Journal of Data Science **12**(2014), 359-375

Nonparametric inference for $P(X < Y)$ with paired variables

José Arturo Montoya[1] and Francisco Javier Rubio[2]
[1] *University of Sonora*
[2] *University of Warwick*

Abstract: We propose two classes of nonparametric point estimators of $\theta = P(X < Y)$ in the case where (X, Y) are paired, possibly dependent, absolutely continuous random variables. The proposed estimators are based on nonparametric estimators of the joint density of (X, Y) and the distribution function of $Z = Y - X$. We explore the use of several density and distribution function estimators and characterise the convergence of the resulting estimators of θ. We consider the use of bootstrap methods to obtain confidence intervals. The performance of these estimators is illustrated using simulated and real data. These examples show that not accounting for pairing and dependence may lead to erroneous conclusions about the relationship between X and Y.

Key words: Bootstrap, dependence, density estimation, distribution estimation, stress-strength model.

1. Introduction

The study of stress–strength models have received considerable attention for many years due to its applicability in diverse areas. The main interest in this kind of models is the quantity $\theta = P(X < Y)$, where X and Y are random variables. In medicine for example, if X and Y are the outcomes of a control and an experimental treatment respectively, the parameter θ can be interpreted as the effectiveness of treatment Y (Ventura et al., 2011). This quantity is also related to the Receiver Operating Characteristic (ROC) curves, where θ is interpreted as an index of accuracy (Zhou, 2008). In engineering and reliability studies θ is also a quantity of interest because it may represent the probability that the strength of a component (Y) exceeds the stress (X) coming from external factors (Kotz et al., 2003).

Stress-strength models were introduced by Birnbaum (1956) who proposed a nonparametric estimator of θ based on the Mann-Whitney statistic for the case where X and Y are independent. There is a large amount of literature related

to the study of point and interval estimation of θ using different approaches (see Kotz et al., 2003 for a good survey on this). For instance, in the case where X and Y are independent, Sun et al. (1998) proposes a Bayesian approach using reference priors; Baklizi and Eidous (2006) propose an estimator based on kernel estimators of the densities of X and Y (which can be straightforwardly generalised to the use of other nonparametric density estimators); Zhou (2008) proposes the use of bootstrap and asymptotic intervals; Jing et al. (2009) estimate θ using the empirical likelihood; Montoya (2008) and Díaz-Francśs and Montoya (2013) propose the use of the profile likelihood for conducting inference about θ; and Ventura et al. (2011) propose the use of Bayesian inference with Jeffreys and matching priors as well as modified profile likelihoods for the cases where X and Y are normal or exponential random variables.

It is important to mention that the parameter θ may not be available in a closed form in many cases (see Azzalini and Chiogna, 2004 and Gupta and Brown, 2001 for an example of this). This makes difficult (if at all feasible) to find a reparameterisation involving θ, which complicates the use of the classical approach. In particular, the use of the profile likelihood might be difficult if this reparameterisation is not available (Díaz-Francś and Montoya, 2013). Alternative inferential approaches that overcome this difficulty are Bayesian inference, nonparametric estimation, and the use of bootstrap methods, which allow for obtaining confidence and credible intervals for the parameter of interest (Baklizi and Eidous, 2006; Zhou, 2008; Rubio and Steel, 2013).

New interest has been focused on the estimation of θ in the case where X and Y are dependent random variables. For example Barbiero (2011) assumes that (X, Y) are jointly normally distributed; Rubio and Steel (2013) suppose that X and Y are marginally distributed as skewed scale mixture of normals and construct the corresponding joint distribution using a Gaussian Copula; Domma and Giordano (2012a) construct the joint distribution of (X, Y) using a Farlie-Gumbel-Morgenstern copula with marginal distributions belonging to the Burr system; Domma and Giordano (2012b) consider Dagum distributed marginals and construct their joint distribution using a Frank copula; among others (Nadarajah, 2005; Gupta et al., 2012). In these papers, the importance of taking the assumption of dependence between X and Y into consideration is illustrated using simulated and real data sets.

We propose two classes of nonparametric estimators of θ for the case where (X, Y) are paired, possibly dependent, continuous random variables. This scenario is of interest since paired observations are produced in many experimental designs (see e.g. Sprott, 2000 and Cox and Reid, 2000 for examples of this). The estimators proposed here are based on nonparametric estimators of the density of (X, Y) and the distribution function of $Z = Y - X$. This approach avoids making

distributional assumptions over (X, Y) and allows for interval estimation of θ via nonparametric bootstrap. In addition, this method can be easily implemented in R using already existing packages. In Section 2 we introduce these estimators and prove some asymptotic properties for the choice of several nonparametric estimators. We also detail how to combine kernel density estimation (KDE) with the methods proposed here. In Section 3 we present two examples, using simulated and real data, which illustrate the importance of accounting for pairing and dependence of the observations when conducting inference about θ.

