Estimating bivariate distributions assuming some form of dependence

Casper Albers

Abstract

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be an independent random sample from a bivariate population with distribution H. The stochastic variables X and Y are assumed to be (positively) associated in some way. To incorporate this assumption, various mathematical-statistical definitions can be used. We prefer the concept of (positive) quadrant dependence. This thesis contains various methods for estimating the distribution function $H(x, y)$.

Two semiparametric methods are developed and a nonparametric method is discussed. The results are not very promising: though those of the semiparametric methods display various similarities, they are considerably different. This might suggest that samples of size 50 are too small to arrive at acceptable estimates, unless restrictive assumptions are imposed.

Contents

Chapter 1

Introduction

Given the outcome (x_i, y_i), $(i = 1, \ldots, n)$ of an independent random sample from a bivariate distribution, satisfying some assumption of (positive) dependence, we develop semi- and nonparametric estimates of this bivariate distribution and also of the corresponding marginals.

In this chapter an introduction and a short explanation of aims are given and followed by a description of the data set we shall use for illustrating the theory. In Chapter 2 we explain some preparations for making inferences about bivariate distributions. In Chapter 3 semiparametric techniques are presented. Chapter 4 is concerned with nonparametric methods. To apply the theory the computer program Matlab from The Mathworks–company is used.

1.1 The picture quality of video fragments

To illustrate our methods we shall use a data set of the Dutch telecommunications company "Koninklijke PTT Nederland" (KPN). This data set was obtained from Fortuin et al.[7].

The data set consisted of $n = 48$ digitally transmitted fragments of videofilm of different quality. Each fragment was characterized by two quality measurement scores, one based on mechanical recording, the other on human observation. One of the purposes is to investigate whether there is some kind of correlation between these two scoring methods. Each video fragment had a duration between 124 and 127 half seconds. At each half second the picture quality of the fragment was measured by an instrument (for which the scores are assumed to be very precise). The quality was scored as one of the integers $0, \ldots, 7$ and was stored as a binary number with three digits. A score of 0 denotes a very good quality, a score of 7 denotes very bad quality. If the 124 to 127 quality-scores for each half second of the fragment are averaged, we obtain the so-called technical quality, which is a number in the interval $[0, 7]$.

Each video fragment was submitted to a panel of 32 judges. They scored each fragment as an integer between 1 and 5, where 1 denotes very bad quality,
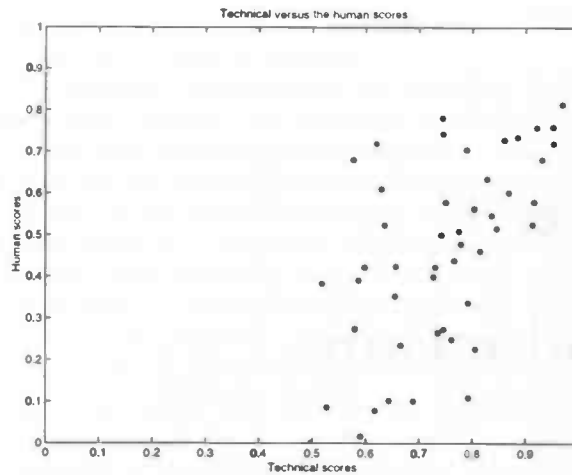
Figure 1.1: The computer-based scores versus the human-based scores

and 5 very good quality. In the data set available to us only the averages of the 32 scores for each fragment were reported. This is a pity because we would have liked to study the inter-observer reliability.

As the technical score is small if the quality is high and the human score behaves in the oppostite manner, we shall apply the transformation

$$x = \frac{7 - \text{technical score}}{7} \quad \text{and} \quad y = \frac{\text{human score} - 1}{4}$$

to our bivariate data set. The distribution of the underlying random variables X and Y on $[0,1] \times [0,1]$ is such that positive dependence seems to be reasonable. A plot of these quality measures is given in Figure 1.1. Although the scores given by the technical instrument lie in principle between 0 and 7, the lowest average score observed was about 3.5 and x was never smaller than $\frac{1}{2}$. In spite of the fact that X and Y are the averages of respectively 124 to 127 and 32 integer-valued random variables, and therefore discrete, we shall continue by regarding X and Y as continuous random variables. This assumption is made for the sake of convenience.

3

Chapter 2

Preparations

In this chapter, we shall look at the bivariate normal distribution which is often used for making (parametric) inferences about bivariate distributions. This is followed by a description of the concept of dependence, with special interest to the concept of positive quadrant dependence. Some methods for testing independence and positive quadrant dependence, using rank correlation coefficients, are evaluated.

2.1 The bivariate normal distribution

Inferences about bivariate distributions are usually based on the assumption of bivariate normality.

Definition 2.1 (Bivariate normal distribution.) *The bivariate normal distribution with* $EX = \xi$, $EY = \eta$, $\mathrm{Var}(X) = \sigma^2$, $\mathrm{Var}(Y) = \tau^2$, $\mathrm{Cov}(X, Y) = \sigma\tau\rho$ *is denoted by* $\mathcal{N}_2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} \sigma^2 & \sigma\tau\rho \\ \sigma\tau\rho & \tau^2 \end{bmatrix} \right)$. *Its density is given by*

$$\varphi(x, y) = \frac{1}{2\pi\sigma\tau\sqrt{1-\rho^2}} \cdot e^{\left[-\frac{1}{2(1-\rho^2)} \left\{ \left(\frac{x-\xi}{\sigma}\right)^2 - 2\rho\left(\frac{x-\xi}{\sigma}\right)\left(\frac{y-\eta}{\tau}\right) + \left(\frac{y-\eta}{\tau}\right)^2 \right\} \right]} \qquad (2.1)$$

This form of the density appeared from the work of Galton who at the end of the nineteenth century studied natural inheritance. Galton presented the data on heights of parents (the mid-parent height) and adult children in the form of a bivariate frequency plot. It is interesting from a causality viewpoint that he plotted the height of children along the x-axis, and not along the y-axis as mathematicians would prefer to do. He described some peculiar features of his data, such as the phenomenon of reversion, which was later referred to as *regression to the mean*, and various types of correlation. Furthermore, Galton noted that the conditional means $E(Y|X = x)$ and $E(X|Y = y)$ seemed to follow straight lines, that the scatter of points is homoscedastic, and that the equiprobability contours are elliptical. With the use of the Cambridge mathematician Dickson,

4

he derived the formula of the bivariate normal distribution which complies with his observations (as can be found in Rao[24]).

Very often the researcher assumes bivariate normality, unless a plot shows that this is clearly inappropiate. The approach based on bivariate normality has the advantage of wide applicability and simplicity. Most of the times this approach gives good results, especially when inferences have to be made about the mean and variance of the actual distribution. Most data, however, are not really generated according to a bivariate normal distribution. Working with the assumption of bivariate normality while it is invalid, may lead to biased opinions. This often leads to the belief that the Galtonian approach has to be replaced by something less restrictive and more complicated.

On the other hand, the main task of the statistician is not to use methods with such complexity and detail that only another statistician can understand his inferences. The statistician has to aim at reasonable results using methods which can be explained to the clients. That is why the possibility of making parametric assumptions will not be ignored. Our purpose is to make inferences on the basis of the outcome of an (independent) random sample from the bivariate distribution $\mathcal{L}(X, Y)$ with distribution function

$$H(x, y) = \mathrm{P}(X \leq x, Y \leq y) \tag{2.2}$$

and density

$$h(x, y) = \frac{\partial^2 H(x, y)}{\partial x \partial y} \tag{2.3}$$

unknown. We do not assume that $H(x, y)$ is close to a bivariate normal distribution. Yet we do assume that X and Y are (positively) correlated in the way to be specified in Section 2.2.

For making inferences based on the outcome of an independent random sample, various approaches have been considered. They can be classified according to whether the approach is parametric, semiparametric, or nonparametric. If normality assumptions have to be made then we, certainly, would have used the usual approach of Galton, elaborated upon by Pearson, Fisher, Rao, Anderson, etcetera. As such assumptions are not realistic for our data set, we shall elaborate on the semiparametric and, ultimately, the nonparametric approach.

