^1$ the existence conditions (EC1) and (EC2) imply the existence of a joint distribution $P \in \mathscr{P}^1$ with given marginal distributions π_X and π_Y and a π-integrable log-density $\varphi = \log(dP/d\pi)$ such that $\psi = \Pi(\varphi \mid \mathscr{A}^1)$.*

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 - at least not for *binary* Y (cf. **1.3**).

The desired log-density φ will be constructed as a limit of an iterative procedure of so-called *marginal fittings*. Since q^X is finite π_X-a.s. by (EC1)$'$, we will take a version of ψ modulo $\underset{\pi}{\sim}$ such that q^X is finite, and by duality from (EC2)$'$ we assume q^Y to be finite, too. Furthermore let $\bar{\beta}$ and $\bar{\gamma}$ denote the functions given by (EC1) and (EC2), which will be chosen such that $\bar{\beta} \in \mathscr{M}_X^1$ and $\bar{\gamma} \in \mathscr{M}_Y^1$ hold, and put

$$\bar{\eta} \;=\; \bar{\beta} + \bar{\gamma} \;\in \mathscr{M}^1, \qquad \bar{\varphi} \;=\; \bar{\eta} + \psi \;=\; \bar{\beta} + \bar{\gamma} + \psi \;\in \mathscr{L}^1 .$$

The corresponding marginal functions (projections) are

$$\bar{\varphi}^X = \bar{\eta}^X = \bar{\beta}\;, \qquad \bar{\varphi}^Y = \bar{\eta}^Y = \bar{\gamma}.$$

Since $\bar{\varphi}$ and the desired φ both have the same projection ψ in $\mathscr{A}^1$, φ is necessarily of the form

$$\varphi \;=\; \eta + \bar{\varphi} \qquad \text{with} \qquad \eta \in \mathscr{M}^1.$$

We are looking for an element $\eta \in \mathscr{M}^1$ such that

$$\bar{f}(\eta) \;=\; \exp\{\,\eta + \bar{\varphi}\} \;=\; \exp\{\eta + \bar{\beta} + \bar{\gamma} + \psi\}.$$

is the π-density of a probability measure P with marginal distributions π_X and π_Y, i.e.

$$\bar{f}^X(\eta) = 1 \quad \pi_X\text{-a.s.}\,, \qquad \bar{f}^Y(\eta) = 1 \quad \pi_Y\text{-a.s.}$$

This element η will be constructed as a limit of a sequence $\eta_n \in \mathscr{M}^1$. The definition of this sequence and the proof of its convergence is divided into several parts.

3.3 Marginal Fitting

As a first step we define an operator M_X for joint distributions P which switches the marginal distribution of X under P to the given marginal π_X. Briefly speaking, the new distribution $M_X(P)$ arises by replacing the marginal distribution P^X via conditioning with π_X. This leaves the odds ratio function unchanged, but typically alters the marginal distribution of Y, too. More precisely, let $P \in \mathscr{P}^1$ be given by a π-density of the form $\bar{f}(\eta) \in \mathscr{F}^1$ with $\eta \in \mathscr{M}^1$. Then the marginal densitiy $\bar{f}^X(\eta)$ is finite π_X-a.s., and we can choose a version modulo $\underset{\pi}{\sim}$ such that

$$0 < \bar{f}^X(\eta) < \infty .$$

From (2.2) we get $\log \bar{f}^X(\eta) \in \mathscr{L}^1_X$ and define the operator by

$$M_X(\eta) = \eta - \log \bar{f}^X(\eta) \in \mathscr{M}^1.$$

The corresponding density $\bar{f}(M_X(\eta))$ is the conditional density of P given X,

$$\bar{f}(M_X(\eta))(x,y) = \bar{f}(\eta)(x,y) \,/\, \bar{f}^X(\eta)(x) .$$

We denote the corresponding joint distribution by $M_X(P)$. In particular, the odds ratio function for $M_X(P)$ and P coincide. The marginal distribution of X under $M_X(P)$ is π_X because the marginal density of $\bar{f}(M_X(\eta))^X$ of X is identically 1.

The definition of $M_X(\eta)$ above is valid for any $\eta \in \mathscr{M}^1$ such that $\bar{f}(\eta)$ is π-integrable (but not necessarily a *probability* density), and the resulting $\bar{f}(M_X(\eta))$ is still a *probability* density.

For any representation

$$\eta = \eta_X + \eta_Y \qquad \text{with} \qquad \eta_X \in \mathscr{L}^1_X,\ \eta_Y \in \mathscr{L}^1_Y$$

we obtain $\bar{f}^X(\eta) = \exp(\eta_X)\, \bar{f}^X(\eta_Y)$. Hence $M_X(\eta)$ depends only through η_Y on η,

$$M_X(\eta) = M_X(\eta_Y).$$

In particular, the decompositions using the projections

$$\eta = \eta^X + \gamma\,, \qquad \gamma = \eta^Y - \int \eta \, d\pi = \Pi(\eta \mid \mathscr{M}^1_Y)$$

$$\eta = \beta + \eta^Y\,, \qquad \beta = \eta^X - \int \eta \, d\pi = \Pi(\eta \mid \mathscr{M}^1_X)$$

yield

$$M_X(\eta^Y) = M_X(\eta) = M_X(\gamma) = \gamma - \log \bar{f}^X(\gamma).$$

Hence the projection of $M_X(\eta)$ onto $\mathscr{L}^1_X$ is

$$[M_X(\eta)]^X = -\log \bar{f}^X(\gamma)\,, \tag{3.8}$$

and (2.2) applied to $\xi = \exp\{\gamma + \bar{\varphi}\}$ yields the important inequality

$$(3.9) \qquad 1 - \bar{f}^X(\gamma) \;\leq\; \left[M_X(\eta)\right]^X \;\leq\; -\bar{\varphi}^X \;=\; -\,\bar{\beta}\,.$$

Furthermore, the operator M_X is evidently idempotent,

$$M_X\big(M_X(\eta)\big) \;=\; M_X(\eta).$$

The dual operator is

$$M_Y(\eta) \;=\; \eta - \log \bar{f}^Y(\eta) \;\in\; \mathscr{M}^1$$

and has the dual properties

$$(3.10) \qquad M_Y(\eta^X) \;=\; M_Y(\eta) \;=\; M_Y(\beta) \;=\; \beta - \log \bar{f}^Y(\beta)\,,$$

$$(3.11) \qquad \left[M_Y(\eta)\right]^Y \;=\; -\,\log \bar{f}^Y(\beta)\,,$$

$$(3.12) \qquad 1 - \bar{f}^Y(\beta) \;\leq\; \left[M_Y(\eta)\right]^Y \;\leq\; -\bar{\varphi}^Y \;=\; -\,\bar{\gamma}\,.$$

Now take an arbitrary $\eta_0 \in \mathscr{M}^1$ such that $\bar{f}(\eta_0) \in \mathscr{F}^1$ is the π-densitiy of a joint distribution $P_0 \in \mathscr{P}^1$. Iterating the operators M_X and M_Y, we obtain a sequence $\eta_n \in \mathscr{M}^1$, recursively given by

$$\eta_{n+1} = M_X\big(M_Y(\eta_n)\big)\,, \qquad\qquad n \geq 0,$$

and an accompanying sequence $\tilde{\eta}_n \in \mathscr{M}^1$, namely

$$\tilde{\eta}_n = M_Y(\eta_n) \qquad \text{with} \qquad \eta_{n+1} = M_X(\tilde{\eta}_n).$$

The sequence $(\eta_n)_{n \geq 0}$ will be called the *marginal fitting sequence* with starting value η_0 and is a straightforward generalization of the sequence used by Sinkhorn (1967) for simple random variables X and Y. Note that by duality we could switch the operators M_X and M_Y in the definition of η_n which would lead to a dual sequence $\tilde{\eta}_n$ for a suitable starting value $\tilde{\eta}_0$. For definiteness, however, we will stick to the choice above.