2. Nonparametric estimators of θ

Let (X, Y) be a pair of absolutely continuous random variables with joint density $f_{X,Y} : \mathbb{R}^2 \to \mathbb{R}_+$. By definition, we have that

$$\theta = \mathbb{P}(X < Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{y} f_{X,Y}(x, y) dx dy. \tag{1}$$

Alternatively, by defining the variable $Z = Y - X$ we obtain

$$\theta = \mathbb{P}(Z > 0) = \int_{0}^{\infty} f_Z(z) dz = 1 - F_Z(0) = S_Z(0), \tag{2}$$

where f_Z, F_Z and S_Z are the density function, the cumulative distribution function, and the survival function of Z, respectively. These equivalent expressions suggest the following nonparametric methods for estimating the parameter θ.

2.1 Estimator I

Let $(\mathbf{x}, \mathbf{y})$ be a sample from (X, Y) of size n and suppose that these observations are collected in couples (x_i, y_i), $i = 1, \ldots, n$. The first proposed estimator, based on expression (1), consists of substituting the density $f_{X,Y}$ by a nonparametric density estimator as follows.

Algorithm 1
1:Using the sample $(\mathbf{x}, \mathbf{y})$ construct a nonparametric estimator $\hat{f}_{X,Y}$ of the density $f_{X,Y}$.
2:Define the estimator $\tilde{\theta} = \int_{-\infty}^{\infty} \int_{-\infty}^{y} \hat{f}_{X,Y}(x, y) dx dy$.

Note that Algorithm 1 involves both a two-dimensional density estimation and the calculation of a double integral. Several nonparametric density estimators can be employed for this purpose such as kernel density estimators (Parzen, 1962),

shape-restricted estimators (Cule et al., 2010), among others (Scott, 1992). This choice has, of course, implications on the performance of the estimators. In Section 2.4 we present some asymptotic properties of $\tilde{\theta}$ for different choices of $\hat{f}_{X,Y}$. The integration step can be conducted using quadrature or Monte Carlo methods.

2.2 Estimator II

Again, let $(\mathbf{x}, \mathbf{y})$ be a sample from (X, Y) of size n and suppose that these observations are collected in couples (x_i, y_i), $i = 1, \ldots, n$. Define the vector of differences $\mathbf{z} = \mathbf{y} - \mathbf{x}$. The second proposed estimator, based on expression (2), is constructed as follows.

Algorithm 2
1:Calculate the differences $\mathbf{z} = \mathbf{y} - \mathbf{x}$.
2:Using the sample $\mathbf{z}$ construct a nonparametric estimator $\hat{F}_Z$ of the distribution function of Z .
3:Define the estimator $\hat{\theta} = 1 - \hat{F}_Z(0)$.

For the nonparametric distribution estimator $\hat{F}_Z$ in step 2 we can employ the empirical cumulative distribution function (ECDF) or the induced distribution estimators obtained by integrating a nonparametric density estimator $\hat{f}_Z$, which lead to $\hat{\theta} = \int_0^\infty \hat{f}_Z(z)dz$. In this line, several univariate nonparametric estimators of $\hat{f}_Z$ can be considered such as kernel density estimators (Parzen, 1962), shape-restricted density estimators (Cule et al., 2010) and smooth shape-restricted estimators (DLumbeng and Rufibach, 2009; Cule et al., 2010; DLumbeng and Rufibach, 2011).

Note that the use of both, Estimator I and Estimator II, avoids making assumptions on the distribution of (X, Y) and the sort of dependence between the variables X and Y. The relationship between these variables, which can be either dependent or independent, is implicitly included in the nonparametric estimators of the density (distribution). In addition, the use of nonparametric bootstrap coupled with either Algorithm 1 or Algorithm 2 allows for obtaining a variety of bootstrap confidence intervals for these estimators (DiCiccio and Efron, 1996).

2.3 Use of Estimator I and Estimator II with kernel density estimators

In this section we discuss the use of KDE in Algorithms I and II. Recall that the use of KDE involve the choice of two elements: a kernel function and a bandwidth parameter (or bandwidth matrix in a multivariate framework). Here,

we present a brief discussion on appropriate choices for these elements in our context.

2.3.1 Estimator I

Let H be a symmetric, positive definite, 2×2 bandwidth matrix and k_2 be a two-dimensional kernel function (Parzen, 1962). Define also $k_H(\mathbf{t}) = (\det H)^{-\frac{1}{2}} k_2 (H^{-\frac{1}{2}}\mathbf{t})$, $\mathbf{t} \in \mathbb{R}^2$. If we consider the use of a KDE in step 1 of Algorithm 1, then the estimator $\tilde{\theta}$ can be written as

$$\tilde{\theta} = \frac{1}{n} \sum_{j=1}^{n} \int_{-\infty}^{\infty} \int_{-\infty}^{y} k_H(x - x_j, y - y_j) dx dy, \tag{3}$$

which can be calculated using quadrature or Monte Carlo methods. The implementation of this estimator requires the specification of the kernel function k_2 and the bandwidth matrix H. A natural first choice is the use of a bivariate Gaussian kernel $\phi_2 = k_2$. The choice of the bandwidth matrix H can be crucial for the performance of KDE, which has fostered an extensive study of several bandwidth matrix estimators (see Duong and Hazelton, 2005 for a good survey on this). However, appropriate bandwidth matrices for estimating the distribution involved in (3) seem to have been little studied to our knowledge. Nevertheless, as a first approach one can consider bandwidth matrix estimators employed in KDE such as the plug-in and cross-validation bandwidth estimators, which are implemented in the R package 'ks' (Duong, 2011).