2.2 Positive quadrant dependence

The two random variables X and Y are said to be *independent* if

$$H(x, y) = H(x, \infty) H(\infty, y) \quad (-\infty < x, y < \infty) \tag{2.4}$$

If they are not independent then they are dependent. Usually they display a specific kind of systematic dependence, e.g. that the 'correlation', or 'association', is positive in some way. There are many ways to describe that there is

a positive correlation between stochastic variables X and Y. A natural way, of course, is to use the correlation coefficient

$$\rho(X,Y) = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}X\text{Var}Y}} \tag{2.5}$$

If the distribution of (X,Y) is bivariate normal then $\rho(X,Y)$ corresponds to the parameter ρ in Definition 2.1. For the nonparametric analogue of positive correlation, we make use of *positive quadrant dependence* (Lehmann[16]). This concept was called *positive stochastic correlation* in Schaafsma[27]. There are some (nonparametric) measures for positive dependence (Kendall's τ, Spearman's ρ, etc.; see Section 2.3), but in this work we will mainly focus on the concept of positive quadrant dependence (in short PQD). The pair (X,Y) is said to be PQD if the probability that they are simultaneously large (or small) is at least as it would have been in the case of independence[21], more precisely

Definition 2.2 (Positive quadrant dependence.) $\mathcal{L}(X,Y)$ *is said to be positively quadrant dependent if*

$$H(x,y) \geq H(x,\infty) \cdot H(\infty,y) \quad \forall x,y \tag{2.6}$$

There is strict positive quadrant dependence if inequality holds for at least one point (x,y). The stochastic pair (X,Y) is negatively quadrant dependent if (2.6) holds with the '$\geq$'-sign replaced by a '$\leq$'-sign.

We see that PQD is indeed a nonparametric concept since for all (X,Y) PQD and for all continuous increasing functions φ and χ we have that the distribution $\mathcal{L}(\varphi(X),\chi(Y))$ is PQD. A concept slightly stronger than PQD is positive regression dependence or, more specifically, *stochastic positive dependence of Y on X* (Lehmann[15]).

Definition 2.3 (Stochastic positive dependence.) *Let Y_x denote a random variable which has the conditional distribution of Y given x as distribution function. There is stochastic positive dependence of Y on X if*

$$\forall x < x' : \quad P(Y_x \leq z) \geq P(Y_{x'} \leq z) \quad (-\infty < z < \infty; -\infty < x < x' < \infty)$$

Lemma 2.4 *The following conditions on $\mathcal{L}(X,Y)$ are equivalent*

(i) $\mathcal{L}(X,Y)$ *is PQD*

(ii) $P(X \leq x, Y \leq y)P(X > x, Y > y) \geq P(X \leq x, Y > y)P(X > x, Y \leq y)$

(iii) $\text{Cov}(\varphi(X),\chi(Y)) \geq 0$ *for all pairs (φ,χ) of nondecreasing functions such that $\varphi(x)$ and $\chi(x)$ have finite second moments*

Proof. See Appendix A.

6

2.3 Testing independence

The assumption of PQD should not be made unless it is not violated by the data. That is why we test the hypothesis that PQD holds. Usually one starts out by testing the hypothesis H_0 that X and Y are independent. Having rejected H_0, one will proceed by testing $H : \mathcal{L}(X,Y)$ is PQD. This hypothesis should, of course, be maintained unless it is rejected at a reasonable level, e.g. $\alpha = 0.05$. There are many tests for testing the hypothesis H_0 of independence against some form of positive dependence([12], [15]).

In his book [12], Kendall describes how to measure the degree of correspondence between variables, using ranks. His starting point is not the theoretical distribution $\mathcal{L}(X,Y)$ (this is unknown and will always remain unknown though it can be 'approximated' or 'estimated') but the empirical data $(x_i, y_i)\ i = 1, \ldots, n$. When individuals are arranged according to some quality, they are said to be ranked. The arrangement as a whole is called a ranking. We write r_i and s_i to denote the ranks of x_i and y_i respectively $(i = 1, \ldots, n)$. For measuring the degree of correspondence, or the intensity of rank correlation, various coefficients have been proposed.

These coefficients have the following properties:

(i) if the agreement between the rankings is perfect, the coefficient should be $+1$, indicating perfect positive correlation

(ii) if the disagreement between the rankings is perfect, the coefficient should be -1, indicating perfect negative correlation

(iii) for other arrangements, the coefficient should lie between these limiting values and, in some intuitive sense, be increasing if the agreement between the ranks is increasing

One of the earliest and most widely used methods of correlation when data are in the form of ranks is due to Spearman(1904). He proposed his rank-order correlation coefficient, which we shall denote by ρ_S. This is the product-moment correlation coefficient between R and S (the 'rank-representations' of X and Y). When we denote $r_i - s_j$ with d_{ij}, we can compute Spearman's ρ using

$$\rho_S = \frac{6 \sum d_{ij}^2}{n(n^2 - 1)} \tag{2.7}$$

To test the statistical significance of this coefficient we can use the following asymptotic test for the null hypothesis of independence

$$t = \rho_S \sqrt{(n-2)/(1 - \rho_S^2)} \sim t_{n-2} \tag{2.8}$$

This test (Lindeman et al.[17], p. 66) provides satisfactory approximations when $n \geq 10$. A useful observation is that ρ_S has *exact* mean 0 and variance $\frac{1}{n-1}$, if the null hypothesis is true. (Note for $n = 2$ that ρ_S is either $+1$ or -1). The approximation based on referring ρ_S to $\mathcal{N}(0, \frac{1}{n-1})$ is almost the same in practice

as that based on (2.8). Since in our case $n = 48$ and the outcome of the test is $t = 8.73$, approximation (2.8) obviously suffices.

Another useful rank correlation coefficient is Kendall's τ, proposed in 1948. This coefficient is based on the extent of agreement between judges in their relative orderings of all possible pairs of individuals. An agreement occurs when both orderings are the same for a pair, the pair of judges can then be said to be concordant. Kendall's τ can be computed by counting the number of concordances (n_c) and the number of discordances (n_d) among all possible pairs, and dividing by the number of pairs,

$$\tau = \frac{n_c - n_d}{\binom{n}{2}} = \frac{4n_c}{n(n-1)} - 1 \tag{2.9}$$

The statistical significance of τ can be tested by computing the variance of τ which under H_0 is equal to

$$\text{var}(\tau) = \frac{2(2n+5)}{9n(n-1)}$$

The test can then be computed as

$$z = \frac{\tau}{\sqrt{\text{var}(\tau)}} \tag{2.10}$$

which has, approximately, the standard normal distribution (when there are no ties) under the null hypothesis of independence (Lindeman et al.[17], p. 69). Again, since $n = 48$ and $z = 8.69$, this approximation also suffices.

An easy way to display τ and ρ_S (and other coefficients) is by using the general correlation coefficient

$$\Gamma = \frac{\sum a_{ij} b_{ij}}{\sqrt{\sum a_{ij}^2 \sum b_{ij}^2}} \tag{2.11}$$

For every pair of individuals an x-score, denoted by a_{ij} will be allocated, subject only to the conditions $a_{ij} = -a_{ji}$ and $a_{ii} = 0$. Simultaneously, y-scores will be allocated and denoted by b_{ij}. Note that Pearson's product-moment correlation-coefficient

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \tag{2.12}$$

arises if one takes $a_{ij} = x_i - x_j$ and $b_{ij} = y_i - y_j$. Kendall's τ, is based on

$$a_{ij} = \begin{cases} +1 & r_i < r_j \\ -1 & r_i > r_j \end{cases}$$

$$b_{ij} = \begin{cases} +1 & s_i < s_j \\ -1 & s_i > s_j \end{cases}$$

and Spearman's ρ is obtained if

$$a_{ij} = r_j - r_i$$

8

$$b_{ij} = s_j - s_i$$

These representations of τ and ρ_S only hold when there are no ties, i.e. no x's or y's with identical values. Generalizations are available to determine representations for the coefficients when observations are tied. However, we only want to get a general idea about the correspondence between X and Y. Since almost no ties are present in the data at hand, the possibility of ties is ignored.