3.4 Maximizing Functional and Convergence Theorem

As we will see later, the desired joint distribution P may also be obtained by maximizing the real-valued functional ℓ defined on $\mathscr{M}^1$ by

$$\ell(\eta) = \int[\eta - \bar{f}(\eta)]\, d\pi = \int \eta\, d\pi - \int \bar{f}(\eta)\, d\pi$$

$$= \int l\big(\eta(x,y) \,|\, \bar{\varphi}(x,y)\big)\, d\pi(x,y) \qquad \text{with}$$

$$l(u\,|\,v) = u - \exp(u+v) \qquad \text{for } u, v \in \mathbb{R}.$$

The functional ℓ generalizes the function considered by Haberman (1974, Theorem 2.6) for (conditional) Poisson distributed Y. Note that $\ell(\eta)$ is finite for π-integrable $\bar{f}(\eta)$ and $-\infty$ otherwise. If $\bar{f}(\eta)$ is a probability density, we get

$$\ell(\eta) = \int \eta\, d\pi - 1 \qquad \text{for} \quad \bar{f}(\eta) \in \mathscr{F}.$$

For fixed v the function $l(u\,|\,v)$ is strictly concave in u and attains its unique maximum for $u = -v$. Hence ℓ is bounded from above

$$\ell(\eta) \le \ell(-\bar{\varphi}) = 1 - \int \bar{\varphi}\, d\pi \tag{3.13}$$

and strictly concave modulo $\underset{\pi}{\sim}$ on the convex set

$$\mathscr{M}^{1\pi} := \{\, \eta \in \mathscr{M}^1 \mid \bar{f}(\eta) \text{ is } \pi\text{-integrable} \,\} = \{\, \eta \in \mathscr{M}^1 \mid \ell(\eta) > -\infty \,\}\,.$$

More precisely, ℓ is concave and the function induced by ℓ on the corresponding quotient space $\tilde{\mathscr{M}}^{1\pi}$ with respect to the equivalence relation $\underset{\pi}{\sim}$ is strictly concave.

For any $\eta \in \mathscr{M}^{1\pi}$ we now establish an important connection between the functional ℓ and the marginal operators M_X and M_Y:

$$\ell(\eta) \le \ell\big(M_X(\eta)\big)\,, \qquad \ell(\eta) \le \ell\big(M_Y(\eta)\big) \tag{3.14}$$

$$\ell(\eta) = \ell\big(M_X(\eta)\big) \quad \Leftrightarrow \quad \bar{f}^X(\eta) = 1 \;\; \pi_X\text{-a.s.} \quad \text{resp.} \quad \eta = M_X(\eta) \;\; \pi\text{-a.s.} \tag{3.15}$$

$$\ell(\eta) = \ell\big(M_Y(\eta)\big) \quad \Leftrightarrow \quad \bar{f}^Y(\eta) = 1 \;\; \pi_Y\text{-a.s.} \quad \text{resp.} \quad \eta = M_Y(\eta) \;\; \pi\text{-a.s.}$$

By duality it suffices to prove the results for M_X. From the definitions of ℓ and M_X we get

$$\ell\big(M_X(\eta)\big) = \int \big(\eta - \log \bar{f}^X(\eta) - 1\big)\, d\pi \qquad \text{and}$$

$$\ell\big(M_X(\eta)\big) - \ell(\eta) = \int \big[\bar{f}(\eta) - \log \bar{f}^X(\eta) - 1\big]\, d\pi$$

$$= \int \big[\bar{f}^X(\eta) - \log \bar{f}^X(\eta) - 1\big]\, d\pi_X\,.$$

Hence the results follow from the inequality $\log(a) \le a - 1$, where equality holds if and only if $a = 1$.

We are now in a position to formulate a convergence result for the iterated marginal fitting procedure.

Convergence-Theorem: *Let* $(\eta_n \in \mathscr{M}^1)_{n\geq 0}$ *be any marginal fitting sequence with starting value* $\eta_0 \in \mathscr{M}^1$ *such that* $\int \bar{f}(\eta_0)\, d\pi = 1$. *If the existence conditions (EC1) and (EC2) hold, then any subsequence* $(\eta_{m(n)})$ *contains a further subsequence* $(\eta_{m'(n)})$ *which converges pointwise as well as in the mean to an element* $\eta \in \mathscr{M}^1$ *having the following properties*

(i) $\qquad M_X(\eta) = \eta \quad \pi\text{-a.s.} \qquad \text{and} \qquad M_Y(\eta) = \eta \quad \pi\text{-a.s.}$

(ii) $\qquad \int \bar{f}(\eta)\, d\pi = 1\,.$

The limit η *is* π*-a.s. independent of the chosen subsequence* $(\eta_{m(n)})$ *and independent of* η_0, *i.e. for two starting values* η_{10}, η_{20} *and any subsequences* $(\eta_{1m(n)})$, $(\eta_{2m(n)})$ *the corresponding limits* η_1 *and* η_2 *coincide* π*-a.s.*

Corollary: *For any starting value* $\eta_0 \in \mathscr{M}^1$ *the marginal fitting sequence* $(\eta_n)_{n\geq 0}$ *converges in the mean to an element* $\eta \in \mathscr{M}^1$ *and*

(iii) $\qquad \big(\ell(\eta_n)\big)_{n\geq 0}$ *is a non-decreasing sequence with limit* $\ell(\eta)$.

Furthermore η *is the* π*-a.s.-unique argument which maximizes the functional* ℓ.

Before providing a proof we reduce the existence theorem to the above convergence theorem. First, $\bar{f}(-\bar{\gamma}) = \exp(\bar{\beta} + \psi)$ is π-integrable by condition (EC1) and hence

$$\eta_0 \;=\; -\,\bar{\gamma} - \log\big[\int \bar{f}(-\bar{\gamma})\, d\pi\big] \;\in\; \mathscr{M}^1$$

provides a starting value with $\int \bar{f}(\eta_0)\, d\pi = 1$. By the convergence theorem we get an element $\eta \in \mathscr{M}^1$ satisfying (i) and (ii). We now show that

$$\varphi \;=\; \eta + \bar{\varphi} \;=\; \eta + \bar{\beta} + \bar{\gamma} + \psi \;\in \mathscr{L}^1$$

is the wanted log-density in the existence theorem. Since $M_X(\eta) = \eta$ resp. $M_Y(\eta) = \eta$ is equivalent to $\bar{f}^X(\eta) = 1$ resp. $\bar{f}^Y(\eta) = 1$, the function $\bar{f}(\eta) = \exp(\varphi)$ is a π-density of a probability measure P with marginal distributions π_X and π_Y. Furthermore the log odds ratio function of P is determined by $\psi = \Pi(\varphi \,|\, \mathscr{A}^1)$, which proves the existence theorem.