2.3.2 Estimator II

Let k_1 be a one–dimensional kernel and $h > 0$ be the corresponding bandwidth (also termed *smoothing parameter*). If we consider the use of a univariate KDE in step 2 of Algorithm 2, then the estimator $\hat{\theta}$ can be written as

$$\hat{\theta} = \frac{1}{nh} \sum_{j=1}^{n} \int_{0}^{\infty} k_1 \left(\frac{z - z_j}{h} \right) dz = 1 - \frac{1}{n} \sum_{j=1}^{n} K_1 \left(\frac{z_j}{h} \right), \tag{4}$$

where $K_1 \left(\frac{z}{h} \right) = \frac{1}{h} \int_0^\infty k_1 \left(\frac{z}{h} \right) dz$. Again, a natural first choice is the Gaussian kernel $\Phi = K_1$. The choice of the bandwidth h in the context of density estimation has been extensively studied, we refer the reader to Jones et al. (1996) for a good survey on this. However, the choice of this parameter in the context of

kernel distribution function estimation has received less attention. Quintela-del-Río and Estévez-Pérez (2012) present a compendium of appropriate bandwidth parameters in the context of kernel distribution estimator. They also implement these in the R package 'kerdiest'.

2.4 Results on the convergence of the proposed estimators

The convergence of Estimator I coupled with KDE is difficult to assess given the limited literature about the choice of appropriate bandwidth matrices for estimating a bivariate distribution. Despite this limitation, one can expect a good performance of this estimator for moderate or large samples and the use of any reasonable bandwidth matrix since kernel estimators converge in terms of the mean square and mean absolute errors to the true density. The following result shows that, even using a diagonal bandwidth matrix, the resulting estimator of θ is weakly consistent under rather mild conditions. The use of more appropriate bandwidth matrices is therefore expected to produce better estimators.

Theorem 1 Suppose that k_2 is bounded on $\mathbb{R}^2$ with

$$L(u) = \sup_{||\mathbf{t}|| \geq u} k_2(\mathbf{t}),$$

for $u \geq 0$. Let $\{h_n\}_{n=1}^{\infty}$ be a sequence of positive numbers such that $\lim_{n\to\infty} h_n = 0$ and $\lim_{n\to\infty} nh_n^2 = \infty$. Define the sequence of bandwidth matrices $H_n = \text{diag}(h_n)$. Suppose also that one of the following conditions holds

(i) $||\mathbf{t}||^2 k_2(\mathbf{t}) \to 0$ as $||\mathbf{t}|| \to \infty$ and $f_{X,Y}$ is almost surely continuous.

(ii) $f_{X,Y}$ is bounded.

(iii) $\int_0^{\infty} uL(u)du < \infty$.

Then, $\tilde{\theta}$ is a weakly consistent estimator of θ, this is, $\tilde{\theta} \xrightarrow{\mathbb{P}} \theta$, as $n \to \infty$.

proof. First, note that

$$\begin{aligned}
|\hat{\theta} - \theta| &= \left| \int_{-\infty}^{\infty} \int_{-\infty}^{y} [\hat{f}_{X,Y}(s,t) - f_{X,Y}(s,t)] dsdt \right| \\
&\leq \int_{-\infty}^{\infty} \int_{-\infty}^{y} \left| \hat{f}_{X,Y}(s,t) - f_{X,Y}(s,t) \right| dsdt \\
&\leq \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left| \hat{f}_{X,Y}(s,t) - f_{X,Y}(s,t) \right| dsdt \\
&\leq \text{MAE}(\hat{f}_{X,Y}, f_{X,Y}),
\end{aligned}$$

where MAE denotes the *mean absolute error* which is also the L_1 distance. Under the stated assumptions we have that $\lim_{n\to\infty} \mathrm{MAE}(\hat{f}_{X,Y}, f_{X,Y}) = 0$, in probability, by the Theorem in Devroye and Wagner (1979).

Although the use of the shape-restricted density estimator in Cule et al. (2010) does not involve a tuning parameter, a study of the asymptotic properties of the induced distribution estimator seems not to have been done yet. However, since this density estimator has smaller mean integrated squared error than those obtained with KDE methods (Cule et al., 2010), the use of this method in Algorithm 1 is also expected to produce good estimators of θ for moderate or large samples.

On the other hand, given the immediate relationship between the Estimator II and the estimation of the distribution F_Z, it follows that the asymptotic properties of $\hat{\theta}$ are inherited from those of the estimator $\hat{F}_Z$ evaluated at 0. Some specific asymptotic results are presented below for different estimators of $\hat{F}_Z(0)$.