Kendall and, in The Netherlands, Van Dantzig, Terpstra, Smid, Ruymgaart, and others have studied the distribution of and the relation between τ (and ρ_S), both in the case of stochastic independence and in the case of dependence. Normal approximations to the distribution of τ and ρ_S under H_0 (with continuity correction) are very accurate if n is large, say $n > 10$. Descriptions of the relation between τ and ρ_S exist, e.g.

$$\frac{3}{2}\tau - \frac{1}{2} \leq \rho_S \leq \frac{1}{2} + \tau - \frac{1}{2}\tau^2 \tag{2.13}$$

for large values of n (see Kendall[12], p.13). This 'interval' for ρ_S should hold when $\tau > 0$.

We have computed our τ and ρ_S and tested the hypothesis of no independence. For our data set about video fragments, we obtained

$$\tau = 0.867$$
$$\rho_S = 0.790$$

In both cases the hypothesis of independence is rejected at 'any' level of significance. We also checked the inequalities in (2.13). It follows that ρ_S lies in the interval $[0.80, 0.99]$. This is not the case, probably because of the presence of tied observations in our data.

2.4   Testing PQD

The assumption of positive quadrant dependence should not be made if it is in clear conflict with the data. To test the hypothesis that $\mathcal{L}(X, Y)$ is PQD we can consider all hypotheses of the form

$$H_{x,y} : \mathrm{P}(X \leq x, Y \leq y) \geq \mathrm{P}(X \leq x)\mathrm{P}(Y \leq y) \tag{2.14}$$

Note that $H = \bigcap_{x,y} H_{x,y}$. To test $H_{x,y}$ at significance level α it is natural to use Fisher's exact test. For such $(x, y) \in R^2$ the corresponding 2×2 table

$$a = \# \{i | x_i \leq x, y_i \leq y\}, \quad b = \# \{i | x_i \leq x, y_i > y\}$$
$$c = \# \{i | x_i > x, y_i \leq y\}, \quad d = \# \{i | x_i > x, y_i > y\}$$

is composed. The hypothesis (2.14) is rejected if and only if

$$T(x, y) := \frac{ad - bc}{\sqrt{\frac{1}{n-1}(a + c)(b + d)(a + b)(c + d)}} < -u_\alpha \tag{2.15}$$

where $u_\alpha = \Phi^{-1}(1-\alpha)$ is the upper α point of the standard-normal distribution.

Our hypothesis H holds iff $H_{x,y}$ holds for all pairs (x,y). By applying (2.15) to all combinations of (x_i, y_i) we perform n^2 tests, each approximately of level α. The overall probability of rejecting $H_{x,y}$ for some (x,y), if H is true, depends on $\mathcal{L}(X,Y)$ and is difficult to determine. One might study the distribution of $\min_{x,y} T(x,y)$ and study whether independence is some sort of 'least favorable' situation. This study goes beyond the aims of this work and will not be done.

The present data is such that we make the assumption that X and Y are positively associated. For the data set about video fragments none of the $H_{x,y}$ is rejected at significance levels above $\frac{1}{2}\%$.

10

Chapter 3

Estimating the bivariate density semiparametrically

This chapter is about estimating a bivariate density using semiparametric techniques. This means that we shall not make the parametric assumption that our bivariate distribution $\mathcal{L}(X,Y)$ is an element from some parametric family $\mathcal{P} = \{P_\theta \mid \theta \in \Theta\}$. For the specification of the dependence structure, however, one parameter will be worked with. Two approaches, one using the bivariate normal distribution, and the other using the bivariate exponential distribution, are considered.

3.1   Introduction

Our aim is to estimate the bivariate density $h(x,y)$ of (X_1, Y_1) on $[0,1] \times [0,1]$ on the basis of the outcome (x_i, y_i) $(i = 1, \ldots, n)$ of a random sample and also to estimate the corresponding distribution function $H(x,y)$. For the marginal distributions of X and Y we use the following notations: $F(x) := H(x, \infty) = P(X \leq x, Y \leq \infty) = P(X \leq x)$ and $G(y) := H(\infty, y)$. The marginal densities are denoted by $f(x) = F'(x)$ and $g(y) = G'(y)$. It is assumed that $\mathcal{L}(X,Y)$ is (strictly) PQD, see Section 2.2. The estimation of the marginal distributions of X and Y will be done nonparametrically. For modelling the dependence between X and Y the parametric assumption will be used in Section 3.3 that (for some $\rho > 0$)

$$\mathcal{L}\left(\begin{array}{c} \Phi^{-1}(F(X)) \\ \Phi^{-1}(G(Y)) \end{array}\right) = \mathcal{N}_2\left(\left(\begin{array}{c} 0 \\ 0 \end{array}\right), \left[\begin{array}{cc} 1 & \rho \\ \rho & 1 \end{array}\right]\right) \tag{3.1}$$

It is trivial that $F(X) \sim \mathcal{U}(0,1)$ and $G(Y) \sim \mathcal{U}(0,1)$ and, hence, the distributions of $\Phi^{-1}(F(X))$ and $\Phi^{-1}(G(Y))$ are $\mathcal{N}(0,1)$. The dependence is modelled efficiently by only one parameter ρ. This is dangerous because reality will almost always be different. If one feels forced to reject the assumption of bivariate

normality then it is difficult to decide upon something else. NEvertheless, in Section 3.4 a alternative approach is considered. This indicates that many possibilities exist, but that it is difficult to choose.

The value of ρ will be estimated from the data. If the dependence is not of the form specified by (3.1), we shall make systematic errors. The data, plotted in Figure 1.1 are such that bivariate normality is not an acceptable assumption. We hope that the more flexible semiparametric model (3.1) will provide useful results.

3.2 Estimating the marginal distributions

For estimating the marginal distribution functions F and G we use two different approaches. The first and simplest approach is to use the empirical marginals (see Grimmett and Stirzaker[9], p. 387) with distribution function

$$\tilde{F}(x) = \frac{1}{n} \#\{i \mid x_i \leq x\} \quad 0 \leq x \leq 1 \tag{3.2}$$

For our data the empirical distribution function $\tilde{G}$ estimating G can be found in the same way. Figure 3.1 displays the empirical distribution functions for the two variables of our data set. Note that $\tilde{F}$ and $\tilde{G}$ are discontinuous functions displaying jumps $\frac{1}{n}$ in the order statistics and being constant elsewhere.
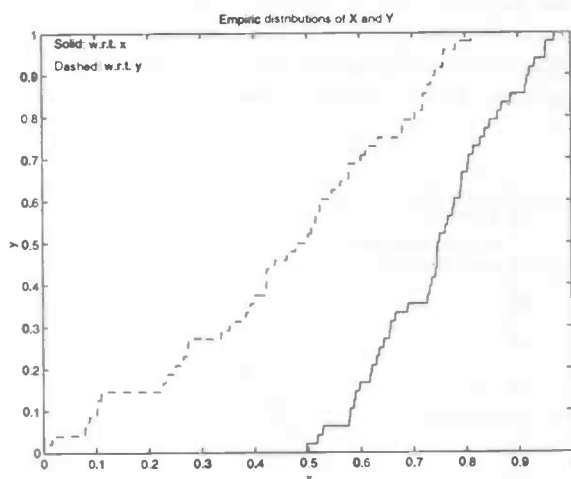