The above argument may be extended in order to show that the limit η in the convergence theorem is π-a.s. independent of the subsequence and the starting value. Indeed, by the uniqueness theorem (and its corollary), the corresponding log-densities $\varphi_i = \eta_i + \bar{\varphi}$ for $i = 1, 2$ coincide π-a.s. and hence $\eta_1 = \eta_2$ π-a.s. $\square$

3.5 Properties of Marginal Fitting Sequences

Before proving the existence of the limit η in the convergence theorem, we establish some important inequalities for marginal fitting sequences $(\eta_n)_{n\geq 0}$ with starting value η_0 and the accompanying sequence $(\tilde{\eta}_n)_{n\geq 0}$ given in **3.3**. In the following the index “n“ always ranges over non-negative integers, i.e $n \geq 0$. An inequality $\xi_1 \leq \xi_2$ resp. $\xi_1 < \xi_2$ for *functions* ξ_1 and ξ_2 here means that $\xi_1(z) \leq \xi_2(z)$ resp. $\xi_1(z) < \xi_2(z)$ holds for all arguments z. Note, that the existence conditions (EC1) and (EC2) will *not* be assumed in this subsection.

We first show that both sequences $\ell(\eta_n)$ and $\ell(\tilde{\eta}_n)$ are non-decreasing and have the same limit s. Indeed, from (3.13) and (3.14) we get by induction

(3.16) $$\ell(\eta_n) \leq \ell(\tilde{\eta}_n) \leq \ell(\eta_{n+1}) \leq \ell(\tilde{\eta}_{n+1}) \leq \ell(-\bar{\varphi})$$

which implies the convergences to $s \in \mathbb{R}$

(3.17) $$\ell(\eta_n) = \alpha_n - 1 \longrightarrow s\,, \qquad \alpha_n = \int \eta_n \, d\pi\,,$$

$$\ell(\tilde{\eta}_n) = \tilde{\alpha}_n - 1 \longrightarrow s\,, \qquad \tilde{\alpha}_n = \int \tilde{\eta}_n \, d\pi\,.$$

Hence (α_n) and $(\tilde{\alpha}_n)$ are non-decreasing too with limit $\alpha = s+1$

$$\alpha_n \uparrow \alpha\,, \qquad \tilde{\alpha}_n \uparrow \alpha\,.$$

Next we provide upper and lower bounds for

$$\eta_{n+1} = M_X(\tilde{\eta}_n) = M_X(\gamma_n) \qquad \text{with} \qquad \gamma_n = \tilde{\eta}_n^Y - \tilde{\alpha}_n = \Pi(\tilde{\eta}_n \,|\, \mathscr{M}_Y^1).$$

From (3.8) we get

$$\eta_{n+1}^X = -\log \bar{f}^X(\gamma_n)\,,$$

and hence the representation

$$\eta_{n+1} = \gamma_n + \eta_{n+1}^X\,.$$

This allows us to derive bounds for η_{n+1} from those for γ_n and η_{n+1}^X. By (3.9), applied to $\tilde{\eta}_n$, we have

(3.18) $$1 - \bar{f}^X(\gamma_n) \leq \eta_{n+1}^X \leq -\bar{\beta}\,,$$

which already provides the upper bound for η_{n+1}^X. For the accompanying sequence $(\tilde{\eta}_n)$ we get from (3.10), (3.11) and (3.12) the dual properties

$$\tilde{\eta}_n = M_Y(\eta_n) = M_Y(\beta_n) \qquad \text{with} \qquad \beta_n = \eta_n^X - \alpha_n = \Pi(\eta_n \,|\, \mathscr{M}_X^1)\,,$$

$$\tilde{\eta}_n^Y = -\log \bar{f}^Y(\beta_n)\,,$$

$$\tilde{\eta}_n = \beta_n + \eta_n^Y\,,$$

(3.19) $$1-\bar f^Y(\beta_n) \le \tilde\eta^Y_n \le -\bar\gamma\,.$$

In the following we use some evident properties (and their duals) of the operator $\bar f$,

$$\bar f(\eta+\eta') = \bar f(\eta)\cdot\exp(\eta')\,, \qquad \eta'\in\mathscr{M}^1,$$
$$\eta\le\eta' \quad\Rightarrow\quad \bar f(\eta)\le\bar f(\eta')\,,\qquad \bar f^Y(\eta)\le\bar f^Y(\eta')\,,$$
$$\bar f^Y(\eta+\gamma) = \bar f^Y(\eta)\cdot\exp(\gamma), \qquad \gamma\in\mathscr{L}^1_Y.$$

From (3.18) and $\alpha_0\le\alpha_{n+1}$ we get

$$\beta_{n+1} = \eta^X_{n+1}-\alpha_{n+1} \le -\bar\beta-\alpha_{n+1} \le -\bar\beta-\alpha_0$$

which implies

$$\bar f^Y(\beta_{n+1}) \le \bar f^Y(-\bar\beta-\alpha_0) = \bar f^Y(-\bar\beta)\cdot\exp(-\alpha_0)$$

and

(3.20) $$-\tilde\eta^Y_{n+1} = \log\bar f^Y(\beta_{n+1}) \le \log\bar f^Y(-\bar\beta)-\alpha_0\,.$$

Putting $q=\exp(\psi)$ we get

$$\bar f(-\bar\beta) = \exp(\bar\gamma+\psi)$$
$$\log\big(\bar f^Y(-\bar\beta)\big) = \log\big[\textstyle\int\exp(\bar\gamma+\psi)\,d\pi_X\big] = \bar\gamma+\log q^Y$$

which, in combination with (3.20) and (3.19), yields

$$(\alpha_0-\bar\gamma-\log q^Y) \le \tilde\eta^Y_{n+1} \le -\bar\gamma\,.$$

This provides the bounds for $\tilde\eta^Y_n$

(3.21) $$\delta_1 := \min(\tilde\eta^Y_0,\alpha_0-\bar\gamma-\log q^Y) \le \tilde\eta^Y_n \le \max(\tilde\eta^Y_0,-\bar\gamma) =: \delta_2\,.$$

From $\tilde\alpha_n\uparrow\alpha$ we obtain the desired bounds for $\gamma_n=\tilde\eta^Y_n-\tilde\alpha_n$,

(3.22) $$\delta_3 := \delta_1-\alpha \le \gamma_n \le \delta_2-\tilde\alpha_0 =: \delta_4\,.$$

Dualizing the arguments leading to (3.21) and (3.22), we get bounds for η^X_n and β_n,

(3.23) $$\delta_5 := \min(\eta^X_0,\tilde\alpha_0-\bar\beta-\log q^X) \le \eta^X_n \le \max(\eta^X_0,-\bar\beta) =: \delta_6\,,$$
$$\delta_7 := \delta_5-\alpha \le \beta_n \le \delta_6-\alpha_0 =: \delta_8\,.$$

Finally, we obtain the bounds for η_n and $\tilde\eta_n$,

(3.24) $$\delta_3+\delta_5 \le \eta_{n+1} = \gamma_n+\eta^X_{n+1} \le \delta_4+\delta_6\,,$$
$$\delta_1+\delta_7 \le \tilde\eta_n = \beta_n+\tilde\eta^Y_n \le \delta_2+\delta_8\,.$$

The above bounds will be used in two ways. First, the bounded functions lie in *compact* sets, e.g. for any $y\in\Omega^Y$ we have

$$\gamma_n(y) \in K_y := [\,\delta_3(y), \delta_4(y)\,] \subset \mathbb{R}$$

and hence

$$\gamma_n \in K := \prod_{y \in \Omega_Y} K_y \subset \mathbb{R}^{\Omega_Y}.$$

Now K is a product of compact sets K_y and by Tychonov's theorem K is compact with respect to the product topology in which convergence of functions is given by *pointwise* convergence.

Second, all bounding functions δ_1, ..., δ_8 are *integrable* if the existence condition (EC1)$'$ and (EC2)$'$ hold, which allows integrating to the limit for pointwise convergent subsequences. Note that the integrability of the *upper* bounds does not require the existence conditions.

