Bayesian Inference for $\mathbb{P}(X < Y)$ Using Asymmetric Dependent Distributions

F. J. Rubio[*] and M. F. J. Steel[†]

November 30, 2011

Abstract

This paper studies the Bayesian inference for $\theta = \mathbb{P}(X < Y)$ in the case where the marginal distributions of X and Y belong to classes of distributions obtained by skewing scale mixtures of normals. We separately address the cases where X and Y are independent or dependent random variables. Dependencies between X and Y are modelled using a Gaussian copula. Noninformative benchmark priors are provided for both scenarios and conditions for the existence of the posterior distribution of θ are presented. We show that the use of the Bayesian models proposed here is also valid in the presence of set observations. Examples using simulated and real data sets are presented.

Key Words: Bayesian inference; Gaussian copula; posterior existence; set observation; skewness; stress-strength model

1 Introduction

Stress–strength models have attracted the attention of statisticians for many years due to their applicability in diverse areas such as medicine, engineering, quality control, among others. For example, if X and Y are the outcomes of a treatment and a control group, respectively, then the quantity $\theta = \mathbb{P}(X < Y)$ can be interpreted as the effectiveness of the treatment (Kotz et al., 2003; Ventura et al., 2011). Another important use of $\theta = \mathbb{P}(X < Y)$ in medicine is related to the analysis of receiver operating characteristic (ROC) curves, where θ naturally appears as an index of diagnostic accuracy (Zhou, 2008). The parameter θ can be seen as a function of the parameters of the distribution of the random vector (X, Y) and can be calculated in closed form for a limited number of cases (Kotz et al., 2003; Nadarajah, 2005; Genç, 2011). There is a large amount of literature about the estimation of θ using different approaches and distributional assumptions on (X, Y) (e.g. Kotz et al., 2003, Greco and Ventura, 2011 and Ventura et al., 2011). For instance, it has been assumed that

(i) X and Y are independent (Zhou, 2008; Ventura et al., 2011).

(ii) The distributions of X and Y share common parameters (Gupta and Peng, 2009).

[*] University of Warwick, Department of Statistics, Coventry, CV4 7AL, UK. E-mail: F.J.Rubio@warwick.ac.uk

[†] University of Warwick, Department of Statistics, Coventry, CV4 7AL, UK. E-mail: M.F.Steel@stats.warwick.ac.uk

(iii) The distributions of X and Y are independent skewed normals (Azzalini and Chiogna, 2004; Gupta and Brown, 2001).

(iv) X and Y are dependent with a bivariate normal distribution (Nandi and Aich, 1994; Barbiero, 2011)

(v) X and Y are conditionally (on certain unobservable variables) independent exponential random variables (Shoukri et al., 2005).

Although closed expressions for the profile likelihood and modified profile likelihood of θ have been calculated for some particular cases (Montoya, 2008; Ventura et al., 2011), it is difficult (if at all feasible) in the general case to find a reparameterization of the model parameters that involves θ (e.g. Azzalini and Chiogna (2004)). This complicates the calculation of the profile likelihood of the parameter θ, and therefore, the interval estimation using the classical approach.

Alternative inferential approaches for estimating this parameter have also been proposed; for example, the use of confidence intervals (see Kotz et al. (2003)), asymptotic confidence intervals and bootstrap (Zhou, 2008), Bayesian inference using reference priors (Sun et al., 1998), modified profile likelihoods and Bayesian inference using Jeffreys priors (Ventura et al., 2011). It is worth pointing out that all these approaches were proposed under specific distributional assumptions on the variables X and Y.

To our knowledge, there is a gap in the cases analysed in the literature. The case where X and Y are dependent and the case where their marginal distributions are skewed with support on $\mathbb{R}$ have been analysed separately. This paper tries to fill this gap by analysing the case where X and Y are dependent with marginal distributions belonging to the class of distributions obtained by skewing scale mixtures of normals. In addition, we address this problem in the context of set observations, which can immediately account for censoring.

In Section 2, we study the case where X and Y are independent with particular focus on the case where their distributions are skewed. We consider skewed distributions obtained with two different skewing mechanisms: two-piece distributions (Fernández and Steel, 1998; Mudholkar and Hutson, 2000; Arellano-Valle et al., 2005) and skew-symmetric distributions (Wang et al., 2004). We propose noninformative benchmark priors and present mild conditions for the existence of the posterior distribution of θ. In Section 3, we study the case where X and Y are dependent random variables with skewed marginal distributions. Dependencies between X and Y are modelled using a Gaussian copula. We provide noninformative benchmark priors and present conditions for the existence of the posterior distribution of θ. In Section 4, we show that the Bayesian models presented here can be used in the presence of set observations. In Section 5, we illustrate the use of these models using simulated and real data sets.

2 Independent case

In this section, we present Bayesian models to conduct inference on $\theta = \mathbb{P}(X < Y)$ in the case where X and Y are independent variables with densities $f_1(\cdot; \boldsymbol{\xi_1})$ and $f_2(\cdot; \boldsymbol{\xi_2})$, respectively. Cumulative distribution functions are denoted with the corresponding uppercase letters throughout. We focus on the case where f_1 and f_2 are skewed distributions and we also present conditions for the existence of the posterior distribution of θ under the use of improper benchmark priors.

If we adopt a product prior structure

$$P_{(\boldsymbol{\xi_1},\boldsymbol{\xi_2})} \propto P_{\boldsymbol{\xi_1}} \times P_{\boldsymbol{\xi_2}}, \tag{1}$$

where $P_{\boldsymbol{\xi_1}}$ and $P_{\boldsymbol{\xi_2}}$ are priors such that the corresponding posteriors are well-defined, then the posterior distribution of θ is well-defined as shown in the next theorem.