The following result shows that the use of the empirical distribution for estimating $\hat{F}_Z(0)$ produces consistent and asymptotically normal estimators of θ.

Theorem 2 Let $\hat{F}_Z$ be the empirical distribution function, then

(i) $\hat{\theta}$ is strongly consistent, this is, $\hat{\theta} \overset{a.s.}{\to} \theta$, as $n \to \infty$.

(ii) The estimator $\hat{\theta}$ is asymptotically normal, this is

$$\sqrt{n}\left(\hat{\theta} - \theta\right) \overset{d}{\to} N\left(0, F_Z(0)(1 - F_Z(0))\right),$$

as $n \to \infty$.

proof. The results follow by the law of large numbers and the central limit theorem (van der Vaart, 1998).

The use of kernel distribution estimators can also produce consistent and asymptotically normal estimators of θ under certain conditions as indicated in the following theorem.

Theorem 3 Assume that F_Z is uniformly Lipschitz on $\mathbb{R}$ and let $\hat{F}_Z$ be a *regular* kernel estimator. This is, there exists a positive sequence $\{h_n\}_{n=1}^{\infty}$ such that $h_n = o\left(n^{-\frac{1}{2}}\right)$ and

$$\int_{|t|>h_n} \frac{1}{h_n} k_1\left(\frac{t}{h_n}\right) dt = o\left(n^{-\frac{1}{2}}\right).$$

Then, it follows that

1. $\hat{\theta}$ is a strongly consistent estimator of θ.

2. $\hat{\theta}$ is asymptotically normal, $\sqrt{n}\left(\hat{\theta}-\theta\right) \xrightarrow{d} N\left(0, F_Z(0)(1-F_Z(0))\right)$.

proof. (i) Using the triangle inequality it follows that $|\hat{F}_Z(0)-F_Z(0)| \leq |\hat{F}_Z(0)-F_n(0)|+|\hat{F}_n(0)-F_Z(0)|$, where F_n is the empirical distribution function. Then, the result follows by Theorem 2.3 from Fernholz (1991) and the law of large numbers. (ii) The asymptotic normallity of $\hat{\theta}$ follows by Corollary 2.4 from Fernholz (1991).

By relaxing the assumptions of the previous theorem it is possible to prove that the use of kernel distribution estimators also produces weakly consistent estimators of θ.

Theorem 4 Suppose that k_1 is bounded in $\mathbb{R}$ with

$$L(u) = \sup_{|t|\geq u} k_1(t),$$

for $u \geq 0$. Let $\{h_n\}_{n=1}^{\infty}$ be a sequence of positive bandwidths such that $\lim_{n\to\infty} h_n = 0$ and $\lim_{n\to\infty} nh_n = \infty$. Suppose also that one of the following conditions holds

(i)$|t|k_1(t) \to 0$ as $|t| \to \infty$ and f_Z is almost surely continuous.

(ii)f_Z is bounded.

(iii)$\int_0^{\infty} L(u)du < \infty$.

Then, $\hat{\theta}$ is a weakly consistent estimator of θ.

proof. First, note that

$$\begin{aligned}
|\hat{\theta}-\theta| &= \left|\hat{F}_Z(0)-F_Z(0)\right| = \left|\int_{-\infty}^{0}[\hat{f}_Z(t)-f_Z(t)]dt\right| \\
&\leq \int_{-\infty}^{0}\left|\hat{f}_Z(t)-f_Z(t)\right|dt \\
&\leq \int_{-\infty}^{\infty}\left|\hat{f}_Z(t)-f_Z(t)\right|dt \\
&\leq \mathrm{MAE}(\hat{f}_Z, f_Z),
\end{aligned}$$

where MAE denotes the *mean absolute error*. Under the stated assumptions we have that $\lim_{n\to\infty} \mathrm{MAE}(\hat{f}_Z, f_Z) = 0$, in probability, by the Theorem in Devroye and Wagner (1979).

Note that Theorems 1 and 4 simply require a well-behaved kernel function and the boundedness of the target density. The assumptions on the bandwidth parameters are also rather mild since most of the popular choices satisfy these conditions.

The use of shape-restricted estimators, by their nature itself, require additional assumptions on the target density. The following result presents such conditions that produce consistent estimators of θ.

Theorem 5 Let $\hat{F}_Z$ be the shape-restricted nonparametric estimator of F_Z proposed in Dümbeng and Rufibach (2009) and suppose that the log-density $\log(f_Z)$ is Lipschitz continuous and $\log(f_Z)'$ is Hölder continuous of order $\beta \in [1, 2]$ on a compact interval $I \subset \mathbb{R}$. Then, $\hat{\theta}$ is a weakly consistent estimator of θ.

proof. The result is a consequence of Corollary 4.2 from Dümbeng and Rufibach (2009).