3.6 Proof of the Convergence Theorem

In the following proof of the convergence theorem the indices $m(n)$ of a subsequence will simply be written as m, i.e. (η_m) is a subsequence of (η_n). Since the sequence (γ_n) lies in a compact set K there exists a (pointwise) convergent subsequence $k = k(n) > 1$ of the given subsequence (η_m), i.e.

$$\gamma_k \longrightarrow \gamma \in K. \tag{3.25}$$

Now γ is measurable and bounded by integrable functions and hence integrable, i.e. $\gamma \in \mathscr{L}^1_Y$. Furthermore the integrable bounds also provide convergence in the mean in (3.25). The convergence of (γ_k) entails several *pointwise* convergences, some of which - in view of the integrable bounds derived in **3.5** - also imply convergence *in the mean*. First, from (3.25) and (3.22) we sucessively get pointwise convergences,

$$\bar{f}(\gamma_k) \longrightarrow \bar{f}(\gamma),$$

$$\bar{f}^X(\gamma_k) \longrightarrow \bar{f}^X(\gamma),$$

$$\eta^X_{k+1} = -\log \bar{f}^X(\gamma_k) \longrightarrow -\log \bar{f}^X(\gamma), \tag{3.26}$$

$$\eta_{k+1} = M_X(\gamma_k) = \gamma_k - \log \bar{f}^X(\gamma_k) \longrightarrow \gamma - \log \bar{f}^X(\gamma) = M_X(\gamma) =: \eta, \tag{3.27}$$

where the limits $-\log \bar{f}^X(\gamma)$ and η are finite and integrable by (3.23) and (3.24). The convergences in (3.26) and (3.27) also hold *in the mean,* and the latter implies

$$\ell(\eta_{k+1}) = \int \eta_{k+1}\, d\pi - 1 \longrightarrow \ell(\eta) = \int \eta\, d\pi - 1. \tag{3.28}$$

Taking (η_{k+1}) as the wanted subsequence $(\eta_{m'(n)})$ it remains to prove the assertions (*i*) and (*ii*). The argument, that η is π-a.s. independent of the subsequence and of the starting value has already been given. First, we establish (*i*). Since

$$M_X(\eta) = M_X(M_X(\gamma)) = M_X(\gamma) = \eta$$

it remains to show $M_Y(\eta) = \eta$ π_X*-a.s.* and this is by (3.15) equivalent to

$$\ell(\eta) = \ell(M_Y(\eta)). \tag{3.29}$$

Now (3.27) and (3.24) provide pointwise convergence (as well as in the mean) of the projections onto $\mathscr{M}^1_X$

$$\beta_{k+1} = \eta^X_{k+1} - \int \eta_{k+1}\, d\pi \longrightarrow \beta := \eta^X - \int \eta\, d\pi \,.$$

Replacing the sequence (γ_k) by (β_{k+1}) in the arguments leading from (3.25) to (3.27) and (3.28), we get

$$\tilde{\eta}_{k+1} = M_Y(\beta_{k+1}) \longrightarrow M_Y(\beta) \in \mathscr{L}^1 \,,$$
$$\ell(\tilde{\eta}_{k+1}) = \int \tilde{\eta}_{k+1}\, d\pi - 1 \longrightarrow \ell(M_Y(\beta)) = \int M_Y(\beta)\, d\pi - 1 \,.$$

By (3.10) we have $M_Y(\eta) = M_Y(\beta)$ and hence

$$\ell(\tilde{\eta}_{k+1}) \longrightarrow \ell(M_Y(\eta)). \tag{3.30}$$

Finally the sequences $\ell(\eta_{k+1})$ and $\ell(\tilde{\eta}_{k+1})$ have the same limit by (3.17), and thus (3.28) and (3.30) establish (3.29) and hence (*i*).

From (*i*) we obtain $\bar{f}^X(\eta) = 1$ π_X-a.s. and hence

$$\int \bar{f}(\eta)\, d\pi \;=\; \int \bar{f}^X(\eta)\, d\pi^X \;=\; 1.$$

This proves (*ii*) and concludes the proof of the theorem.

To establish the *corollary,* we take η as the limit provided by the theorem for the original sequence (η_n) viewed as its own subsequence. Now any other subsequence of (η_n) contains a further subsequence converging in the mean to a limit which coincides with the above η π-a.s. and hence also converges in the mean to η. This proves convergence in the mean of (η_n) to the limit η and hence

$$\ell(\eta_n) \;=\; \int \eta_n\, d\pi - 1 \;\longrightarrow\; \int \eta\, d\pi - 1 \;=\; \ell(\eta) \,.$$

The sequence $\ell(\eta_n)$ is non-decreasing by (3.16) which proves (*iii*).

It remains to show, that η is the π-a.s.-unique argument which maximizes the functional ℓ. From (*iii*) and (3.16) we first conclude

$$-\infty \;<\; \ell(\eta_0) \;\leq\; \ell(\eta) \,.$$

Now for any $\eta^* \in \mathscr{M}^1$ with $\ell(\eta) \leq \ell(\eta^*)$ we establish $\ell(\eta) = \ell(\eta^*)$ as follows. First

$\ell(\eta^*)$ is *finite* and hence $\bar{f}(\eta^*)$ is π-integrable. Taking $\eta'_0 = M_X(\eta^*)$ as a starting value for a new marginal fitting sequence (η'_n) we already know that that (η'_n) converges in the mean to a limit $\eta' \in \mathscr{M}^1$. From (3.14) and (*iii*) - applied to η'_n - we get

$$\ell(\eta) \;\leq\; \ell(\eta^*) \;\leq\; \ell(\eta'_0) \;\leq\; \ell(\eta')\,.$$

However, by the theorem η and η' coincide π-a.s., and hence $\ell(\eta) = \ell(\eta')$ which proves $\ell(\eta) = \ell(\eta^*)$. As already mentioned in **3.4**, the functional ℓ is strictly concave modulo $\underset{\pi}{\sim}$ which makes η the π-*a.s.-unique* argument maximizing ℓ. $\square$

4. Applications

The complete characterization of the association has interesting consequences and we now look at two typical applications. The first application briefly deals with *association models*, which specify a model for the association function (i.e. odds ratios) of X and Y but leave the marginal distributions completely arbitrary. Motivated by linear logistic regression models we specialize to *log-bilinear associations* (i.e. a bilinear log odds ratio function) for which the existence conditions (EC1) and (EC2) may easily be checked. A somewhat converse approach leads to the second application where we start with given multivariate *normal marginal* distributions for X and Y and characterize *joint normal* distributions for (X,Y) as those with a *log-bilinear association.* As a consequence we obtain a variety of *non-normal* joint distributions with marginal normal distributions. More generally the existence theorem may be used to specify any wanted joint distribution via the marginal distributions and the association function.

4.1 Association Models

From the uniqueness theorem we conclude that the joint distribution P of (X,Y) is determined by their 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 *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 $f = dP/d\pi$ with respect to the product $\pi = \pi_X \times \pi_Y$ of the marginals may be written in terms of the log-density $\varphi = \log f$ and the log odds ratio $\psi^\circ = \log OR$ as (cf. **2.5**)

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

Here ψ° is restricted to a subspace $\Psi^\circ \subset \mathscr{A}^0$ specifying the model and $\alpha^\circ \in \mathbb{R}$ as well as the functions β° and γ° are completely arbitrary. Identifiability of the parameters 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 $\mathscr{A}^0$ 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 $\theta \in \Theta$ (which often has *finite* dimension), i.e. $\Psi^\circ = \{\, \psi^\circ_\theta \,|\, \theta \in \Theta \}$. Assuming the log-density ψ to be π-integrable, the model can be rewritten (cf. **2.6**) as

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

with ψ restricted to a subspace $\Psi \subset \mathscr{A}^1$, and arbitrary $\alpha \in \mathbb{R}$, $\beta \in \mathscr{M}_X^1$ and $\gamma \in \mathscr{M}_Y^1$. An important point, however, is that allowing arbitrary α, β and γ is no guarantee that the model 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 the given marginal distributions may be obtained for suitable values of the "nuisance" parameters α, β and γ.