Theorem 1 *Let X and Y be two independent random variables with distributions $f_1(\cdot;\boldsymbol{\xi_1})$ and $f_2(\cdot;\boldsymbol{\xi}_2)$, respectively. Let $\mathbf{x} = (x_1, ..., x_m)$ and $\mathbf{y} = (y_1, ..., y_n)$ be two independent samples from X and Y. Then, the posterior distribution of θ, using the product prior structure (1), is proper if the corresponding posteriors of $\boldsymbol{\xi_1}$ and $\boldsymbol{\xi_2}$ are proper.*
Proof. *See Appendix.*

An example of this is the use of the Jeffreys prior of $\boldsymbol{\xi_1}$ and $\boldsymbol{\xi_2}$ under a normal or exponential sampling model as in Ventura et al. (2011). A second example of this is the use of reference priors for $\boldsymbol{\xi_1}$ and $\boldsymbol{\xi_2}$ under a Weibull sampling model as studied in Sun et al. (1998). In the following sections, we study the cases where the marginal distributions of X and Y belong to the family of skewed scale mixtures of normals obtained by two different skewing mechanisms.

2.1 Two-Piece marginals

Let s_1 and s_2 be two symmetric densities with support on $\mathbb{R}$, location parameters $\mu_j \in \mathbb{R}$ and scale parameters $\sigma_j \in \mathbb{R}^+$, $j = 1, 2$ respectively. Let X and Y be two independent continuous random variables with densities given respectively by (Arellano-Valle et al., 2005)

$$
\begin{aligned}
f_1(x;\mu_1,\sigma_1,\gamma_1) &= \frac{2}{\sigma_1[a(\gamma_1)+b(\gamma_1)]}\left[s_1\left(\frac{x-\mu_1}{\sigma_1 b(\gamma_1)}\right)I_{(-\infty,\mu_1)}(x) + s_1\left(\frac{x-\mu_1}{\sigma_1 a(\gamma_1)}\right)I_{[\mu_1,\infty)}(x)\right], \\
f_2(y;\mu_2,\sigma_2,\gamma_2) &= \frac{2}{\sigma_2[a(\gamma_2)+b(\gamma_2)]}\left[s_2\left(\frac{y-\mu_2}{\sigma_2 b(\gamma_2)}\right)I_{(-\infty,\mu_2)}(y) + s_2\left(\frac{y-\mu_2}{\sigma_2 a(\gamma_2)}\right)I_{[\mu_2,\infty)}(y)\right]
\end{aligned} \quad (2)
$$

where $\gamma_j \in \Gamma$ and Γ depends on the choice of $\{a(\cdot), b(\cdot)\}$ where $a(\cdot)$ and $b(\cdot)$ are positive and differentiable functions. The main examples found in the literature are $\{a(\gamma), b(\gamma)\} = \{\gamma, 1/\gamma\}$, $\gamma > 0$ (Fernández and Steel, 1998) and $\{a(\gamma), b(\gamma)\} = \{1-\gamma, 1+\gamma\}$, $\gamma \in (-1, 1)$ (Mudholkar and Hutson, 2000). The densities f_1 and f_2 can be interpreted as skewed versions of s_1 and s_2, and are often called "two-piece" distributions. If we measure skewness using the measure in Arnold and Groeneveld (1995) (which is defined as one minus twice the probability mass to the left of the mode and takes values in $[-1, 1]$), Rubio and Steel (2011) find that for these distributions this skewness measure becomes

$$AG = AG(\gamma_j) = \frac{a(\gamma_j) - b(\gamma_j)}{a(\gamma_j) + b(\gamma_j)}, \quad j = 1, 2.$$

Therefore, we can see that γ_j controls the allocation of mass each side of the mode of the transformed distribution. This result lets us interpret the parameter γ_j as a skewness parameter for the typical choices of $\{a(\cdot), b(\cdot)\}$ found in the literature.

For the purpose of conducting Bayesian inference for the parameter $\theta = \mathbb{P}(X < Y)$ we consider the priors

$$p(\mu_j, \sigma_j, \gamma_j|\alpha_j, \beta_j) \propto \frac{1}{\sigma_j}\frac{|a'(\gamma_j)b(\gamma_j) - a(\gamma_j)b'(\gamma_j)|}{[a(\gamma_j) + b(\gamma_j)]^{\alpha_j+\beta_j}}a(\gamma_j)^{\alpha_j-1}b(\gamma_j)^{\beta_j-1}, \quad j = 1, 2. \quad (3)$$

The structure of these priors is the product of the Independence Jeffreys prior for a symmetric location-scale model and a Beta(α_j, β_j) distribution on the parameter $(AG(\gamma_j) + 1)/2$ (Rubio and Steel, 2011). Note that if $\alpha_j = \beta_j = 1$, then the latter prior is equivalent to setting a uniform prior over the measure of skewness AG. Conditions for the existence of the posterior distribution of θ using this prior are given in the following corollary.

Corollary 1 *Let* $\mathbf{x} = (x_1, ..., x_m)$ *and* $\mathbf{y} = (y_1, ..., y_n)$ *be two independent samples from the models in* (2) *and* (3)*, where* s_1 *and* s_2 *are scale mixtures of normals. Then, the posterior distribution of* θ *is proper for any parameterization* $\{a(\cdot), b(\cdot)\}$ *if* $m, n \geq 2$ *and all the observations are different.*
Proof. *The result is a consequence of Theorem 1 above and Theorem 6 from Rubio and Steel (2011).*

If the samples $\mathbf{x}$ and $\mathbf{y}$ contain repeated observations, then, using the proof of Theorem 6 from Rubio and Steel (2011), we find that condition (4.5) in Theorem 2 from Fernández and Steel (1999) also has to be satisfied by the mixing probabilities of X and Y. In the case of a two-piece normal sampling model, it suffices to have two different observations in each sample (see Theorem 3 from Fernández and Steel, 1999).

2.2 Skew-symmetric marginals

We now consider the case where X and Y are independent random variables with skew-symmetric distributions as in Wang et al. (2004). Let s_1 and s_2 be two symmetric densities with support on $\mathbb{R}$, location parameters $\mu_j \in \mathbb{R}$, scale parameters $\sigma_j \in \mathbb{R}^+$, $j = 1, 2$ respectively, and define

$$\begin{aligned} f_1(x; \mu_1, \sigma_1, \pi_1) &= \frac{2}{\sigma_1} s_1\left(\frac{x-\mu_1}{\sigma_1}\right) \pi_1\left(\frac{x-\mu_1}{\sigma_1}\right), \\ f_2(y; \mu_2, \sigma_2, \pi_2) &= \frac{2}{\sigma_2} s_2\left(\frac{y-\mu_2}{\sigma_2}\right) \pi_2\left(\frac{y-\mu_2}{\sigma_2}\right), \end{aligned} \tag{4}$$

where $\pi_j(\cdot)$ are functions that satisfy $0 \leq \pi_j(x) \leq 1$ and $\pi_j(-x) = 1 - \pi_j(x)$. We use parametric skewing functions $\pi_j(\cdot; \lambda_j)$, $\lambda_j \in \Lambda_j$, and adopt the prior structure

$$p(\mu_j, \sigma_j, \lambda_j) \propto \sigma_j{}^{-1} p_{\lambda_j}(\lambda_j), \;\; j = 1, 2, \tag{5}$$

where p_{λ_j} is an integrable function over Λ_j. Conditions for the existence of the posterior distribution of θ using this prior are given in the following corollary.