The results presented in this section show that both estimators have good asymptotic performance under mild conditions. An important difference between Estimator I and Estimator II is that the former involves a two-dimensional density (distribution) estimation while the latter involves a one-dimensional density (distribution) estimation. This represents an advantage of Estimator II over Estimator I since the convergence rate of the resulting estimators as well as the ease of implementation are tied to the dimensionality of the problem. However, an interesting feature of Estimator I is that it can be implemented in the context of censored and missing observations since the use of KDE in these contexts has been studied, for example, in Titterington and Mill (1983) and Wells and Yeo (1996).

3 Examples

In this section, we present two examples that illustrate the implementation of the estimators proposed in Section 2. In the first example we use a sample simulated from a bivariate sinh-arcsinh distribution (Jones and Pewsey, 2009). As detailed in Jones and Pewsey (2009), this distribution contains parameters that control skewness, kurtosis and correlation of the marginals. This example illustrates the influence of the assumptions of pairing and dependence on the bootstrap distributions of the corresponding estimators in terms of their location and spread. In the second example we use a real data set and show that not including the assumptions of pairing and dependence may lead to opposite conclusions about the relationship between X and Y. In both examples, we consider the following 9 types of estimators of θ:

(i) **Kernel 2D.** Based on Algorithm 1, this estimator employs a two-dimensional Gaussian kernel density estimator with the bandwidth matrix *Hscv* implemented in the R package 'ks' (Duong, 2011). The required integration step is conducted using quadrature methods.

(ii) **MLE 2D.** This estimator is based on Algorithm 1. The corresponding

two-dimensional density estimation is conducted using the shape-restricted estimator from Cule et al. (2010). This estimator is also implemented using the command *dlcd* from the R package 'LogconcDEAD' (Cule et al., 2009). The integration of this density is conducted using a Monte Carlo method.

(iii) **SMLE 2D.** This estimator is based on Algorithm 1. The corresponding two-dimensional density estimation is conducted using the smooth shape-restricted estimator (Cule et al., 2010) implemented in the command *dslcd* from the R package 'LogconcDEAD' (Cule et al., 2009). The integration of this density is conducted using a Monte Carlo method.

(iv) **MLE 1D.** Estimates $\hat{F}_Z$ in Algorithm 2 by integrating the shape-restricted density estimator proposed in Cule et al. (2010). The density estimation is implemented using the command *dlcd* from the R package 'LogconcDEAD' (Cule et al., 2009).

(v) **Kernel 1D.** Based on Algorithm 2, this estimator employs a Gaussian kernel distribution estimator with the bandwidth *ALbw* implemented in the R package 'kerdiest' (Quinteladel- Río and Estévez-Perez, 2012).

(vi) **SMLE 1D.** Estimates $\hat{F}_Z$ in Algorithm 2 by integrating a smoothed shape-restricted density estimator (Cule et al., 2010) implemented in the command *dslcd* from the R package 'LogconcDEAD' (Cule et al., 2009).

(vii) **ECDF.** This estimator employs the empirical distribution function for estimating $\hat{F}_Z$ in Algorithm 2.
In order to assess the impact of the assumptions of pairing and dependence in the estimation of θ, we also consider the following estimators:

(viii) **Independent.** (Baklizi and Eidous, 2006) This estimator assumes that X and Y are independent variables and that the corresponding samples are unpaired. The estimator is defined as

$$\theta^\star = \int_{-\infty}^{\infty} \int_{-\infty}^{y} \hat{f}_X(x)\hat{f}_Y(y)dxdy, \tag{5}$$

where $\hat{f}_X$ and $\hat{f}_Y$ are Gaussian kernel density estimators obtained with samples of X and Y respectively. For both KDE we employ the bandwidth $h = \left(\frac{4\hat{\sigma}^5}{3n}\right)^{\frac{1}{5}}$, where $\hat{\sigma}$ is the sample standard deviation and n is the sample size. This bandwidth is known as the Silverman's rule of thumb.

(ix) **Paired.** (Baklizi and Eidous, 2006). This estimator is the same as (5) but assuming that the samples of X and Y are paired. This additional assumption is taken into consideration in the bootstrap methods used to calculate confidence intervals for $\theta^\star$.

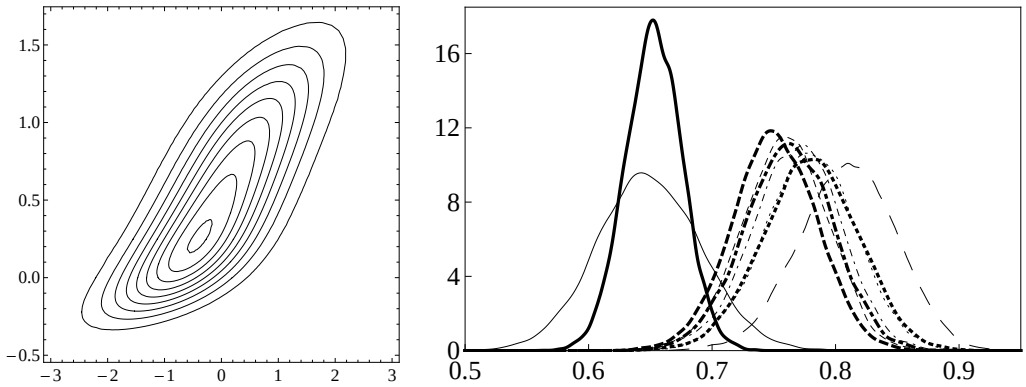
Bootstrap samples and bootstrap confidence intervals (Normal, Basic, Percentile and BCa) are obtained using the R packages 'boot' (Canty and Ripley, 2012) and 'simpleboot' (Peng, 2008). R source code for these examples is available upon request.