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

$$\log f^{|X}(y\,|\,x) \;=\; \beta_X^\circ(x) + \gamma^\circ(y) + \psi^\circ(x,y) \qquad \text{with}$$

$$\beta_X^\circ(x) \;=\; -\log \int \exp\left[\gamma^\circ(y) + \psi^\circ(x,y)\right] d\pi_Y(y)$$

or, even simpler, by the log-density *ratio*

$$\log\left(f^{|X}(y\,|\,x)\,/\,f^{|X}(y^\circ\,|\,x)\right) \;=\; \gamma^\circ(y) + \psi^\circ(x,y)\ . \tag{4.1}$$

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 following distributions:

- the conditional distribution of Y given $X = x_i$,
- the conditional distribution of X given $Y = y_i$,
- the joint distribution of X and 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* for arbitrary Ω_Y. Statistical inference for these models, however, is outside the scope of this paper.

4.2 Output with Finite Range

If Ω_Y is finite with $K>1$ elements ($K=1$ is trivial) we may with no loss in generality assume $\Omega_Y=\{0,1,...,K\}$. Then $\mathcal{L}(Y\,|\,X=x)$ 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

$$\operatorname{logit}\pi_k(x) \;=\; \log\left(\pi_k(x)\,/\,\pi_0(x)\right)$$

of the probability vector $\pi(x)$, the association model (4.1) with reference value $y^\circ=0$ reduces to a *logistic regression model*,

$$\operatorname{logit}\pi_k(x) \;=\; \gamma_k^\circ + \psi_k^\circ(x)\,, \qquad\qquad k=1,...,K,$$

where the argument y is replaced by an index k. 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 (the superscript "T" denotes the transpose). Although typically the observation $x=(x_1,...,x_t)$ 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, ... of a „continuous component" x_1 as well as indicator variables $I\{x_2=l\}$ for levels $l=1,...,L$ of a „discrete component" x_2. Introducing indicator variables $h_k(y)=\;I\{y=k\}$ for all values of 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 $\theta_1,...,\theta_K$. 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 $g : \Omega_X \longrightarrow \mathbb{R}^{k_x}$ and $h : \Omega_Y \longrightarrow \mathbb{R}^{k_y}$. Then a *log-bilinear association model* for (X, Y) with respect to (g, h) is given by a $k_x \times k_y$ matrix A such that the log odds ratio function for the joint distribution of (X, Y) is of the form

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

To check the existence conditions for ψ° we first express the projection $\psi = \Pi(\psi^\circ \,|\, \mathscr{A}^1)$ in terms of the expectations $\mu_U = \mathrm{E}(U)$ and $\mu_V = \mathrm{E}(V)$ (which are assumed to exist) as

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

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

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

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

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

Hence the existence condition (EC1)$'$ is here 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 cumulant generating function κ_U of U as

$(\mathrm{EC2})_A$ *The expectation of* $\kappa_U(A[\,V - \mu_V\,])$ *exists (i.e. is finite).*

Since the above considerations depend only on the distributions π_U and π_V of U and V (rather than on π_X and π_Y) we may forget X and Y and start instead with the corresponding log-bilinear association model for the joint distribution of the transformed pair (U, V)

$$\psi^\circ_{UV}(u,v) \;=\; (u - u^\circ)^T A\, (v - v^\circ) \qquad \text{for all } u,\, v$$

which induces the above model for ψ°.

4.4 Multivariate Normal Distributions

Let us now investigate the situation with multivariate *normal* marginal distributions

$$\mathscr{L}(X) = \pi_X = N_{k_x}(\mu_x, \Sigma_x) , \qquad \mathscr{L}(Y) = \pi_Y = N_{k_y}(\mu_y, \Sigma_y) ,$$

in more detail. For simplicity we first assume (the general case is considered later), that the $k_x \times k_x$ resp. $k_y \times k_y$ covariance matrix Σ_x resp. Σ_y is *non-singular* so that X and Y have positive densities f^x_λ resp. f^y_λ with respect to Lebesgue's measure λ^{k_x} in $\mathbb{R}^{k_x}$ resp. λ^{k_y} in $\mathbb{R}^{k_y}$ given by

$$\begin{aligned} \log f^x_\lambda(x) &= -\tfrac{1}{2}\,(x-\mu_x)^T\Sigma_x^{-1}(x-\mu_x) + c_x\,, &\qquad c_x &= -\tfrac{1}{2}\log\left[(2\pi)^{k_x}\det(\Sigma_x)\right] ,\\ \log f^y_\lambda(y) &= -\tfrac{1}{2}\,(y-\mu_y)^T\Sigma_y^{-1}(y-\mu_y) + c_y\,, &\qquad c_y &= -\tfrac{1}{2}\log\left[(2\pi)^{k_y}\det(\Sigma_y)\right] . \end{aligned}$$

Our aim is to characterize the *joint multivariate normal* distributions in $\mathscr{P}^1$ in terms of their association (i.e. odds ratio) functions. Now the product-measure $\pi = \pi_X \times \pi_Y$ is the multivariate normal distribution of dimension $k = k_x + k_y$ with expectation $\mu = (\mu_x, \mu_y)$ and nonsingular covariance matrix $\operatorname{diag}(\Sigma_x, \Sigma_y)$. Since π has a positive density with respect to Lebesgue's measure λ^k we have $\pi \ll \lambda^k \ll \pi$ and hence the class $\mathscr{P}$ consists of all joint distributions P having a positive density with respect to λ^k.

First we take a multivariate normal distribution $P = N_k(\mu, \Sigma) \in \mathscr{P}^1$ and derive its association (odds ratio) function. Since $P \in \mathscr{P}^1$ must have a λ^k-density f_λ, the covariance matrix Σ is nonsingular, and Σ resp. $Q = \Sigma^{-1}$ may be written as

(4.2) $$\begin{aligned} \Sigma &= \begin{bmatrix} \Sigma_x & \Sigma_{xy} \\ \Sigma_{xy}^T & \Sigma_y \end{bmatrix}, &\qquad Q &= \begin{bmatrix} Q_x & Q_{xy} \\ Q_{xy}^T & Q_y \end{bmatrix},\\ Q_x &= \left(\Sigma_x - \Sigma_{xy}\Sigma_y^{-1}\Sigma_{xy}^T\right)^{-1}, &\qquad Q_{xy} &= -\,Q_x\Sigma_{xy}\Sigma_y^{-1},\\ Q_y &= \Sigma_y^{-1} - \Sigma_y^{-1}\Sigma_{xy}^T Q_{xy}\,. \end{aligned}$$

From

$$\log f_\lambda(x,y) = $$
$$-\tfrac{1}{2}\left[(x-\mu_x)^T Q_x (x-\mu_x) + 2(x-\mu_x)^T Q_{xy}(y-\mu_y) + (y-\mu_y)^T Q_y (y-\mu_y)\right] + \mathit{const}$$

we get the log odds ratio function of P as

(4.3) $$\psi^\circ_P(x,y) = -\,(x-x^\circ)^T Q_{xy}(y-y^\circ)$$

which is *bilinear* for the choices $x^\circ = 0$ and $y^\circ = 0$. The projection of ψ°_P onto the association space $\mathscr{A}^1$ is given by

$$\psi_P(x,y) \;=\; -\,(x-\mu_x)^T Q_{xy}(y-\mu_y)$$

and coincides with ψ_P° if the expectations are taken as reference values.