Figure 3.1: Empirical distributions of X and Y

The second approach is based on the idea that the true distribution functions F and G will be continuous and even differentiable with derivatives $f = F'$ and $g = G'$. The theory of nonparametric density estimates can then be applied (see Silverman[30]). We shall use the new and somewhat peculiar method described

in De Bruin and Schaafsma[5]. This semi-Bayesian method provides a smooth estimate of the inverse of the distribution of a (univariate) random variable. Let $x_{[1]}, \ldots, x_{[n]}$ denote the ordered outcomes of the sample from the distribution F. For the support of X we write $[x_{[0]}, x_{[n+1]}]$ and we assume that the values $x_{[0]}$ and $x_{[n+1]}$ have been prescribed. The method provides

$$B_n^{(1)}(p) = \sum_{i=0}^{n+1} x_{[i]} \left(\begin{array}{c} n+1 \\ i \end{array} \right) p^i (1-p)^{n+1-i} \qquad (3.3)$$

where $B_n^{(1)}(p)$ is an estimate of $B^{(1)}(p) = F^{-1}(p)$. Analogously, $B_n^{(2)}$ is constructed to estimate $G^{-1}(p)$. For our data set the observations are presented as crosses and circles, and the two estimated distribution functions, $\hat{F} = B_n^{(1)-1}$ and $\hat{G} = B_n^{(2)-1}$, are given by the dotted curves in Figure 3.2. To start with, the choice $[x_{[0]}, x_{[n+1]}]=[0, 1]$ is made (see Section 1.1). We see that these results are nice and smooth, but we also see that some 'tails' are unsatisfactorily large, especially the right-hand tail of Y and the left-hand tail of X. Improvement is possible by using a more precise specification of the supports which, of course, should extend beyond $[x_{[1]}, x_{[n]}] \times [y_{[1]}, y_{[n]}]$. De Bruin and Schaafsma[5] give various methods for specifying the supports, depending on whether the support must be finite or not. One of the suggestions is to use $x_{[0]} = x_{[1]} - (x_{[2]} - x_{[1]})$ and $x_{[n+1]} = x_{[n]} - (x_{[n]} - x_{[n-1]})$. For X this provides $[0.4755, 0.9860]$, but for Y the interval $[-0.0160, 0.8438]$ obtained will be modified by taking 0 as the left boundary of the support. The estimates $\hat{F}$ and $\hat{G}$ for the marginal distribution functions provided by this procedure are given by the solid curves in Figure 3.2. It is obvious that both methods for estimation differ very little in the 'middle' of the distributions. The modified method is preferred because it seems more accurate.
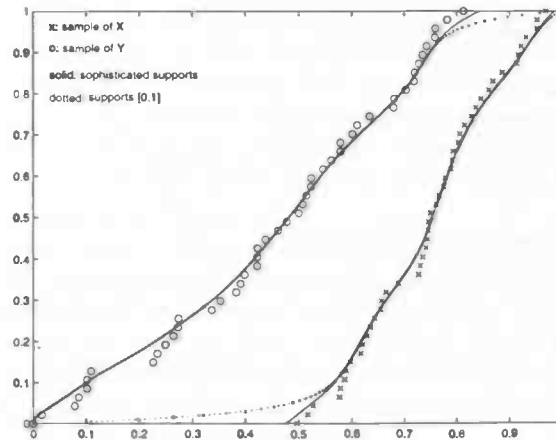

Figure 3.2: Estimates of $F(x)$ and $G(Y)$ based on $B_n^{(1)}$ and $B_n^{(2)}$

13

3.3 Estimating $\mathcal{L}(X, Y)$ using the normal distribution

Formula (3.1) displays how we will model the dependence between X and Y in a parametric way. We shall now construct the estimate $\hat{\rho}$ for the product-moment correlation coefficient ρ by using the transformations

$$(x_i, y_i) \mapsto (u_i, v_i) := (\Phi^{-1}(\hat{F}(x_i)), \Phi^{-1}(\hat{G}(y_i))) \ (i = 1, \ldots, n) \tag{3.4}$$

and computing the sample covariance providing

$$\hat{\rho} = \frac{1}{n} \sum_{i=1}^{n} u_i v_i \tag{3.5}$$

As the points with smallest and largest ranks play a very important role in this product-moment correlation coefficient, it is pertinent to use the most appropriate supports. If this is done as indicated, we find $\hat{\rho} = 0.4894$.

The distribution $\mathcal{L}(X, Y)$ will be estimated by computing the distribution of

$$\begin{pmatrix} \tilde{X} \\ \tilde{Y} \end{pmatrix} = \begin{pmatrix} \tilde{F}^{-1}(\Phi(Z_1)) \\ \tilde{G}^{-1}(\Phi(Z_2)) \end{pmatrix} \tag{3.6}$$

where

$$\mathcal{L} \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} = \mathcal{N}_2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} 1 & \hat{\rho} \\ \hat{\rho} & 1 \end{bmatrix} \right) \tag{3.7}$$

Note that it is not obligatory to take for $\tilde{F}$ and $\tilde{G}$ the estimated distribution functions of the previous section. As the joint density of (Z_1, Z_2) is given by

$$\varphi(z_1, z_2) = \frac{1}{2\pi\sqrt{1 - \hat{\rho}^2}} \cdot \exp\left[-\frac{1}{2(1 - \hat{\rho}^2)} \left\{ z_1^2 - 2\hat{\rho} z_1 z_2 + z_2^2 \right\} \right] \tag{3.8}$$

the estimated joint distribution of $\tilde{X}$ and $\tilde{Y}$ has distribution function $\hat{H}_n(x, y) =$

$$\int_{-\infty}^{\Phi^{-1}(\hat{F}(x))} \int_{-\infty}^{\Phi^{-1}(\hat{G}(y))} \frac{1}{2\pi\sqrt{1 - \hat{\rho}^2}} \cdot e^{\left[-\frac{1}{2(1 - \hat{\rho}^2)} \left\{ u^2 - 2\hat{\rho}uv + v^2 \right\} \right]} \, dv du \tag{3.9}$$

and density $\hat{h}_n(x, y) =$

$$= \frac{1}{2\pi\sqrt{1 - \hat{\rho}^2}} \frac{\partial^2}{\partial x \partial y} \int_{-\infty}^{\Phi^{-1}(\hat{F}(x))} \int_{-\infty}^{\Phi^{-1}(\hat{G}(y))} e^{\left[\frac{-1}{2(1 - \hat{\rho})^2} \left\{ u^2 - 2\hat{\rho}uv + v^2 \right\} \right]} \, du dv$$

$$= \frac{1}{2\pi\sqrt{1 - \hat{\rho}^2}} e^{\left[\frac{-1}{2(1 - \hat{\rho}^2)} \left\{ (\Phi^{-1}(\hat{F}(x)))^2 - 2\hat{\rho}(\Phi^{-1}(\hat{F}(x)))(\Phi^{-1}(\hat{G}(y))) + (\Phi^{-1}(\hat{G}(y)))^2 \right\} \right]}$$

$$\cdot \frac{\partial}{\partial x}(\Phi^{-1}(\hat{F}(x))) \frac{\partial}{\partial y}(\Phi^{-1}(\hat{G}(y))) \tag{3.10}$$

$$= \frac{1}{\sqrt{1 - \hat{\rho}^2}} e^{\left[\frac{-1}{2(1 - \hat{\rho}^2)} \left\{ (\Phi^{-1}(\hat{F}(x)))^2 - 2\hat{\rho}(\Phi^{-1}(\hat{F}(x)))(\Phi^{-1}(\hat{G}(y))) + (\Phi^{-1}(\hat{G}(y)))^2 \right\} \right]}$$

$$\cdot e^{-\left[\frac{1}{2}(\hat{F}(x))^2 + \frac{1}{2}(\hat{G}(y))^2 \right]} \hat{f}(x) \hat{g}(x) \tag{3.11}$$
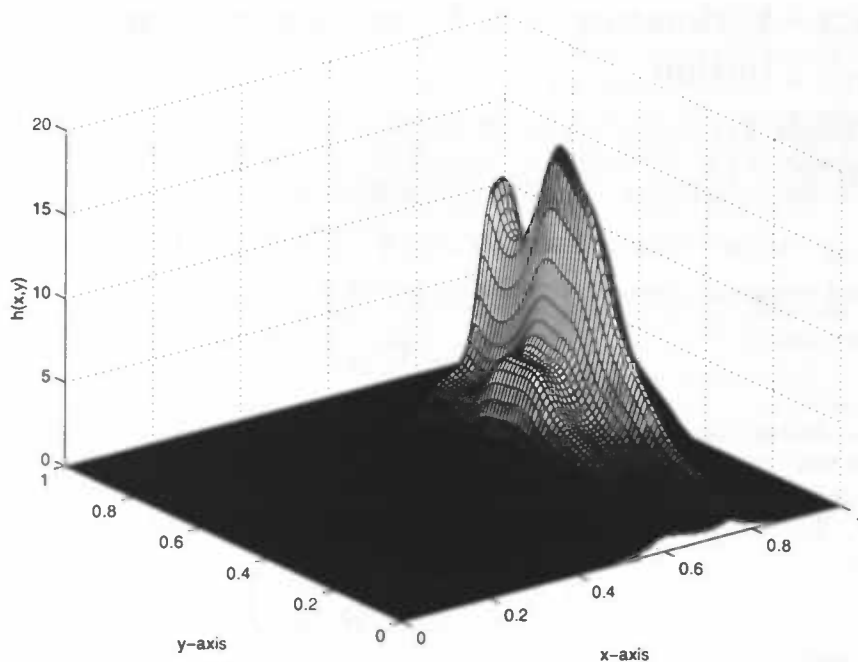
Figure 3.3: Estimate of the bivariate density using the $\mathcal{N}_2$ distribution