Corollary 2 *Let* $\mathbf{x} = (x_1, ..., x_m)$ *and* $\mathbf{y} = (y_1, ..., y_n)$ *be two independent samples from the model* $(4) - (5)$*, where* s_1 *and* s_2 *are scale mixtures of normals. Then, the posterior distribution of* θ *is proper if* $m, n \geq 2$ *and all the observations are different.*
Proof. *See Appendix.*

Again, if the samples $\mathbf{x}$ and $\mathbf{y}$ contain repeated observations, then, by inequality (14) in the Appendix, we also need condition (4.5) in Theorem 2 from Fernández and Steel (1999) for the mixing probabilities of X and Y. In the case of a skew-symmetric normal sampling model, where s_1 and s_2 are normal, it suffices to have two different observations in each sample (see Theorem 3 from Fernández and Steel, 1999 and Liseo and Loperfido, 2006).

A particular case of model (4) is the Azzalini skew-normal (Azzalini, 1985), which is obtained by setting $\pi_j(x; \lambda_j) = \Phi(\lambda_j x)$, $\lambda_j \in \mathbb{R}$, and $s_1 = s_2 = \phi$, where Φ and ϕ are the standard normal CDF and PDF, respectively. This model is frequently used in applications and will be considered for the examples in Section 5 together with the prior

$$p(\mu_j, \sigma_j, \lambda_j) \propto \sigma_j{}^{-1} p_J(\lambda_j), \;\; j = 1, 2. \tag{6}$$

The structure of this prior, using the Jeffreys prior of λ_j derived in the model without location and scale parameters for $p_J(\lambda_j)$, was proposed in Liseo and Loperfido (2006), who also prove existence of the posterior under this prior. Bayes and Branco (2007) show that the Jeffreys prior of λ_j can be approximated by a Student t distribution with $1/2$ degrees of freedom.

3 Dependent case

In this section, we focus on Bayesian inference for $\theta = \mathbb{P}(X < Y)$ in the case where X and Y are dependent random variables with marginal distributions $f_1(\cdot; \boldsymbol{\xi_1})$ and $f_2(\cdot; \boldsymbol{\xi_2})$, respectively. We pay special attention to the case where the marginal distributions are skewed and we use a Gaussian copula for modelling dependencies between X and Y. The density of the Gaussian copula is given by

$$\begin{aligned} s(x, y; \boldsymbol{\xi_1}, \boldsymbol{\xi_2}, \rho) &= \frac{1}{\sqrt{1-\rho^2}} \exp\left[-\frac{V^T(R^{-1} - I)V}{2}\right] \\ &\times f_1(x; \boldsymbol{\xi_1}) f_2(y; \boldsymbol{\xi_2}), \end{aligned} \tag{7}$$

where

$$R = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix},$$

is a correlation matrix with $\rho \in (-1, 1)$ and $V = (\Phi^{-1}[F_1(x; \boldsymbol{\xi_1})], \Phi^{-1}[F_2(y; \boldsymbol{\xi_2})])^T$. This copula presents some appealing features like being comprehensive, symmetric (in the sense that positive and negative dependence is treated equally) and also that the Spearman's measure of association, $r_\rho \in (-1, 1)$, can be calculated in closed form as (Carta and Steel, 2011)

$$r_\rho = \frac{6}{\pi} \arcsin\left(\frac{\rho}{2}\right).$$

We adopt a product prior structure

$$P_{(\boldsymbol{\xi_1}, \boldsymbol{\xi_2}, \rho)} \propto P_{\boldsymbol{\xi_1}} \times P_{\boldsymbol{\xi_2}} \times P_\rho, \tag{8}$$

where $P_{\boldsymbol{\xi_1}}$ and $P_{\boldsymbol{\xi_2}}$ are priors such that the corresponding posteriors are well-defined and P_ρ is a distribution with density $p_\rho(\rho)$ which satisfies

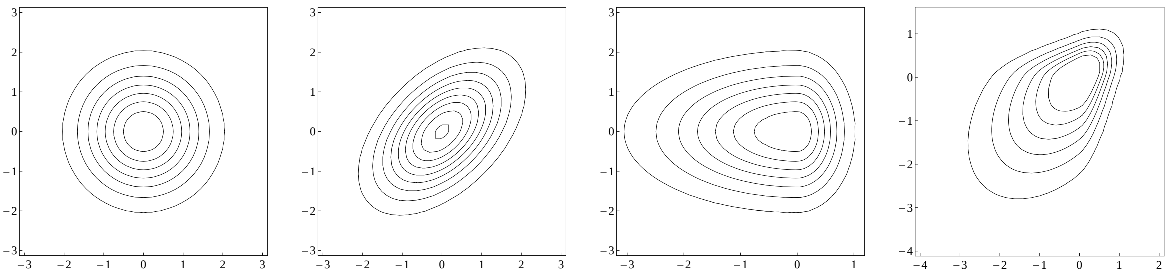
$$\int_{-1}^{1} \frac{p_\rho(\rho)}{\sqrt{1-\rho^2}} d\rho < \infty. \tag{9}$$

In particular, any bounded prior density on ρ satisfies this condition. The posterior distribution of θ is well-defined for this Bayesian model as shown in the next theorem.