Let us now show that any joint distribution $P \in \mathscr{P}^1$ is *multivariate normal* if and only if its log odds ratio function is *bilinear*. More precisely, we prove the

Characterization Theorem: *Any joint distribution $P \in \mathscr{P}^1$ with marginal normal distributions $\pi_X = N_{kx}(\mu_x, \Sigma_x)$ and $\pi_Y = N_{ky}(\mu_y, \Sigma_y)$ is itself multivariate normal 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 $\psi^\circ(x,y) = x^T A y$ for all x and y.*
Note: *The existence conditions $(EC1)_A$ and $(EC2)_A$ hold if both π_X and π_Y are normal.*

Proof (non-singular case): We first provide a proof for the case that both Σ_x and Σ_y are non-singular from which the general case is derived later. It was just shown above that the log odds ratio function for any such normal P is bilinear.

Conversely, for a given $P \in \mathscr{P}^1$ with normal marginals π_X, π_Y and a bilinear log odds ratio function $\psi^\circ(x,y) = x^T A y$ we have to show that P is *normal*. One possibility is to find the correponding covariance matrix $\mathrm{cov}(X,Y) = \Sigma_{xy}$ as a solution of the equation $A = Q_x \Sigma_{xy} \Sigma_y^{-1}$, where Q_x given above also depends on Σ_{xy}. However, we will apply the theorems of convergence and uniqueness to show that P is normal. In view of **2.4** it is sufficient to consider the special case with $\mu_x = 0$ and $\mu_y = 0$, for which ψ° coincides with its projection: $\psi = \Pi(\psi^\circ \mid \mathscr{A}^1) = \psi^\circ$.

We first note that the assumptions for the convergence theorem are easily verified using the cumulant generating function κ_X of X. Indeed, the expectation of

$$\kappa_X(AY) = \tfrac{1}{2}\, Y^T A^T \Sigma_x A Y$$

exists which establishes $(\mathrm{EC2})_A$. Condition $(\mathrm{EC1})_A$ follows by duality.

We now construct a normal distribution P_0 with log odds ratio function ψ° which not necessarily has the wanted marginal distributions π_X and π_Y. Taking P_0 as a starting value in the iterated marginal fitting procedure, we get in the limit the joint distribution P which will turn out to be normal.

To obtain P_0 take any $a > 0$ such that $C = a^{-1} A$ is a correlation matrix, i.e. $|C_{ij}| < 1$ for all i, j. Now for standard normally distributed vectors $U \sim N_{kx}(0,I)$ and $V \sim N_{ky}(0,I)$ (I denotes a unit matrix) there exists a joint normal distribution P_C such that C is the covariance matrix of (U,V) under P_C. By (4.3) and (4.2) the log odds ratio function of P_C is $\psi_C(u,v) = u^T (BA)\, v$ with

$$B \;=\; a^{-1}(I - C\,C^T)^{-1}.$$

The joint distribution P_0 of (BU, V) under P_C is normal $N_k(0, \Sigma_0)$, and its log odds ratio function turns out to be ψ° (cf. **2.4**). Since $\psi^\circ = \psi$, the log-density $\varphi_0 = \log(dP_0 / d\pi)$ is of the form $\varphi_0 = \eta_0 + \bar{\eta} + \psi$ with $\eta_0 \in \mathscr{M}^1$ and $\bar{\eta} = \bar{\beta} + \bar{\gamma}$ taken from the conditions (EC1), (EC2), cf. **3.2**.

Consider next the iterated marginal fitting sequence $\eta_n = M_X M_Y(\eta_{n-1})$ and the corresponding distributions P_n with log-densities

$$\varphi_n = \log(dP_n / d\pi) = \eta_n + \bar{\eta} + \psi\,.$$

Now for any *normal* centered distribution $P' = N(0, \Sigma)$ the (conditional) distributions $M_X(P')$ and $M_Y(P')$ are normal and centered, too. Hence all P_n are normal $N(0, \Sigma_n)$ and from the convergence theorem we obtain a subsequence $m = m(n)$ such that (η_m) converges (pointwise) to an $\eta \in \mathscr{M}^1$. This implies convergence of the log-densities

$$\varphi_m \;=\; \eta_m + \bar{\eta} + \psi \quad\longrightarrow\quad \eta + \bar{\eta} + \psi \;=: \varphi.$$

Thus the log-densities $\varphi_{\lambda m}$ with respect to Lebesgue's measure λ converge, too,

$$(i) \qquad \varphi_{\lambda m} \;=\; \varphi_m + \log f_\lambda^x + \log f_\lambda^y \quad\longrightarrow\quad \varphi + \log f_\lambda^x + \log f_\lambda^y \;=: \varphi_\lambda$$

Since φ is a log-density of a distribution P_φ with marginals π_X, π_Y and log odds ratio function $\psi^\circ = \psi$, the uniqueness theorem yields $P = P_\varphi$. Hence, for P to be normal, it remains to show that φ_λ is a log-density of a *normal* distribution. Now for any $z = (x, y)$ we have

$$\varphi_{\lambda m}(z) \;=\; -\tfrac{1}{2}\, z^T Q_m z + c_m\,, \qquad Q_m = \Sigma_m^{-1}\,, \qquad c_m = -\tfrac{1}{2} \log[(2\pi)^k \det(\Sigma_m)]\,.$$

Then, by (i)

$$(ii) \qquad c_m = \varphi_{\lambda m}(0) \quad\longrightarrow\quad \varphi_\lambda(0) =: \varphi,$$

$$2[c_m - \varphi_{\lambda m}(z)] = \; z^T Q_m z \quad\longrightarrow\quad 2[c - \varphi_\lambda(z)] \qquad \text{for all } z.$$

Taking z as the unit vector u_i (with $u_{ii} = 1$ and $u_{ij} = 0$ for $i \neq j$) we conclude that the diagonal elements Q_{mii} converge and for $z = u_i + u_j$ we obtain convergence of

$$(u_i + u_j)^T Q_m (u_i + u_j) \;=\; Q_{mii} + 2 Q_{mij} + Q_{mjj}$$

and hence of Q_{mij}. This proves that (Q_m) converges to a matrix Q and

$$\varphi_{\lambda m}(z) = -\tfrac{1}{2}\, z^T Q_m z + c_m \quad\longrightarrow\quad \varphi_\lambda(z) = -\tfrac{1}{2}\, z^T Q z + c$$

for any z. Hence φ_λ is a log-density of a normal distribution, provided Q is non-singular and positive-definite. Now (ii) implies the convergence of

$$\log[\det(Q_m)] = -\log[\det(\Sigma_m)] \quad\longrightarrow\quad a$$

and hence

$$\det(Q_m) \quad \longrightarrow \quad \det(Q) = \exp(a) > 0$$

which proves that Q is non-singular. Since all Q_m are positive-definite, so is their limit Q, which completes the proof. □

As a supplement to the above proof let us remark that the odds ratio function $\exp\{\psi^\circ(x,y)\} = \exp(x^T A y)$ is in general *not* π-*i*ntegrable, i.e. condition (EC3) may be false although (EC1) and (EC2) are satisfied. Indeed, take X and Y both with the same standard multivariate normal distribution $N_k(0,I)$, and, for fixed $t \in \mathbb{R}$, choose A such that $A^T A = 2\,t I$ is a multiple of the unit-matrix I. Then

$$\begin{aligned} \int \exp(x^T A y)\; d\pi(x,y) &= \int \Big[\int \exp(x^T A y)\; d\pi_X(x) \Big]\, d\pi_Y(y) \\ &= \int m_X(A y)\; d\pi_Y(y)\;, \end{aligned}$$

where m_X is the moment-generating function of X, and thus

$$m_X(A y) \;=\; \exp\big(\tfrac{1}{2}\, y^T A^T A y\big) \;=\; \exp\big(t\, y^T y\big)\,.$$

Hence the expectation of $m_X(AY)$ is the moment-generating function of $Y^T Y$ (which has a χ^2_k-distribution) evaluated at t. Thus for $t < \frac{1}{2}$

$$\int \exp(x^T A y)\; d\pi(x,y) \;=\; (1-2t)^{-k/2}\,,$$

but for $t \geq \frac{1}{2}$ the integral is infinite.