Here, $\hat{F}$ and $\hat{G}$ are smooth estimators of F and G, not necessarily equal to the earlier mentioned estimators $\tilde{F}$ and $\tilde{G}$. Expression (3.11) is complex but suitable for computation. The precise shape of $\hat{h}_n(x, y)$ depends to a considerable extent on the approach we use for estimating the marginal distributions F and G.

To compute the estimates of F, G, f and g for the data set of video fragments, the Bernstein polynomial estimates for F^{-1} and G^{-1} were computed in 1600 equidistant point on $[0, 1]$. By linear interpolation and numerical differentiation, the estimates for F and G and f and g were computed. These estimates are sufficiently accurate for the calculation of h_n. A 3d-surface plot of this estimated bivariate density is given in Figure 3.3.

Figure 3.4 is the 'view from above' of Figure 3.3. When the points with the same color are connected, one gets the euqiprobability curves. These are the level curves corresponding to a certain value of $\hat{h}_n(x, y)$. The probability that an observation (x, y) lies inside an area I is, of course, $\int_I h_n(x, y)\, \mathrm{d}x\mathrm{d}y$. For the level curves corresponding to $\hat{h}_n = 1.3, 5.0, 8.2$, and 13.8, these probabilities are 93%, 37%, 20%, and 6.8%, respectively. The corresponding observed frequencies (see Figure 1.1) are 88%, 48%, 27%, and 13%.

As can be seen from these plots, our estimate for the bivariate density has a shape more detailed and sophisticated than when the density is estimated by a bivariate normal one, but still has the nice smoothness property. It is
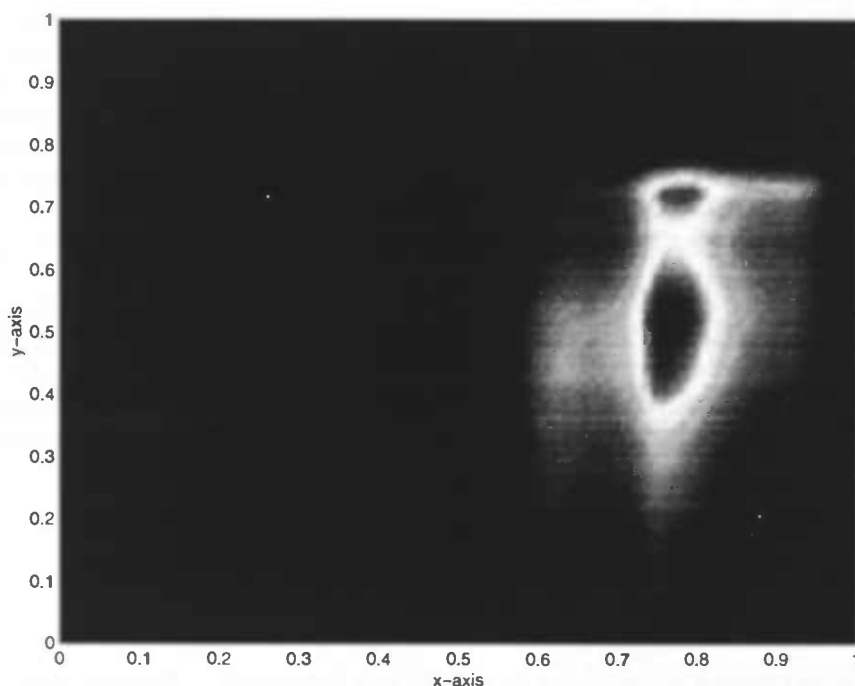
Figure 3.4: Height plot corresponding to Figure 3.3

an interesting question whether the bimodality is real or apparent. For that purpose the statistical accuracy of the estimates should be studied. This goes beyond the present work.

3.4 Estimating $\mathcal{L}(X, Y)$ using the bivariate exponential distribution

Similarly to the previous section, the bivariate density is estimated using a parametrisation only for the correlation. This time a transformation to exponentially distributed variables is used. We define the following mapping

$$(x_i, y_i) \mapsto (u_i, v_i) := (-\log(1 - x_i), -\log(1 - y_i)) \ (i = 1, \ldots, n) \qquad (3.12)$$

The u_i and v_i can be considered as taken from random variables U and V, both having the standard-exponential density with mean and variance one.

The distribution $\mathcal{L}(X, Y)$ will be estimated by computing the distribution of

$$\begin{pmatrix} \tilde{X} \\ \tilde{Y} \end{pmatrix} = \begin{pmatrix} \tilde{F}^{-1}(1 - e^{-U}) \\ \tilde{G}^{-1}(1 - e^{-V}) \end{pmatrix} \qquad (3.13)$$

16

where the joint distribution of (U, V), denoted with $H_{U,V}$, is one with both marginals exponentially (mean 1) distributed. There are several bivariate distributions with such exponential marginals, see for example Gupta et al.[10]. We have chosen to work with the bivariate exponential distribution introduced by Marshall and Olkin[19] (see also [2], [4] and [10]). This is one of the most frequently used bivariate exponential distributions, and implying it also takes the PQD into account (see section 4.2). The distribution has the form

$$H_{U,V}(u, v) = e^{(-\lambda_1 u - \lambda_2 v - \lambda_{12} \max(u,v))} \tag{3.14}$$

and is called the $\text{BVE}(\lambda_1, \lambda_2, \lambda_{12})$ distribution. The marginal distributions are $U \sim \text{Exp}(\lambda_1 + \lambda_{12})$ and $V \sim \text{Exp}(\lambda_2 + \lambda_{12})$ (Basu[2]). The correlation coefficient is equal to $\rho = \lambda_{12}/(\lambda_1 + \lambda_2 + \lambda_{12})$(Brady et al.[4]). Since U and V are constructed such that they both follow an exponential distribution with mean 1, we have that $\lambda_1 = \lambda_2 = 1 - \lambda_{12}$. The correlation coefficient ρ can be estimated from the data and therefore we can use the estimates

$$\hat{\lambda}_{12} = \frac{2\hat{\rho}}{1 + \hat{\rho}}$$

and

$$\hat{\lambda}_1 = \hat{\lambda}_2 = \frac{1 - \hat{\rho}}{1 + \hat{\rho}}$$

For our data $\hat{\rho} = 0.62$, $\hat{\lambda}_1 = \hat{\lambda}_2 = 0.23$, and $\hat{\lambda}_{12} = 0.77$. The joint density of (U, V) is given by

$$h_{U,V}(u, v) = \begin{cases} \frac{1}{2}(1 + \lambda_1)e^{[-\lambda_1 u - \lambda_1 v - (1 - \lambda_1)\max(u,v)]} & u, v \geq 0 \\ 0 & \text{elsewhere} \end{cases} \tag{3.15}$$

and (when $0 \leq \lambda_1 \leq 1$) this is indeed a bivariate probability distribution function. So the estimated distribution function of $\hat{X}$ and $\hat{Y}$ is