3.1 Simulated data

In this example we use a simulated sample of size $n = 100$ from a bivariate sinh-arcsinh distribution (Jones and Pewsey, 2009) with parameters $(\sigma_1, \sigma_2, \rho, \epsilon_1, \epsilon_2, \delta_1, \delta_2) = (1, 1, 0.75, 0, 1, 1, 2)$. Figure 1a shows a contour plot of the corresponding density. This is a complex scenario where the entries present departure from normality and dependence. The population correlation coefficient of this sample is 0.737 and the theoretical correlation is 0.743. The parameter θ in this family of distributions is not generally tractable. The theoretical value of θ, obtained by numerical integration, is 0.78. Figure 1b shows the bootstrap distribution of the estimators of θ previously described. We can observe a considerable influence of the assumptions of pairing and dependence in the location and spread of the bootstrap distributions of the estimators of θ. We can also notice the influence of these assumptions in the point estimators and bootstrap confidence intervals shown in Table 1. In this case, not including these assumptions leads to underestimating θ. Finally, we can observe that the estimator ECDF is slightly larger than the others, which seems to be a result of its discrete nature.

Table 1: Simulated data: Estimators and 95% bootstrap confidence intervals.

Estimator	$\hat{\theta}$	Normal	Basic	Percentile	BCa
Independent	0.65	$(0.560, 0.724)$	$(0.559, 0.723)$	$(0.568, 0.732)$	$(0.562, 0.727)$
Paired	0.65	$(0.606, 0.695)$	$(0.606, 0.696)$	$(0.607, 0.697)$	$(0.604, 0.694)$
ECDF	0.81	$(0.734, 0.886)$	$(0.740, 0.890)$	$(0.730, 0.880)$	$(0.720, 0.870)$
Kernel 1D	0.77	$(0.695, 0.838)$	$(0.697, 0.840)$	$(0.701, 0.843)$	$(0.688, 0.833)$
Kernel 2D	0.75	$(0.674, 0.807)$	$(0.674, 0.807)$	$(0.683, 0.816)$	$(0.668, 0.804)$
MLE 1D	0.78	$(0.707, 0.853)$	$(0.709, 0.854)$	$(0.705, 0.850)$	$(0.701, 0.847)$
MLE 2D	0.77	$(0.695, 0.845)$	$(0.697, 0.846)$	$(0.703, 0.853)$	$(0.685, 0.840)$
SMLE 1D	0.77	$(0.704, 0.844)$	$(0.707, 0.847)$	$(0.694, 0.835)$	$(0.694, 0.835)$
SMLE 2D	0.78	$(0.690, 0.830)$	$(0.692, 0.832)$	$(0.690, 0.830)$	$(0.682, 0.824)$



(a) (b)

Figure 1: (a) Contour plot: sinh-arcsinh distribution; (b) Simulated data: bootstrap distributions of the estimators. Independent (solid line), Paired (solid bold line), ECDF (long-dashed line), Kernel 1D (dashed line), Kernel 2D (dashed bold line), MLE 1D (dotted line), MLE 2D (dotted bold line), SMLE 1D (dotted-dashed line), SMLE 2D (dotted-dashed bold line).

3.2 Real data

In this section we study the data set presented in Venkatraman and Begg (1996), which contains 72 lesion scores obtained using both a clinical scheme without a dermoscope (X Test), and a dermoscopic scoring scheme (Y Test). Their main interest is to assess the information provided by the use of the dermoscope. Here, we analyse the subset of 51 non-diseased patients (diagnosed using a biopsy) and compare the nonparametric inferences for θ obtained using the estimators described in the introduction of this section. It is important to note that the population correlation coefficient of this sample is 0.794, which suggests that the entries are correlated. Table 2 shows point estimators and four types of bootstrap confidence intervals of θ. Figure 2 shows the bootstrap distributions of the estimators of θ. We can note a discrepancy of the point estimators under the assumptions of dependence and independence of the tests. Interval inference is also different; in the cases where pairing and dependence are not considered we can observe that the value $\theta = 0.5$ is included in some of the bootstrap confidence intervals, leading to different conclusions about the relationship of the tests. This is in line with the conclusions in Rubio and Steel (2013) and emphasises the importance of the dependence and pairing assumptions.

Table 2: Melanoma data: Estimators and 95% bootstrap confidence intervals.