Theorem 2 *Let* $\mathbf{x} = ((x_1, y_1), ..., (x_n, y_n))$ *be a sample from* (X, Y)*, distributed as in* (7)*. Then, the posterior distribution of* θ*, using the prior structure* (8)*, is proper if the corresponding posteriors of* $\boldsymbol{\xi_1}$ *and* $\boldsymbol{\xi_2}$ *are proper.*
Proof. *See Appendix.*

3.1 Two-piece marginals

Consider the case where X and Y are dependent random variables with marginal distributions given by (2). The dependency between X and Y is modelled with a Gaussian copula as in (7). Figure 1 shows some contour plots obtained for this copula density using the parameterization in Mudholkar and Hutson (2000), $\{a(\gamma), b(\gamma)\} = \{1-\gamma, 1+\gamma\}$ and $s_1 = s_2 = \phi$. By appropriately choosing the parameters γ_1, γ_2 and ρ, we can assign a wide range of shapes to the density. The mode of the density is not affected by changes in the parameters, in line with the mode-preserving property of the two-piece skewing mechanism.



(a) (b) (c) (d)

Figure 1: Contour plots: two-piece skew-normal marginals with $\mu_1 = \mu_2 = 0$, $\sigma_1 = \sigma_2 = 1$ and (a) $\gamma_1 = \gamma_2 = 0$, $\rho = 0$; (b) $\gamma_1 = \gamma_2 = 0$, $\rho = 0.5$; (c) $\gamma_1 = 0.5$, $\gamma_2 = 0$, $\rho = 0$; (d) $\rho = \gamma_1 = \gamma_2 = 0.5$.

For the parameters of this model, we adopt the product prior structure

$$\begin{aligned} p(\mu_1, \mu_2, \sigma_1, \sigma_2, \gamma_1, \gamma_2, \rho | a_1, b_1, a_2, b_2) \quad &\propto \quad \frac{1}{\sigma_1} \frac{|a'(\gamma_1)b(\gamma_1) - a(\gamma_1)b'(\gamma_1)|}{[a(\gamma_1) + b(\gamma_1)]^{\alpha_1+\beta_1}} a(\gamma_1)^{\alpha_1-1} b(\gamma_1)^{\beta_1-1} \\ &\times \quad \frac{1}{\sigma_2} \frac{|a'(\gamma_2)b(\gamma_2) - a(\gamma_2)b'(\gamma_2)|}{[a(\gamma_2) + b(\gamma_2)]^{\alpha_2+\beta_2}} a(\gamma_2)^{\alpha_2-1} b(\gamma_2)^{\beta_2-1} \\ &\times \quad p_\rho(\rho), \end{aligned} \tag{10}$$

where $p_\rho(\rho)$ is a prior that satisfies (9). The structure of this prior is the product of the priors in (3) and the prior of the correlation coefficient, ρ. Of particular interest can be the case $\alpha_j = \beta_j = 1$ together with

$$p_\rho(\rho) \propto \frac{1}{1 - \left(\frac{\rho}{2}\right)^2}, \tag{11}$$

which corresponds to $AG \sim U(-1, 1)$ for both marginals and $r_\rho \sim U(-1, 1)$.

If $((x_1, y_1), ..., (x_n, y_n))$ is a sample from (X, Y) distributed as in (7) with f_1 and f_2 given by (2), where s_1 and s_2 scale mixtures of normals, then the posterior distribution of θ, using the prior (10), is proper if $n \geq 2$ and all the observations are different. This follows using Theorem 2 and Corollary 1. The additional condition in Section 2.1 applies in the context of repeated observations.

3.2 Skew-symmetric marginals

Here we focus on the case where X and Y are dependent random variables with skew-symmetric marginal distributions (4). Figure 2 shows some contour plots obtained for the copula density in (7) with Azzalini skew-normal marginals. By varying the parameters, it is possible to cover a wide range of shapes, but note that there is a shift of the mode relative to the symmetric case.

For the parameters of this model, we adopt a product structure for the prior

$$p(\mu_1, \mu_2, \sigma_1, \sigma_2, \lambda_1, \lambda_2, \rho) \quad \propto \quad \frac{1}{\sigma_1 \sigma_2} p_{\lambda_1}(\lambda_1) p_{\lambda_2}(\lambda_2) p_\rho(\rho), \tag{12}$$

where $p_\rho(\rho)$ is a prior that satisfies (9) and $p_{\lambda_j}(\lambda_j)$ are integrable functions over Λ_j, $j = 1, 2$. The structure of this prior is the product of the priors in (5) and the prior of the correlation coefficient, ρ.

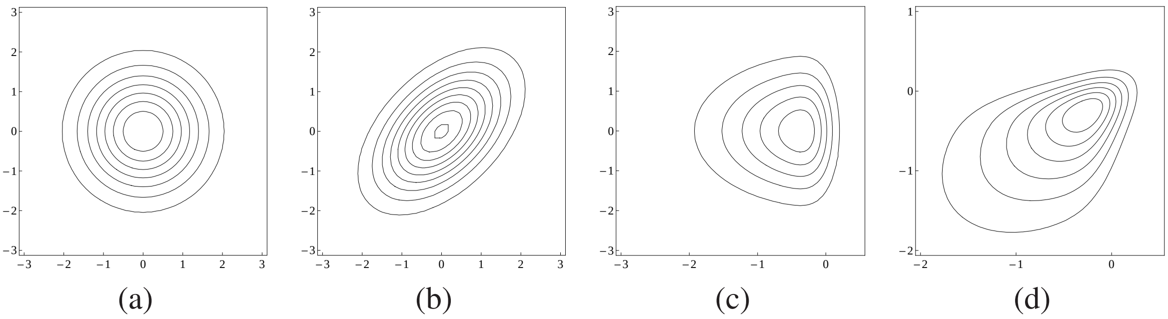


Figure 2: Contour plots: Azzalini skew-normal marginals with $\mu_1 = \mu_2 = 0$, $\sigma_1 = \sigma_2 = 1$ and (a) $\lambda_1 = \lambda_2 = 0$, $\rho = 0$; (b) $\lambda_1 = \lambda_2 = 0$, $\rho = 0.5$; (c) $\lambda_1 = -5$, $\lambda_2 = 0$, $\rho = 0$; (d) $\lambda_1 = \lambda_2 = -5$, $\rho = 0.5$.