$$\hat{H}(x, y) = \int_0^{-\log(1 - \hat{F}(x))} \int_0^{-\log(1 - \hat{G}(y))} (1 + \hat{\rho})^{-1}\exp\left[-\frac{1 - \hat{\rho}}{1 + \hat{\rho}}(u + v)\right.$$
$$\left. - \frac{2\hat{\rho}}{1 + \hat{\rho}}\max(u, v)\right] dudv \tag{3.16}$$

with density $\hat{h}(x, y) = \frac{\partial^2}{\partial x \partial y}\hat{H}(x, y) =$

$$= (1 + \hat{\rho})^{-1}\left[\exp\left\{-\frac{1 - \hat{\rho}}{1 + \hat{\rho}}(-\log(1 - \hat{F}(x) - \log(1 - \hat{G}(y)) - \frac{2\hat{\rho}}{1 + \hat{\rho}}\cdot\right.\right.$$
$$\left.\max(-\log(1 - \hat{F}(x)), -\log(1 - \hat{G}(y)))\right\}\right] \cdot \frac{\partial}{\partial x}\left(-\log(1 - \hat{F}(x))\right)$$
$$\cdot \frac{\partial}{\partial y}\left(-\log(1 - \hat{G}(y))\right) \tag{3.17}$$

$$= (1 + \hat{\rho})^{-1}\left[\exp\left\{-\frac{1 - \hat{\rho}}{1 + \hat{\rho}}(-\log(1 - \hat{F}(x) - \log(1 - \hat{G}(y)) - \frac{2\hat{\rho}}{1 + \hat{\rho}}\cdot\right.\right.$$
$$\left.\max(-\log(1 - \hat{F}(x)), -\log(1 - \hat{G}(y)))\right\}\right] \cdot \frac{\hat{f}(x)}{1 - \hat{F}(x)}\frac{\hat{g}(y)}{1 - \hat{G}(y)} \tag{3.18}$$
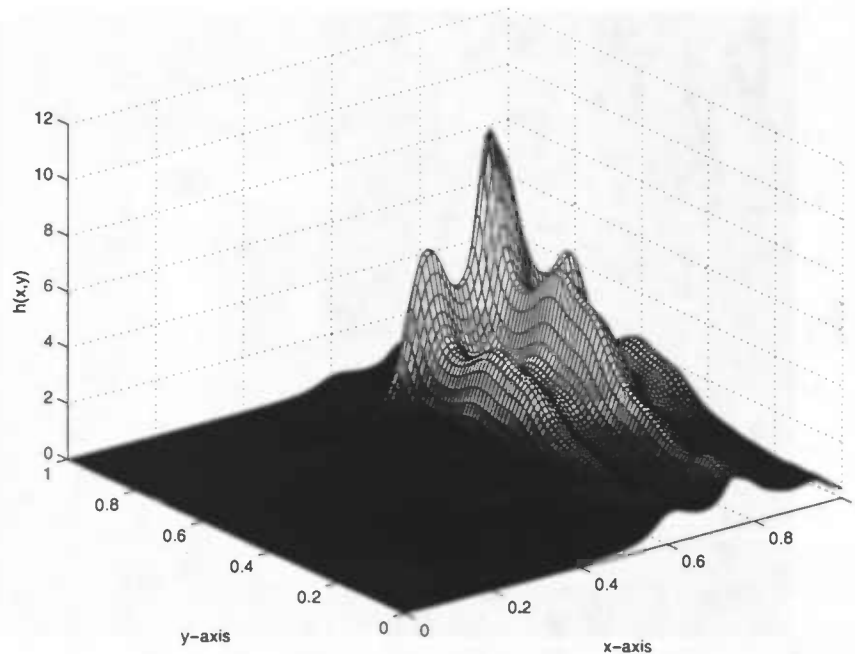
Figure 3.5: Estimated bivariate density using the bivariate exponential distribution

Just like in the previous section, it is assumed that $\hat{F}$ and $\hat{G}$ are smooth estimators for F and G. Simultaneously to the approach in that section, we have calculated $\hat{F}, \hat{G}, \hat{f}, \hat{g}$, and $\hat{\rho}$.

In Figure 3.5 the 3d-surface plot of our estimated bivariate density is given. This is again a nice, smooth density, but with more peaks than in Figure 3.3. Figure 3.6 displays the 'overview' corresponding to Figure 3.5. Again, we have calculated some estimated probabilities corresponding to some equiprobability curves. The probabilities corresponding to the curves for respectively $\hat{h} = 1.4, 3.2, 5.0$, and 7.3 are respectively $75\%, 37\%, 14\%$, and 2.8%. The corresponding frequencies (see Figure 1.1) are $79\%, 44\%, 17\%$, and 2.1%.

3.5 Conclusions

The methods described in Sections 3.3 and 3.4 both have only one parameter on which the inferences are dependent. Although the results of both methods have some similarities, they are, unfortunately, quite different. The estimated probabilities that an observation lies inside some equiprobability curve also differs much from observed frequencies. It seems that the family of models that is used has a very strong influence on the inference. One might want to reduce
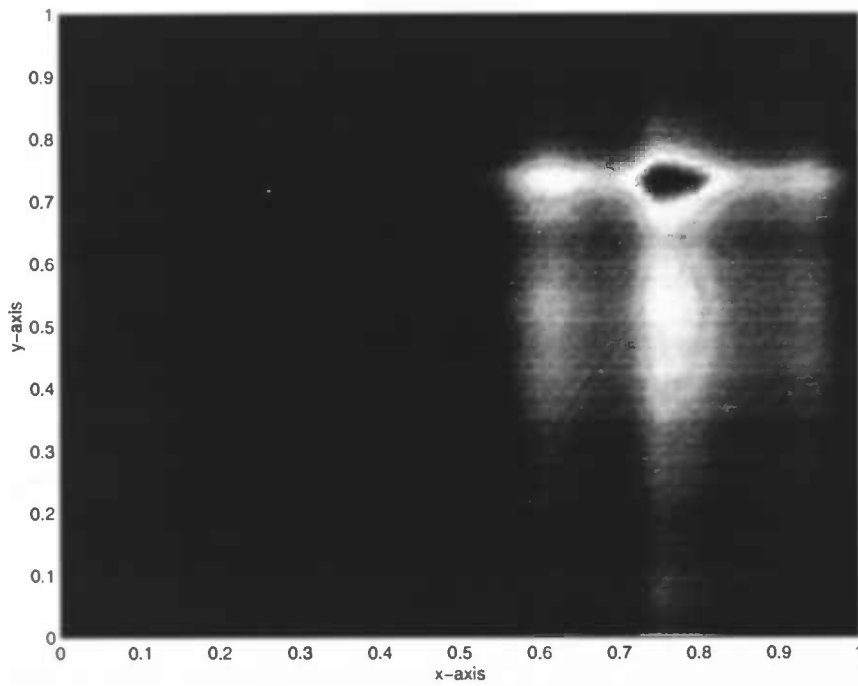
18

Figure 3.6: Height plot corresponding to Figure 3.5

this problem by introducing more parameters, but this results into a parametric, complex model. The two methods shall give more similar results when the sample size is larger than that we used ($n = 48$).

Chapter 4

Nonparametric dependence concepts

In this chapter some nonparametric extensions are considered. Nonparametric inferences are inferences for which the family of distributions is the family of *all* possible probability distributions.

One estimate for the bivariate distribution function is easily obtained: use the (bivariate) empirical distribution function

$$\tilde{H}(x,y) = \frac{1}{n} \#\{i|x_i \leq x, y_i \leq y\} \quad 0 \leq x, y \leq 1 \tag{4.1}$$

For our data set about video fragments, the bivariate empirical distribution function is displayed in Figure (4.1). Nonparametric density estimates can be obtained along the lines described in the literature: kernel methods, wavelets, etc.

In De Bruin and Schaafsma[5], a method is derived to obtain an estimator for the quantile function $F^{-1}(x)$ in the univariate case. Attempts to make a 2-dimensional generalization failed. In this chapter we discuss some relevant literature which, hopefully, will result in something useful in the future.