Estimator	$\hat{\theta}$	Normal	Basic	Percentile	BCa
Independent	0.55	$(0.469, 0.678)$	$(0.467, 0.672)$	$(0.450, 0.656)$	$(0.474, 0.691)$
Paired	0.55	$(0.498, 0.597)$	$(0.497, 0.596)$	$(0.501, 0.601)$	$(0.499, 0.598)$
ECDF	0.69	$(0.559, 0.813)$	$(0.569, 0.823)$	$(0.549, 0.804)$	$(0.529, 0.784)$
Kernel 1D	0.64	$(0.525, 0.737)$	$(0.525, 0.738)$	$(0.528, 0.741)$	$(0.519, 0.732)$
Kernel 2D	0.62	$(0.514, 0.720)$	$(0.511, 0.719)$	$(0.526, 0.733)$	$(0.512, 0.718)$
MLE 1D	0.65	$(0.543, 0.776)$	$(0.544, 0.776)$	$(0.532, 0.765)$	$(0.537, 0.768)$
MLE 2D	0.65	$(0.523, 0.769)$	$(0.524, 0.772)$	$(0.524, 0.772)$	$(0.513, 0.762)$
SMLE 1D	0.64	$(0.538, 0.756)$	$(0.539, 0.757)$	$(0.527, 0.744)$	$(0.533, 0.749)$
SMLE 2D	0.63	$(0.519, 0.746)$	$(0.523, 0.748)$	$(0.511, 0.736)$	$(0.512, 0.737)$

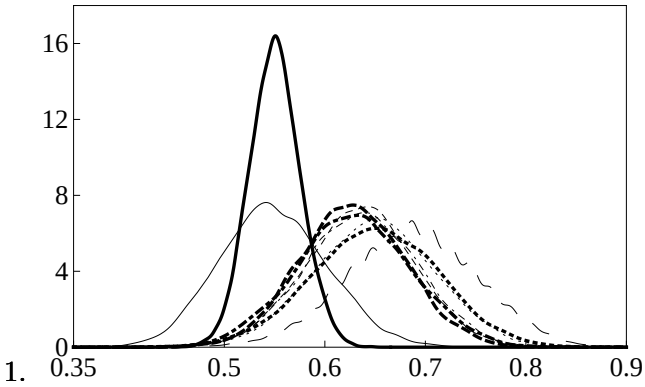


Figure 2: Melanoma data: bootstrap distributions of the estimators. Independent (solid line), Paired (solid bold line), ECDF (long-dashed line), Kernel 1D (dashed line), Kernel 2D (dashed bold line), MLE 1D (dotted line), MLE 2D (dotted bold line), SMLE 1D (dotted-dashed line), SMLE 2D (dotted-dashed bold line).

4 Discussion

We introduced two classes of nonparametric estimators of $\theta = P(X < Y)$ for the case of paired, possibly dependent, observations. The proposed estimators avoid making assumptions on the distribution and the dependence structure of (X, Y) which are implicitly considered by nonparametrically estimating either the joint distribution of (X, Y) or the distribution of the difference $Z = Y - X$. We proved that the combination of the proposed approach with several nonparametric distribution estimators produces estimators of θ with appealing asymptotic properties. In addition, we have shown that confidence intervals for θ, based on these estimators, can be obtained using bootstrap methods that are easy to

implement using already existing R packages. The nonparametric distribution estimators explored in the context of Estimator I perform similarly. They are also comparable in terms of their ease of implementation and the required CPU usage. In the context of Estimator II, we empirically found that the estimators of θ based on smooth distribution estimators exhibit a slightly better performance than those based on discrete distribution estimators such as the empirical distribution. The example presented in Section 3.2 show that not accounting for dependence between X and Y may lead to opposite conclusions about $\theta = 0.5$, and consequently about the relationship between these variables.

References

Azzalini, A. and Chiogna, M. (2004). Some results on the stress-strength model for skew normal variates. *Metron* **LXII**, 315-326.

Baklizi, A. and Eidous, O. (2006). Nonparametric estimation of $P(X < Y)$ using kernel methods. *Metron* **LXIV**, 47-60.

Barbiero, A. (2011). Interval estimators for reliability: the bivariate normal case. *Journal of Applied Statistics* **39**, 501-512.

Birnbaum, Z. M. (1956). On a use of the Mann–Whitney statistic, *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability* **Volume 1: Contributions to the Theory of Statistics** & University of California Press, Berkeley. 13-17 , &University of California Press, Berkeley.

Canty, A. and Ripley, B. D. (2012). *boot: Bootstrap R (S–Plus) Functions.* **R package version 1.3–5**.

Cox, D. R. and Reid, N. (2000). *The Theory of the Design of Experiments* . Chapman & Hall/CRC, Boca Raton, FL.

Cule, M., Gramacy, Robert and Samworth, R. (2009). LogConcDEAD: an R package for maximum likelihood estimation of a multivariate log-concave density. *Journal of Statistical Software* **29(2)**,
URL http://www.jstatsoft.org/v29/i02/.

Cule, M. L., Samworth, R. J. and Stewart, M. I. (2010), Maximum likelihood estimation of a multi-dimensional log-concave density. *Journal Royal Statistical Society* **B 72**, 545–600.