So far we have assumed that the covariance matrices Σ_x und Σ_y of X and Y are non-singular. We now extend the above results to the more general case when Σ_x and Σ_y have any rank $r_x > 0$ and $r_y > 0$. Then there exists a $k_x \times r_x$ matrix C_x of rank r_x such that $\Sigma_x = C_x C_x^T$ and for a standardized $U \sim N_{r_x}(0,\, I)$ the linear transform $g(U) = C_x U + \mu_x$ is distributed as $N_{k_x}(\mu_x,\, \Sigma_x)$. Therefore we may assume $X = g(U)$ and, by duality, $Y = h(V) = C_y V + \mu_y$ for a $V \sim N_{r_y}(0, I)$.

Note that g and h have inverses, i.e. $g^{-1}(x) = C_x^-(x - \mu_x)$ with $C_x^- = (C_x^T C_x)^{-1} C_x^T$.

Now any joint normal distribution $P = N_k(\mu, \Sigma)$ for (X,Y) is the image measure of a joint distribution $P' = N_r(\mu', \Sigma')$ for (U,V) under the transformation given by g and h. Furthermore we have $\pi \ll P \ll \pi$ if and only if $\pi' := \pi_U \times \pi_V \ll P' \ll \pi'$. The latter condition is equivalent to $\operatorname{rank}(\Sigma') = r_x + r_y$. Since Σ and Σ' have the same rank, we conclude

$$N_k(\mu, \Sigma) \in \mathscr{P} \qquad \Leftrightarrow \qquad \operatorname{rank}(\Sigma) = r_x + r_y\,.$$

Note that $P \in \mathscr{P}$ will not have a density with respect to Lebesgue's measure λ^k

unless Σ is non-singular.

By (4.3) the log odds ratio function of P' with respect to $u^\circ = g^{-1}(x^\circ)$ and $v^\circ = h^{-1}(y^\circ)$ is given by

$$\psi^\circ_{P'}(u,v) \;=\; -(u-u^\circ)^T Q_{uv}(v-v^\circ)$$

which yields the log odds ratio function of P (cf. **2.4**) as

$$\psi^\circ_P(x,y) \;=\; -(x-x^\circ)^T C_x^{-T} Q_{xy} C_y^{-}(y-y^\circ).$$

Hence even for a *singular* covariance matrix Σ the log odds ratio function ψ°_P is bilinear (assuming $x^\circ = 0$ and $y^\circ = 0$). The characterization theorem can now be established for this more general situation.

Proof (characterization theorem: general case): For normal P we just showed that the log odds ratio function is bilinear. Conversely, suppose $P \in \mathscr{P}^1$ has normal marginals π_X, π_Y of X, Y and a bilinear log odds ratio function $\psi^\circ(x,y) = x^T A y$ with respect to $x^\circ = 0$, $y^\circ = 0$. We have to show that P is normal and may again assume $\mu_x = 0$ and $\mu_y = 0$ in view of **2.4**. Consider the joint distribution P' of $U = C_x^- X$ and $V = C_y^- Y$ which has normal marginals and a bilinear log odds ratio function given by $\psi^{\circ\prime}(u,v) = u^T (C_x^T A C_y) v$. From the non-singular case of the theorem we conclude that P' is normal and hence the joint distribution P of (X,Y) is normal, too. □

Having characterized the joint distributions $P \in \mathscr{P}^1$ which are *normal*, we are now in a position to obtain a variety of *non-normal* joint-distributions $P \in \mathscr{P}^1$ by simply specifying a *non-bilinear* log odds ratio function ψ° which satisfies the assumptions of the existence theorem, e.g. any *bounded* ψ° non-bilinear function. Taking any such non-bilinear function, say $\psi^\circ(x,y) = \sin(x^T A y)$ to be definite, we obtain a *non-normal* joint distribution P whose correlation matrix $\varrho(X,Y)$ exists. Now there also is a joint *normal* distribution P' with the same correlation matrix and marginals. Hence P and P' differ, but have the same correlation matrix and the same marginal distributions. As already remarked in the introduction, this shows that even for *normal marginals* of (X,Y) the correlation matrix does not characterize the association between X and Y.

5. Conclusions

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 $OR(X,Y)$. Specifying each of these three parts separatly yields (under mild integrability conditions) a unique joint distribution. 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.

In particular, *measures* of association as well as *models* for the association may be derived. A *measure* of association is any real-valued functional of the odds ratio function, e.g. the expectation of $OR(X,Y)$ or $\log OR(X,Y)$. *Association models* are (semi-parametric) models which only specify the association structure via the odds ratio function and leave the marginal distributions arbitrary. An important example are *bilinear* log odds ratio functions for vector-space-valued X and Y. Statistical inference of association models may be based on samples from the joint or the conditional distribution of either variable given the other.

Our characterization also allows *classifications* as well as *constructions* of joint distributions according to their odds ratio function and their .marginal distributions. Take for example random vectors X and Y with normal *marginal* distributions. Then their *joint* distribution has shown to be normal if and only if their log odds ratio is a bilinear function. However, a variety of non-normal joint distributions may be specified using non-bilinear log odds ratio functions.

References

Bauer, H. (1991). *Wahrscheinlichkeitstheorie (4. Auflage)*. W. de Gruyter, Berlin.

Bauer, H. (1992). *Maß- und Integrationstheorie (2. Auflage)*. W. de Gruyter, Berlin.

Billingsley, P. (1986). *Probability and Measure (Second Edition)*. J. Wiley & Sons, New York.

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

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

Haberman, S.J. (1974). *The Analysis of Frequency Data.* The University of Chicago Press, Chicago.

Plackett, R.L. (1974). *The Analysis of Categorical Data.* Griffin, London.

Sinkhorn, R. (1967). *Diagonal Equivalence to Matrices With Prescribed Row and Column Sums.* Amer. Math. Mon. 74, 402 - 405

Date: 15-Nov-2000 (printed edition)
24-Sept-2004 (PDF-file, latest version with minor corrections)

Stand Januar 2001 ISSN 0173 - 685 X

Vertrieb der Hefte 4, 14, 23, 26 durch Universitätsbuchhandlung, Bibliothekstr. 3, D-28359 Bremen. Vertrieb der übrigen Hefte (soweit nicht vergriffen) durch die Autoren oder FB 3 Mathematik/Informatik Universität Bremen, Postfach 330440, D-28334 Bremen.