4.1 Concepts describing bivariate positive dependence

In this master thesis, we mainly focused on positive quadrant dependence to describe some sort of positive relationship between two variables, since this concept appeared to us to be the 'most natural' way to describe such dependence. Of course there are many other ways to define some sort of positive bivariate dependence. While the concept of independence is mathematically defined by an equality relation, the violation of this equality by definition signifies dependence. In Kotz et al.[14], seven different methods are evaluated. Amongst these methods are
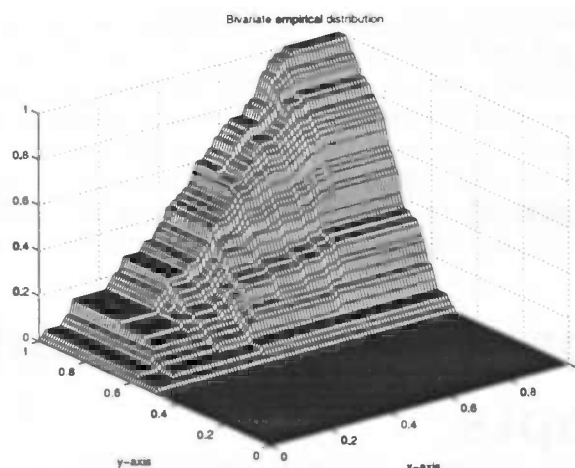
20

Figure 4.1: Bivariate empirical distribution of X and Y

Covariance The covariance between the two variables is non-negative.

PQD X and Y are PQD (see Section 2.2).

Association (X, Y) are said to be associated if for all non-descreasing functions
$\text{Cov}(\varphi(X, Y), \chi(X, Y)) \geq 0$.

Furthermore, four stronger concepts (left- and right-tail dependence, row- / column-regression dependence, and total dependence of order s) of dependence are reviewed. It is trivial that Association implies PQD, and that PQD implies non-negativity of the covariance. Kotz et al.[14] carried out an extensive computer simulation where they checked these seven concepts 3000 times for 3×3 matrices P, where the $p_{ij} = \text{P}(X = j, Y = i)$ are uniform random. In 16.8% of the times the generated data were PQD, which coincides with the theoretic probability of $\frac{1}{6}$(Kotz et al.[14]). In all the simulations where PQD was obtained, Association was also obtained (and of course the converse holds too). So in practice, it seems that PQD and Association are almost the same and the choice between them does not affect the inferences much.

4.2 Estimation using copulas and t-norms

To find nonparametric methods that take positive association into account, the concepts of *copula* and *t-norm* may be helpful. An introduction into these concepts is given on the basis of Schweizer and Sklar[28]. See also IMS Lecture Notes, Volume 28, 'Distributions with fixed marginals and related topics' ([20], [21], [29]). We restricted the attention to distributions on the unit square because they can be made that way.

21

A function T from $S \times S$ onto S is called a binary operation on S. It is called *associative* if $T(T(x,y),z) = T(x,T(y,z))$ $\forall x, y, z$ in S. We shall restrict ourselves to $S = [0,1)$.

Definition 4.1 (t-norm.) *A triangular norm (or t-norm) is an associative binary operation on $[0,1]$ that satisfies the axioms*

(i) $T(x_1, y_1) \leq T(x_2, y_2)$ $\forall x_1 \leq x_2, y_1 \leq y_2$

(ii) $T(x, 1) = T(1, x) = x$

(iii) $T(x, y) = T(y, x)$

in each point.

A t-norm may be visualized as a surface over the unit square that contains the skew quadrilateral whose vertices are the coordinates $(0,0,0), (1,0,0), (1,1,1)$, and $(0,1,0)$. The term triangular norm originates from this visualization.

Three common examples of t-norms are W, Π and M:

$$
\begin{aligned}
W(x,y) &= \max(x + y - 1, 0) \\
\Pi(x,y) &= xy \\
M(x,y) &= \min(x, y)
\end{aligned}
$$

For the joint distribution function H with marginals F and G, there is a function C from the unit square onto the unit interval such that

$$H(x,y) = C(F(x), G(y)), \quad \forall x, y \tag{4.2}$$

This function is continuous when the marginal distributions of H are continuous. Such a function C is called 2-copula, or 2-dimensional copula, ([28], [29]. Since we are only interested in the bivariate case, we delete the prefix 2-). Copula's are 2-dimensional distribution functions with uniformly distributed marginals, and they are often used in transformation models - just like as in Sections 3.3 and 3.4. We obtain different functions H when we use different functions C, so we can incorporate initial 'knowledge' about the bivariate distribution by choosing C.

It follows that each copula concerned with a distribution with continuous marginals, is uniformly continuous on its domain. It also follows that the t-norms M, Π, and W are copulas, and for any copula C we have $W \leq C \leq M$. That is why the t-norm W (M) is sometimes called the lower (upper) Fréchet bound (Marshall[20] and Nelsen[21]). In general, a t-norm is a copula if and only if it satisfies the Lipschitz condition

$$T(a, y) - T(b, y) \leq a - b, \quad a \leq b \tag{4.3}$$

For the proof we refer to Schweizer and Sklar[28], p. 86.

For continuous marginals, copulas are unique. Marshall[20] states that if H, with discontinuous marginals, is PQD, then among the various copulas of H,

there is at least one that is PQD. He also states that $C(F,G)$ has a nonnegative correlation for all F and G if and only if C is PQD. This follows immediately from Hoeffding's lemma (see also Appendix A. So, when we want to incorporate positive quadrant dependence of our two random variables, we need to take a copula which itself is PQD too. After some basic calculations it can be seen that W is negative quadrant dependent, Π is PQD (but not strict), and M is strict PQD. When the assumption that F and G are PQD is made, $H(x,y) = \min(F(x),G(y))$ is a nonparametric estimation that takes this assumption into account.

4.3 The relationship between copulas and Chapter 3

In Section 3.3 we have used bivariate normal distribution for estimating our bivariate distribution, with $\hat{\rho}$ as the parameter describing dependence. This corresponds to the copula

$$C_\rho(u,v) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}} \cdot e^{\left[-\frac{1}{2(1-\rho^2)}\left\{u^2 - 2\rho uv + v^2\right\}\right]} dvdu \quad (4.4)$$

(see Formula 3.9) where $C_\rho \in \mathcal{C} = \{C_\rho | \rho \in [-1, +1]\}$. In Section 3.4 the BVE-distribution is used, corresponding to the copula

$$\tilde{C}_\rho(u,v) = \int_0^u \int_0^v (1+\hat{\rho})^{-1} \exp\left[-\frac{1-\hat{\rho}}{1+\hat{\rho}}(u+v) - \frac{2\hat{\rho}}{1+\rho}\max(u,v)\right] dudv \quad (4.5)$$

where $\tilde{C}_\rho \in \tilde{\mathcal{C}} = \{\tilde{C}_\rho | \rho \in [-1, +1]\}$. So in both models, there is only one parameter, ρ, on which the estimates depend (therefore these models are called semiparametric models).

As was stated in the previous section, $M((x,y) = \min(F(x),G(y))$ is a bivariate PQD distribution. This also explains our choice for the bivariate exponential distribution of Mashall and Olkin (Formula (3.14) in Section 3.4). Note that the copula $M(x,y) = \min(x,y)$ is similar to the part of Formula (3.14) describing the dependence between X and Y (since (X,Y) is transformed into (e^{-U}, e^{-V}) the maximum must be taken instead of the minimum).

23

4.4 Ordening the data to obtain PQD

Let us now think of the case where a distribution function H on $[0,1] \times [0,1]$ is given, with marginals uniform, but not satisfying the requirement of PQD ($H(x,y) \geq xy, \forall x,y$). The goal is to find a modification J of H such that the marginals remain uniform, but that the PQD-requirement is satisfied. We want to find the J that is 'as close as possible' to H, according to some specified dissimilarity measure.