Devroye, L. P. and Wagner, T. J. (1979). The L_1 convergence of kernel density estimates. *Annals of Statistics* **7**, 1136-1139.

Díaz–Francés, E. and Montoya J. A. (2013). The simplicity of likelihood based inferences for $P(X < Y)$ and for the ratio of means in the exponential model. *Statistical Papers* **54**, 499-522.

DiCiccio, T. J. and Efron, B. (1996). Bootstrap confidence intervals. *Statistical Science* **3**, 189-228.

Domma, F. and Giordano, S. (2012a). A copula–based approach to account for dependence in stress–strength models. *Statistical Papers*,forthcoming.

Dümbgen, L. and Rufibach, K. (2011). logcondens: Computations Related to Univariate Log–Concave Density Estimation.*Journal of Statistical Software* **39**, 1–28. URL http://www.jstatsoft.org/v39/i06/.

Dümbgen, L. and Rufibach, K. (2009). Maximum likelihood estimation of a log-concave density and its distribution function: Basic properties and uniform consistency. *Bernoulli* **15**, 40-68.

Duong, T. (2011). ks: Kernel smoothing. **R package version 1.8.5.** http://CRAN.R-project.org/package=ks.

Duong, T. and Hazelton, L. H. (2005). Cross-validation Bandwidth Matrices for Multivariate Kernel Density Estimation. *Scandinavian Journal of Statistics* **32**, 485-506.

Fernholz, L. T. (1991). Almost Sure Convergence of Smoothed Empirical Distribution Functions. *Scandinavian Journal of Statistics* **18**, 225-262.

Gupta, R. C. and Brown, N. (2001). Reliability studies of the skew–normal distribution and its application to a strength–stress model. *Communications in Statistics: Theory and Methods* **30**, 2427-2445.

Gupta R. C., Ghitany, M. E. and Al–Mutairi, D. K. (2012). Estimation of reliability from a bivariate log–normal data. *Journal of Statistical Computation and Simulation*, forthcoming.

Jing, B. Y., Yuan, J. and Zhou, W. (2009). Jackknife empirical likelihood. *Journal of the American Statistical Association* **104**, 1224-1232.

Jones, M. C., Marron, J. and Sheather, S. (1996). A Brief Survey of Bandwidth Selection for Density Estimation. *Journal of the American Statistical Association* **91**, 401-407.estimation. Comput. Statist. 91, 337-381.

Jones, M. C. and Pewsey A. (2009). Sinh–arcsinh distributions. *Biometrika* **96**, 761-780.

Kotz, S., Lumelskii, S. and Pensky, M. (2003). *The Stress–Strength Model and its Generalizations.* Theory and Applications. Singapore: World Scientific.

Montoya, J. A. (2008). *La verosimilitud perfil en la Inferencia Estadística.* PhD Thesis, Centro de Investigación en Matemáticas A. C., México.

Nadarajah, S. (2005). Reliability for some bivariate beta distributions. *Mathematical Problems in Engineering* **2**, 101-111.

Parzen, E. (1962). On estimation of a probability density function and mode. *Annals of Mathematical Statistics* **33**, 1065-1076.

Peng, R. D. *simpleboot: Simple Bootstrap Routines* **R package version 1.1–3**.

Quintela-del-Río, A. and Estéevez-Pérez, G. (2012). Nonparametric Kernel Distribution Function Estimation with kerdiest: An R Package for Bandwidth Choice and Applications. *Journal of Statistical Software* **50**, 1-21.

Rubio, F. J. and Steel, M. F. J. (2013). Bayesian inference for $P(X < Y)$ using asymmetric dependent distributions. *Bayesian Analysis* **8**, 43-62.

Scott, D.W. (1992). *Multivariate Density Estimation: Theory, Practice, and Visualization.* John Wiley, New York.

Sprott, D. A. (2000). *Statistical Inference in Science.* Springer, New York.

Sun, D., Ghosh, M. and Basu, A. P. (1998). Bayesian analysis for a stress–strength system under noninformative priors. *The Canadian Journal of Statistics* **45**, 258-266.

van der Vaart, A. W. (1998). *Asymptotic statistics.* Cambridge University Press, UK.

Venkatraman, E. S. and Begg, C. B. (1996). A distribution–free procedure for comparing operating characteristic curves from a paired experiment. *Biometrika* **83**, 835-848.

Ventura, L. and Racugno, W. (2011). Recent advances on Bayesian inference for $P(X < Y)$. *Journal of the American Statistical Association* **436**, 1566-1574.

Zhou, W. (2008). Statistical inference for $P(X < Y)$. *Statistics in Medicine* **27**, 257-279.

Received June 9, 2012; accepted September 16, 2012.

J. A. Montoya
University of Sonora, México
Blvd. Luis Encinas y Rosales S/N
Col. Centro, Hermosillo, Sonora, Mexico.

F. J. Rubio
University of Warwick
Coventry CV4 7AL, UK
Francisco.Rubio@warwick.ac.uk