If $((x_1, y_1), ..., (x_n, y_n))$ is a sample from (X, Y), distributed as in (7) with f_1 and f_2 given by (4), where s_1 and s_2 scale mixtures of normals, then the posterior distribution of θ, using the prior (12), is proper if $n \geq 2$ and all the observations are different. This follows from Theorem 2 and Corollary 2. The extra condition in Section 2.2 applies if the sample contains repeated observations.

4 Set observations

A common phenomenon in reliability and survival analysis is the presence of set observations under a continuous sampling model. A set observation y is produced when a measurement is recorded as a set of positive probability, this is

$$\mathbb{P}[\text{Observing } y] = \mathbb{P}[y \in S] > 0,$$

where S is a Borel set. This corresponds to any observation in practice recorded with finite precision, and also includes left, right and interval censoring. When the quantitative effect of censoring is significant, this must be formally taken into account in the model (Heitjan, 1989). In the following theorem, conditions for the existence of the posterior distribution of θ using the Bayesian models from Section 2 in the context of set observations are presented.

Theorem 3 *Let $\mathbf{x} = (x_1, ..., x_m)$ and $\mathbf{y} = (y_1, ..., y_n)$ be two independent samples from the model $(2)-(3)$ or $(4)-(5)$, where s_1 and s_2 are scale mixtures of normals. Then, the posterior distribution of θ is proper if $m, n \geq 2$ and there exist two pairs of sets, say (S_i, S_j) and (S'_i, S'_j), such that*

$$\begin{aligned} \inf_{x_i \in S_i, x_j \in S_j} |x_i - x_j| &> 0, \\ \inf_{y_i \in S'_i, y_j \in S'_j} |y_i - y_j| &> 0. \end{aligned} \tag{13}$$

Proof. *See Appendix.*

Thus, whenever each sample of set observations contains at least two intervals that do not overlap, we can conduct Bayesian inference on θ. In practice, of course, this is very likely to be satisfied for any samples that we would seriously think of analysing.

Using this result, we can also prove the existence of the posterior distribution of θ for the copula models presented in Section 3 under the same conditions on the set observations.

Corollary 3 *Let* $((x_1, y_1), \ldots, (x_n, y_n))$ *be a sample from* (X, Y)*, distributed as in* (7) *with* f_1 *and* f_2 *given by* (2) *or* (4)*, where* s_1 *and* s_2 *are scale mixtures of normals. Then, the posterior distribution of* θ *using the prior* (10) *or* (12)*, respectively, is proper if* $n \geq 2$ *and there exist two pairs of sets, say* (S_i, S_j) *and* (S'_i, S'_j)*, such that* (13) *is satisfied.*
Proof. *Using inequality* (15)*, we have that it suffices to prove the existence of the posterior of* θ *in the independent case and the result follows by Theorem 3.*

5 Examples

In this section, three examples are presented to illustrate the use of the Bayesian models for $\theta = \mathbb{P}(X < Y)$ in different scenarios: independent observations, dependent observations and set observations. Throughout, in order to obtain inferences for θ, we consider the use of the marginal sampling models (2) and (4) with $s_1 = s_2 = \phi$. In the case of the two-piece marginal we adopt the parameterization $\{a(\gamma), b(\gamma)\} = \{1-\gamma, 1+\gamma\}$, $\gamma \in (-1, 1)$, and use the prior in (3) and (10) with $\alpha_j = \beta_j = 1$. We compare this model with the Bayesian model with Azzalini skew-normal marginals and the prior in (6) and (12). For the dependent cases, modelled as in (7), we use the prior on ρ in (11). Using a Metropolis-Hastings algorithm, a posterior sample of size $10,000$ of the corresponding model parameters was simulated using a burn-in period of $50,000$ iterations and a thinning of 100 iterations. Then, through numerical integration, the corresponding posterior sample of θ was calculated.

5.1 Independent case

5.1.1 Simulated data

First, we present an example using simulated data which illustrates the importance of taking departures from symmetry into account, particularly in the case where X and Y display quite different skewness properties. Two independent samples of size 250 from the two-piece skew-normal model were drawn with $\mu_i = 10, \sigma_i = 1, i = 1, 2$ and X generated with $\gamma_1 = 0.75$ and Y using $\gamma_2 = -0.75$. Using these data, a posterior sample of θ for the following three Bayesian models was simulated: (i) (2)-(3), with $\alpha_j = \beta_j = 1$; (ii) (4)-(6); and (iii) a normal sampling model for X and Y together with the independence Jeffreys prior $p(\mu_1, \mu_2, \sigma_1, \sigma_2) \propto \sigma_1^{-1}\sigma_2^{-1}$. Figure 3 shows the posterior distribution of θ for these models. We can observe a clear discrepancy between the inference obtained with the symmetric and the asymmetric sampling models. Properly accounting for skewness centers the inference nicely around the theoretical value (calculated using unidimensional numerical integration) for θ of 0.9646. In addition, both skewed models produce very similar inference about θ.

In the applications with real data, the skewness properties of X and Y are much more similar, and thus the inference on θ is not as crucially affected by allowing for skewness in the marginals. Of course, inference related to the marginals themselves will typically be more sensitive to the modelling of skewness.

5.1.2 Body measurements

An important goal of forensic studies is to determine the gender of adults given their skeletal remains (Heinz et al., 2003). Therefore, it is important to assess if certain body measurements are informative about the gender. Here we analyse the variable "Chest depth between spine and sternum at nipple level, mid-expiration" from the data set presented in Heinz et al. (2003). This

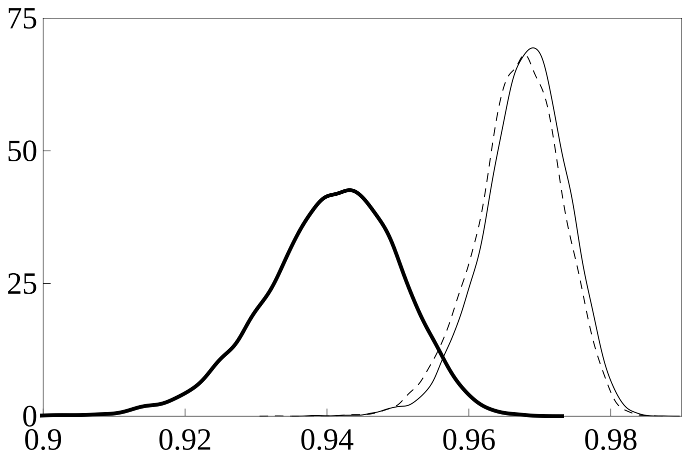


Figure 3: Simulated data: posterior distribution of θ, two-piece skew-normal (solid line), Azzalini skew-normal (dashed) and normal (bold).

sample consists of 507 measurements taken on physically active adults, 260 females and 247 males. In this case, it seems reasonable to assume independence between the measurements on females and males given that no relationship between the individuals is known. In addition, the histograms in Figure 4 suggest departure from symmetry. Figure 5 shows the posterior distributions of $\theta = P(\text{female chest depth} < \text{male chest depth})$. This figure indicates that this variable can be informative about the gender given that the posterior of θ assigns most of the mass to values bigger than 0.5. Both models produce similar inferences about θ.