1. Ulrich Krause (1976): Strukturen in unendlichdimensionalen konvexen Mengen, 74 S.
2. Fritz Colonius, Diederich Hinrichsen (1976): Optimal control of heriditary differential systems. Part I, 66 S.
3. Günter Matthiessen (1976): Theorie der heterogenen Algebren, 88 S.
4. H. Wolfgang Fischer, Jens Gamst, Klaus Horneffer (1976): Skript zur Analysis, Band 1 (11. Auflage 2000), 286 S.
5. Wolfgang Schröder (1977): Operator-algebraische Ergodentheorie für Quantensysteme, 59 S.
6. Rolf Röhrig, Michael Unterstein (1977): Analyse multivariabler Systeme mit Hilfe komplexer Matrixfunktionen, 216 S.
7. Horst Herrlich, Hans-Eberhard Porst, Rudolf-Eberhard Hoffmann, Manfred Bernd Wischnewsky (1976): Nordwestdeutsches Kategorienseminar, 193 S.
8. Fritz Colonius, Diederich Hinrichsen (1977): Optimal Control of Hereditary Differential Systems. Part II: Differential State Space Description, 36 S.
9. Ludwig Arnold (1977): Differentialgleichungen und Regelungstheorie, 185 S.
10. Rudolf Lorenz (1977): Iterative Verfahren zur Lösung großer, dünnbesetzter symmetrischer Eigenwertprobleme, 104 S.
11. Konrad Behnen, Hans-Peter Kinder, Gerhard Osius, Rüdiger Schäfer, Jürgen Timm (1977): Dose-Response-Analysis, 206 S.
12. Hans-Friedrich Münzner, Dieter Prätzel-Wolters (1978): Minimalbasen polynomialer Moduln, Strukturindizes und BRUNOVSKY-Transformationen, 53 S.
13. Konrad Behnen (1978): Vorzeichen-Rangtests mit Nullen und Bindungen, 53 S.
14. H. Wolfgang Fischer, Jens Gamst, Klaus Horneffer, Eberhard Oeljeklaus (1978): Skript zur Linearen Algebra, Band 1 (13. Auflage 2000), 249 S.
15. Günter Ludyk (1978): Abtastregelung zeitvarianter Einfach- und Mehrfachsysteme, 54 S.
16. Momme Johs Thomsen (1977): Zur Theorie der Fastalgebren, 146 S.
17. Klaus Horneffer, Horst Diehl (1978): Modellrechnungen zur anaeroben Reduktionskinetik des Cytochroms P-450, 34 S.
18. Horst Herrlich, Rudolf-Eberhard Hoffmann, Hans-Eberhard Porst, Manfred Bernd Wischnewsky (1979): Structure of Topological Categories, 252 S.
19. Hans-Friedrich Münzner, Dieter Prätzel-Wolters (1979): Geometric and moduletheoretic approach to linear systems. Part I: Basic categories and functors, 28 S.
20. Hans-Friedrich Münzner, Dieter Prätzel-Wolters (1979): Geometric and moduletheoretic approach to linear systems. Part II: Moduletheoretic characterization and reachability, 28 S.
21. Eckart Beutler, Hans Kaiser, Günter Matthiessen, Jürgen Timm (1979): Biduale Algebren, 165 S.
22. Horst Diehl, Detlef Harbach, Jürgen Timm (1980): Planung und Auswertung von Atomabsorptions-Spektrometrie-Untersuchungen mit der Additionsmethode, 44 S.
23. H. Wolfgang Fischer, Jens Gamst, Klaus Horneffer (1981): Skript zur Analysis, Band 2 (7. Auflage 2001), 299 S.
24. Horst Herrlich (1981): Categorical Topology 1971-1981, 105 S.
25. Horst Herrlich, Rudolf-Eberhard Hoffmann, Hans-Eberhard Porst, Manfred Bernd Wischnewsky (1981): Special Topics in Topology and Category Theory, 108 S.

26. H. Wolfgang Fischer, Jens Gamst, Klaus Horneffer (1984): Skript zur Linearen Algebra, Band 2 (7. Auflage 1999), 257 S.
27. Rudolf-Eberhard Hoffmann (1982): Continuous Lattices and Related Topics, 314 S.
28. Horst Herrlich, Rudolf-Eberhard Hoffmann, Hans-Eberhard Porst (1987): Workshop on Category Theory, 169 S.
29. Harald Boehme (1987): Zur Berufspraxis des Diplommathematikers, 16 S.
30. Jürgen Timm (1986): Mathematische Modelle der Dosis-Wirkungsanalyse bei den experimentellen Untersuchungen der Arbeitsgruppe zur karzinogenen Belastung des Menschen durch Luftverunreinigung, 65 S.
31. Dieter Denneberg (1988): Mathematik für Wirtschaftswissenschaftler. I. Lineare Algebra, 97 S.
32. Peter E. Crouch, Diederich Hinrichsen, Anthony J. Pritchard, Dietmar Salamon (1988, previous edition University of Warwick 1981): Introduction to Mathematical Systems Theory, 244 S.
33. Gerhard Osius (1989): Some Results on Convergence of Moments and Convergence in Distribution with Applications in Statistics, 27 S.
34. Dieter Denneberg (1989): Verzerrte Wahrscheinlichkeiten in der Versicherungsmathematik, Quantilsabhängige Prämienprinzipien, 24 S.
35. Eberhard Oeljeklaus (1989): Birational splitting of homogeneous Albanese bundles, 30 S.
36. Gerhard Osius, Dieter Rojek (1989): Normal Goodness-of-Fit Tests for Parametric Multinomial Models with Large Degrees of Freedom, 38 S.
37. Dieter Denneberg (1990): Mathematik zur Wirtschaftswissenschaft. II. Analysis, 59 S.
38. Ulrich Krause, Cornelia Zahlten (1990): Arithmetik in Krull monoids and the cross number of divisor class groups, 29 S.
39. Dieter Denneberg (1990): Subadditive Measure and Integral, 39 S.
40. Ulrich Krause, Peter Ranft (1991): A limit set trichotomy for monotone nonlinear dynamical systems, 31 S.
41. Angelika van der Linde (1992): Statistical analyses with splines: are they well defined? 22 S.
42. Dieter Denneberg (1992): Lectures on non-additive measure and integral (new edition: Non-additive measure and integral. TDLB 27, Kluwer Academic, Dordrecht (1994)), 114 S.
43. Gerhard Osius (1993): Separating Agreement from Association in Log-linear Models for Square Contingency Tables With Applications, 23 S.
44. Hans-Peter Kinder, Friedrich Liese (1995): Bremen-Rostock Statistik Seminar, 5. - 7. März 1992, 110 S.
45. Dieter Denneberg (1995): Extension of a measurable space and linear representation of the Choquet Integral, 30 S.
46. Dieter Denneberg, Michael Grabisch (1996): Shapley value and interaction index, 20 S.
47. Angelika Bunse-Gerstner, Heike Faßbender (1996): A Jacobi-like method for solving algebraic Riccati equations on parallel computers, 24 S.
48. Hans-Eberhard Porst editor (1997): Categorical methods in algebra and topology - a collection of papers in honour of Horst Herrlich, 498 S.
49. Angelika van der Linde, Gerhard Osius (1997): Estimation of nonparametric risk functions In matched case-control studies, 28 S.
50. Angelika van der Linde (1997): Estimating the smoothing parameter in generalized spline-based regression, 46 S.
51. Ursula Müller, Gerhard Osius (1998): Asymptotic normality of goodness-of-fit statistics for sparse Poisson data, 15 S.
52. Ursula Müller (1999): Nonparametric regression for threshold data, 18 S.
53. Gerhard Osius (2000): The association between two random elements – A complete characterization in terms of odds ratios, 32 S.