The data are transformed in some way to a $n \times n$ matrix, say M. For this matrix $\sum_{i=1}^{n} m_{ij} = n$ and $\sum_{j=1}^{n} m_{ij} = n$ ('uniform marginals') hold, but $\sum_{i=1}^{x} \sum_{i=1}^{y} m_{ij} \geq xy, \forall x,y \in \{1,\ldots,n\}$ ('PQD') does not hold. Now we want a modification N of M such that summation over each row and column still gives n, and that $\sum_{i=1}^{x} \sum_{i=1}^{y} n_{ij} \geq xy, \forall x,y \in \{1,\ldots,n\}$. N is made PQD by tranfering some value, say a, from element $n_{i,k}$ to $n_{i,l}$ and transfering a from $n_{j,l}$ to $n_{j,k}$. This process is repeated several times for specific values of a,i,j,k, and l coming from a specified algoritm. Scarsini[26] explains such an algoritm, using the earlier mentioned copula $M(x,y) = \min(x,y)$, but this transformation is not 'minimal'. The goal is to find a transformation N of M such that the difference between N and M is minimal, conditional to N being PQD. After such a matrix N has been composed, it has to be transformed back to a bivariate distribution. Some smoothing methods have to be used, but this brings the complication that the smoothened distribution following from N might not be PQD. Hopefully, we eventually obtain a bivariate distribution function J that satisfies the PQD requirement and is 'close' to our original distribution function H.

At the moment, we can not solve this problem. A good investigation of related literature is needed before we might be able to estimate the bivariate density according to the nonparametric method described above.

Appendix A

Proof of Lemma 2.4

That condition (i) is equivalent to (ii) is obvious, so we only have to proof that (i) (and thus also (ii)) are equivalent to (iii). As already stated in section 2.2, PQD is invariant under increasing transformations. The same holds thus if decreasing functions are applied to both coordinates. Remaining for us to prove is

$$(X, Y) \text{ is PQD} \quad \Leftrightarrow \quad \text{cov}(r(X), s(Y)) \geq 0, \forall r, s \text{ nondecreasing}$$

Since $\text{cov}(X, Y) = \text{E}(XY) - \text{E}X \, \text{E}Y$, stating that $\text{cov}(X, Y) \geq 0$ is equivalent to stating that $\text{E}(XY) \geq \text{E}X \, \text{E}Y$. The proof was found in Lehmann([16], p. 1139-1140), who made use of the following lemma of Hoeffding

Lemma A.1 *If H denotes the joint distribution of X and Y, then*

$$\text{E}(XY) - \text{E}X \, \text{E}Y = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [H(x, y) - H(x, \infty)H(\infty, y)] \, dx dy \qquad (A.1)$$

provided the expectations on the left hand side exist.

Proof. Let (X_1, Y_1), (X_2, Y_2) be independent and each distributed according to H. Then

$$2 [\text{E}(X_1 Y_1) - \text{E}X_1 \text{E}Y_1] = \text{E}[(X_1 - X_2)(Y_1 - Y_2)] =$$

$$\text{E} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[1_{\{u \leq X_1\}} - 1_{\{u \leq x_2\}}\right] \left[1_{\{v \leq Y_1\}} - 1_{\{v \leq Y_2\}}\right] du dv \qquad (A.2)$$

The first step is in the line of analogous steps made in Section 2.3. Since we assume that $\text{E}|XY|, \text{E}|X|, \text{E}|Y|$ are finite, we can take the expectation under the integration sign. After a few simple calculations we obtain twice the right hand side of (A.1), which completes the proof.

That PQD of (X, Y) implies $\text{cov}(X, Y) \geq 0$ follows immediately from this lemma. Suppose now that the covariance is zero, and that (X, Y) is PQD. This means that $H(x, y) = H(x, \infty)H(\infty, y)$ (except possibly on a set of Lebesque measure zero). Cumulative distribution functions are continuous on the right,

25

and this means that if two distributions agree a.e. w.r.t. Lebesgue measure, they must agree everywhere. Thus X and Y must be independent, and this completes the proof.

Bibliography

[1] Atkinson, K.E. (1989), *An introduction to numerical analysis*, (second edition), Wiley, New York.

[2] Basu, A.P. (1990), *A survey of some inference problems for dependent systems*, IMS Lecture Notes, Monograph Series Vol. 16, 35–44.

[3] Bickel, P.J., C.A.J. Klaassen, Y. Ritov, and J.A. Wellner (1993) *Efficient and adaptive estimation for semiparametric models*, John Hopkins Press, Baltimore/London.

[4] Brady, B. and N.D. Singpurwalla (1990), *Stochastically monotone dependence*, IMS Lecture Notes, Monograph Series Vol. 16, 93–102.

[5] Bruin, R. de and W. Schaafsma (1994), *A semi-Bayesian method for nonparametric density estimation*, University of Groningen.

[6] Dehling, H.G. and J.N. Kalma (1995), *Kansrekening, het zekere van het onzekere*, Epsilon Uitgaven, Utrecht.

[7] Fortuin, E., E. Hülsmann and J.K. Ng (1997), *(no title)*, University of Groningen.

[8] Gibbons, J.D. and S. Chakraborti (1992), *Nonparametric statistical inference*, Dekker, New York.

[9] Grimmett, G.R. and D.R. Stirzaker (1992), *Probability and random processes*, (second edition), Oxford University Press, New York.

[10] Gupta, P.L. and R.D. Gupta (1990), *Relative errors in reliable measures*, IMS Lecture Notes, Monograph Series Vol. 16, 251–256.

[11] Höeffding, W. (1948), *A nonparametric test of independence*, Annals of Mathematical Statistics 19, 546–557.

[12] Kendall, M.G. (1975), *Rank correlation methods (fourth edition, second impression)*, Griffin, London.

[13] Kendall, M.G. (1980), *Multivariate analysis (second edition)*, Griffin, London.

[14] Kotz, S., Q. Wang and K. Hung (1990), *Interrelations among various definitions of bivariate positive dependence*, IMS Lecture Notes, Monograph Series Vol. 16, 333–349.

[15] Lehmann, E.L. (1959), *Testing linear hypotheses*, Wiley, New York.

[16] Lehmann, E.L. (1966), *Some concepts of dependence*, Annals of Mathematical Statistics 37, 1137–1153.

[17] Lindeman, R.H., P.F. Merenda and R.Z. Gold (1980), *Introduction to bivariate and multivariate analysis*, Scott, Foresman and Company, Dallas.

[18] Lindgren, B.W. (1993), *Statistical Theory, fourth edition*, Chapman and Hall, New York.

[19] Marshall, A.W. and I. Olkin (1967) *A multivariate exponential distribution*, Journal of the American Statistical Association 62, 30–44.

[20] Marshall, A.W. (1996), *Copulas, marginals, and joint distibution functions*, IMS Lecture Notes, Monograph Series Vol. 28, 213–222.

[21] Nelsen, R.B. (1996), *Nonparametric measures of multivariate association*, IMS Lecture Notes, Monograph Series Vol. 28, 223–232.

[22] Puri, M.D. (1970), *Nonparametric techniques in statistical inference*, Cambridge University Press, Cambridge.

[23] Puri, M.L. and P.K. Sen (1971), *Nonparametric methods in multivariate analysis*, Wiley, New York.

[24] Rao, C.R. (1981), *Multivariate analysis; some reminiscences on its origin and development*, T.V. talk at the University of Connecticut, Storrs, Connecticut.

[25] Ruymgaart, F.H. (1973), *Asymptotic theory of rank tests for independence*, Mathematical Centre, Amsterdam.

[26] Scarsini, M. (1990), *An ordering of dependence*, IMS Lecture Notes, Monograph Series Vol. 16, 403–414.

[27] Schaafsma, W. (1966), *Hypothesis testing problems with the alternative restricted by a number of parameters*, Noordhoff, Groningen.

[28] Schweizer, B. and A. Sklar (1983), *Probabilistic Metric Spaces*, North Holland, New York.

[29] Sklar, A. (1996), *Random variables, distribution functions and copulas – a personal look backward and forward*, IMS Lecture Notes, Monograph Series Vol. 28, 1–14.

[30] Silverman, B.W. (1986), *Density estimation for statistics and data analysis*, Chapman and Hall, London.