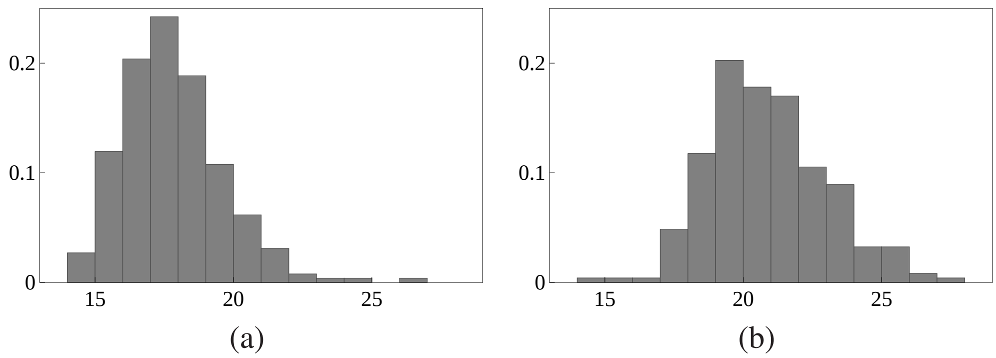


Figure 4: Histograms of Chest depth data: (a) females; (b) males.

5.2 Dependent case

We now analyse 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. This data set was also considered in Gupta and Peng (2009) using bootstrap and asymptotic confidence intervals but assuming independence between the X Test and the Y Test. This assumption is somewhat restrictive because each pair of observations was measured in the same patient. In fact, the population correlation coefficient is 0.794 and we can observe this positive correlation in the scatter plot in Figure 6. Here, we analyse the subset

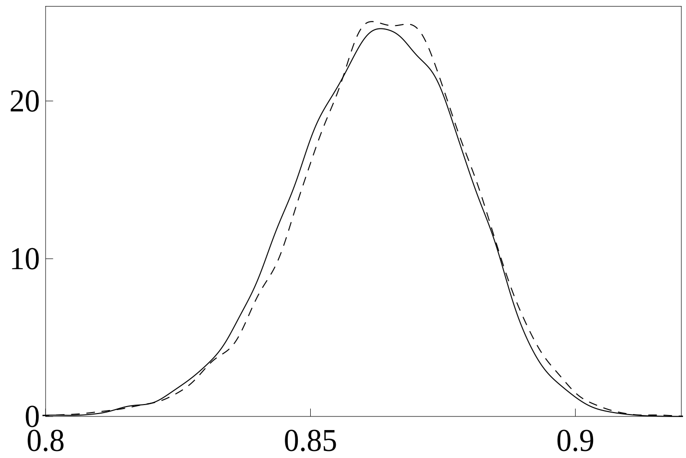


Figure 5: Chest depth data: posterior distribution of θ, two-piece skew-normal (solid line) and Azzalini skew-normal (dashed).

of 51 non-diseased patients (diagnosed using a biopsy) and compare the Bayesian inferences obtained under both assumptions: independence and dependence of the tests. Figure 7 shows the posterior distributions of $\theta = P(\text{Y Test} < \text{X Test})$ for both scenarios. We see that the conclusions are substantially affected by taking the dependence of the variables into account. In contrast, both marginal specifications lead to similar results, as in the previous application.

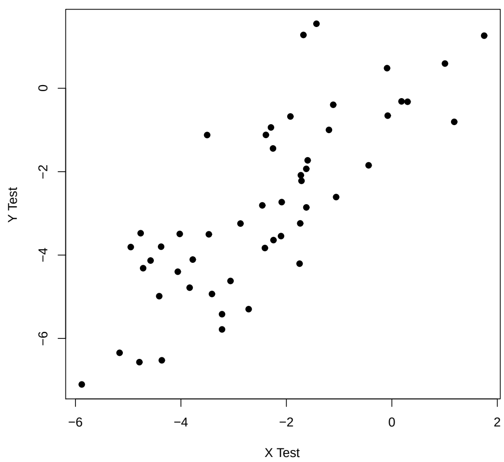


Figure 6: Melanoma data: scatter plot.

5.3 Set observations

To illustrate the use of the Bayesian models for θ in the presence of censoring, we consider the breast cancer data set from Finkelstein and Wolfe (1985). This data set contains the times until cosmetic deterioration, determined by evaluation of breast retraction, observed for two treatments (46 observations for the first treatment and 48 observations for the second one): Radiotherapy (R) and Radiotherapy + Chemotherapy (RC). The presence of cosmetic dete-

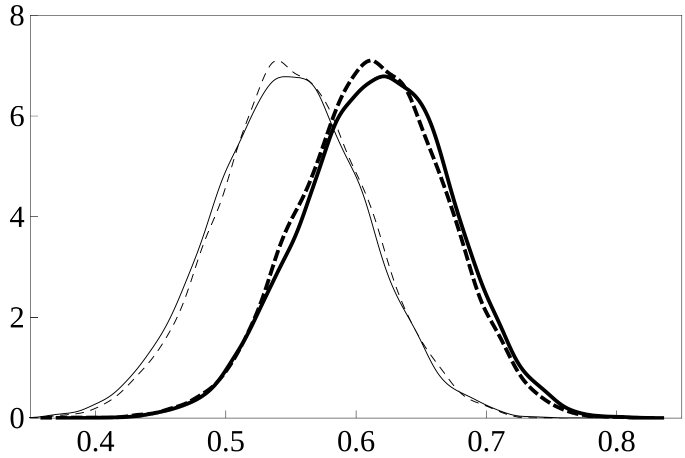


Figure 7: Melanoma data: posterior distributions of θ; two-piece skew-normal independent case (solid line), Azzalini skew-normal independent case (dashed), two-piece skew-normal dependent case (bold) and Azzalini skew-normal dependent case (bold dashed).

rioration is observed in between two appointments, so that the observations are recorded as intervals. The assumption of independence between X and Y seems to be reasonable here, but we do take the censoring into account. Since these observations are positive and some of them are close to zero, we analyse the logarithm of the original observations. Figure 8 shows the posterior distribution of $\theta = P(\mathrm{R} < \mathrm{RC})$. The posterior mass is clearly concentrated on values smaller than 0.5. This is in line with the conclusion in Finkelstein and Wolfe (1985) that the group receiving both radiotherapy and chemotherapy experiences an earlier cosmetic deterioration.

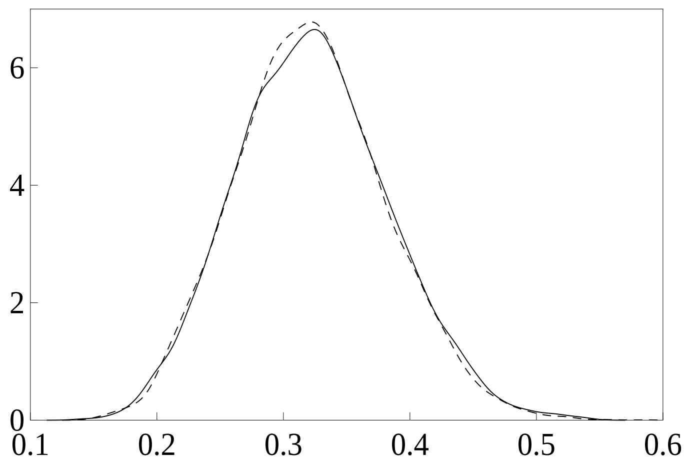


Figure 8: Breast cancer data: posterior distributions of θ; two-piece skew-normal model (solid line), Azzalini skew-normal model (dashed).

6 Conclusions

We have presented Bayesian models for the parameter $\theta = P(X < Y)$ in the case where the marginal distributions of X and Y belong to the family of skewed scale mixtures of normals. In general, the Bayesian approach overcomes the classical issue regarding the need for an explicit transformation involving this parameter of interest. This allows us to study this problem in more complex scenarios such as the case where X and Y are dependent variables and the context of set observations. Section 5 illustrates, through different examples using simulated and real data sets, the relevance of including these assumptions into the model.

Despite the similarities of the inference using two-piece marginals and skew symmetric marginals observed in the examples, simulating from the posterior distribution of θ using two-piece distributions tends to be easier than with skew-symmetric distributions. The reason may be the ill-behaved likelihood function obtained with these skewing functions (Arnold et al., 1993; Ley and Paindaveine, 2010).

Finally, let us mention two natural directions in which the results presented here can be extended. Firstly, Theorems 1 and 2 can immediately be applied to contexts with different marginal distributional assumptions. Secondly, we can consider the use of other bivariate copulas (e.g. Archimedean copulas) for modelling dependencies between X and Y, which would require the specification of appropriate priors.

Appendix: Proofs

Proof of Theorem 1

Using that the transformation from $(\boldsymbol{\xi_1}, \boldsymbol{\xi_2})$ to θ,

$$\theta = \int_{\mathbb{R}} F_1(y; \boldsymbol{\xi_1}) f_2(y, \boldsymbol{\xi_2}) dy,$$

is a measurable function of the parameters (Enis and Geisser, 1971), we get that the posterior distribution of θ is also well-defined if the posterior of $\boldsymbol{\xi_1}$ and $\boldsymbol{\xi_2}$ is well-defined.

Proof of Corollary 2

The result follows using the upper bound

$$f_j(x|\mu_j, \sigma_j, \pi_j) \leq \frac{2}{\sigma_j} s_j \left(\frac{x - \mu_j}{\sigma_j} \right), \tag{14}$$

together with Theorem 1 above and the properness of the posterior of the parameters in the symmetric case under the use of this prior structure (Theorem 1, Fernández and Steel (1999)).

Proof of Theorem 2

First of all, note that

$$s(x, y; \boldsymbol{\xi_1}, \boldsymbol{\xi_2}, \rho) \leq \frac{1}{\sqrt{1 - \rho^2}} f_1(x; \boldsymbol{\xi_1}) f_2(y; \boldsymbol{\xi_2}). \tag{15}$$

Using this bound together with the properness of the posterior of $\boldsymbol{\xi_1}$ and $\boldsymbol{\xi_2}$, we have that the posterior distribution of $(\boldsymbol{\xi_1}, \boldsymbol{\xi_2}, \rho)$ is proper, using the prior in (8). Now, using that the transformation from $(\boldsymbol{\xi_1}, \boldsymbol{\xi_2}, \rho)$ to θ

$$\theta = \int_{\mathbb{R}} \int_{-\infty}^{y} s(x, y; \boldsymbol{\xi_1}, \boldsymbol{\xi_2}, \rho) dx dy,$$

is a measurable function of the parameters (Enis and Geisser, 1971), we get that the posterior distribution of θ is also well-defined.

Proof of Theorem 3

Using Theorem 1 we have that it suffices to prove existence of the posterior distribution of the model parameters.

For model $(2) - (3)$, let s_1 be a scale mixture of normals with λ_j the mixing variable associated with x_j and where the λ_j's are independent random variables defined on $\mathbb{R}^+$ with distribution P_{λ_j}. We get an upper bound for the marginal distribution of $(x_1, ..., x_m)$ proportional to

$$\int_{S_1 \times \cdots \times S_m} \int_{\mathbb{R}_m^+} \int_{\Gamma_1} \int_0^\infty \int_{-\infty}^\infty \left(\prod_{j=1}^m \lambda_j^{\frac{1}{2}} \right) \frac{\sigma_1^{-(m+1)}}{[a(\gamma_1) + b(\gamma_1)]^m} \exp \left[-\frac{1}{2\sigma_1^2 h(\gamma_1)^2} \sum_{j=1}^m \lambda_j (x_j - \mu_1)^2 \right]$$
$$\times \quad p_{\gamma_1}(\gamma_1)\, d\mu_1 d\sigma_1 d\gamma_1 dP_{(\lambda_1,\ldots,\lambda_m)} d\mathbf{x},$$

where $h(\gamma_1) = \max\{a(\gamma_1), b(\gamma_1)\}$ and $p_{\gamma_1}(\gamma_1)$ is the factor dependent of γ_1 in (3). Consider the change of variable $\vartheta = \sigma_1 h(\gamma_1)$ and rewrite the upper bound as follows

$$\int_{\Gamma_1} \left[\frac{h(\gamma_1)}{a(\gamma_1) + b(\gamma_1)} \right]^m p_{\gamma_1}(\gamma_1)\, d\gamma_1 \int_{S_1 \times \cdots \times S_m} \int_{\mathbb{R}_m^+} \int_0^\infty \int_{-\infty}^\infty \left(\prod_{j=1}^m \lambda_j^{\frac{1}{2}} \right) \vartheta^{-(m+1)}$$
$$\times \quad \exp \left[-\frac{1}{2\vartheta^2} \sum_{j=1}^m \lambda_j (x_j - \mu_1)^2 \right] d\mu_1 d\vartheta dP_{(\lambda_1,\ldots,\lambda_m)} d\mathbf{x}.$$

The integral with respect to γ_1 is finite for any m and by Theorem 4 from Fernández and Steel (1999) we have that the integral in $(\mu_1, \vartheta, \lambda_1, ..., \lambda_m, \mathbf{x})$ is finite if (13) is satisfied. Analogously for $\mathbf{y}$.

For model $(4) - (5)$, using inequality (14) we find that, for skew symmetric scale mixtures of normals sampling models, the posterior of θ exists whenever the posterior distribution of the parameters in the symmetric case exists. Thus, by Theorem 4 from Fernández and Steel (1999) we find that this happens whenever (13) is satisfied.

References

Arellano-Valle, R. B., Gómez, H. W. and Quintana, F. A. (2005). Statistical inference for a general class of asymmetric distributions. *Journal of Statistical Planning and Inference* 128: 427–443.

Arnold, B. C. and Groeneveld, R. A. (1995). Measuring skewness with respect to the mode. *The American Statistician* 49: 34–38.

Arnold, B. C., Beaver, R. J., Groeneveld, R. A. and Meeker, W. Q. (1993). The nontruncated marginal of a truncated bivariate normal distribution. *Psychometrika* 58: 471-478.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics* 12: 171-178.

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

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

Bayes, C. L. and Branco, M. D. (2007). Bayesian inference for the skewness parameter of the scalar skew-normal distribution. *Brazilian Journal of Probability and Statistics* 21: 141–163.

Carta, A. and Steel, M. F. J. (2011). Modelling multi-output stochastic frontiers using copulas. *Computational Statistics and Data Analysis*, forthcoming.

Enis, P. and Geisser, S. (1971). Estimation of the probability that $Y < X$. *Journal of the American Statistical Association* 66: 162–168.

Fernández, C. and Steel, M. F. J. (1998). On Bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association* 93: 359–371.

Fernández, C. and Steel, M. F. J. (1999). On the dangers of modelling through continuous distributions: a Bayesian perspective (with discussion). In *Bayesian Statistics* 6. J. O. Berger, J. M. Bernardo, A.P. Dawid and A. F. M. Smith

Finkelstein, D. M. and Wolfe, R. A. (1985). A semiparametric model for regression analysis of interval-censored failure time data. *Biometrics* 41: 933–945.

Genç, A.I. (2011). Estimation of $P(X > Y)$ with Topp-Leone distribution. *Journal of Statistical Computation and Simulation*, forthcoming.

Greco, L. and Ventura, L. (2011). Robust inference for the stress-strength reliability. *Statistical Papers* 52: 773–788.

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. and Peng, C. (2009). Estimating reliability in proportional odds ratio models. *Computational Statistics & Data Analysis* 53: 1495–1510.

Heinz G., Peterson, L. J., Johnson, R. W. and Kerk, C. J. (2003). Exploring relationships in body dimensions. *Journal of Statistics Education* (online only) 11(2). www.amstat.org/publications/jse/v11n2/datasets.heinz.html.

Heitjan, D. F. (1989). Inference from grouped continuous data: A review. *Statistical Science* 4, 164–183.

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

Ley, C. and Paindaveine, D. (2010). On the singularity of multivariate skew-symmetric models. *Journal of Multivariate Analysis* 101: 1434–1444.

Liseo, B. and Loperfido, N. (2006). A note on reference priors for the scalar skew-normal distribution. *Journal of Statistical Planning and Inference* 136: 373-389.

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

Mudholkar, G. S. and Hutson, A. D. (2000). The epsilon-skew-normal distribution for analyzing near-normal data. *Journal of Statistical Planning and Inference* 83: 291–309.

Nadarajah, S. (2005). Reliability for some bivariate gamma distributions. *Mathematical Problems in Engineering* 2005: 151–163.

Nandi, S. B. and Aich, A. B. (1994). A note on confidence bounds for $P(X > Y)$ in bivariate normal samples. *Sankhyā, Ser. B* 56: 129–136.

Rubio, F. J. and Steel M. F. J. (2011). Inference in two-piece location-scale models with Jeffreys priors. *CRiSM* working paper 11-13, University of Warwick.

Shoukri, M. M., Chaudary, M. A. and Al-Halees, A. (2005). Estimating $P(Y < X)$ when X and Y are paired exponential variables. *Journal of Statistical Computation and Simulation* 75: 25-38.

Sun, D., Ghosh, M. and Basu, A. P. (1998). Bayesian analysis for a stress–strength system under noninformative priors. *The Canadian Journal of Statistics* 26: 323–332.

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 $P(X < Y)$. *Bayesian Analysis* 6: 118.

Wang, J., Boyer, J. and Genton M. C. (2004). A skew symmetric representation of multivariate distributions. *Statistica Sinica* 14: 1259–1270.

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