A Note on Learning Dependence under Severe Uncertainty

Matthias C.M. Troffaes[1], Frank P.A. Coolen[1], and Sébastien Destercke[2]

[1] Durham University, UK
[2] Université de Technologie de Compiègne, France

Abstract. We propose two models, one continuous and one categorical, to learn about dependence between two random variables, given only limited joint observations, but assuming that the marginals are precisely known. The continuous model focuses on the Gaussian case, while the categorical model is generic. We illustrate the resulting statistical inferences on a simple example concerning the body mass index. Both methods can be extended easily to three or more random variables.

Keywords: bivariate data, categorical data, copula, Gaussian copula, robust Bayesian, imprecise probability.

1 Introduction

Sklar's theorem [10] states that any multivariate distribution of k variables can be expressed through a density on $[0,1]^k$ with uniform marginals—this density is called a copula—and the marginal distributions of each of the variables. For this reason, copulas [7] have become an indispensible tool to model and learn statistical dependence in multivariate models: they allow of estimation of the dependence structure, separately from the marginal structure.

Estimating dependence requires joint observations, which in many cases are only available in small amounts, while substantial amounts of marginal data may be available. For example, when studying the reliability of a system, it is common to have good information about the reliability of each system component, yet to have only little information about joint failures [11]. Imprecise probabilities provide one possible theoretical basis for dealing with small sample sizes, by representing knowledge as a *set* of distributions [3,13,1], rather than a single distribution necessarily based on somewhat arbitrary assumptions [6].

Copulas and imprecise probabilities have been studied in the literature by various researchers. The Fréchet–Hoeffding copula bounds, which represent completely unknown dependence, are used for instance in probabilistic arithmetic [16] and p-boxes [4,12]. One theoretical difficulty is that there is no straightforward imprecise equivalent of Sklar's theorem, say, expressing any set of joint distributions as a sets of copulas along with a set of marginal distributions [9]: it appears that, when working with sets of distributions, separating the dependence structure from the marginal structure is a lot more difficult in general.

A. Laurent et al. (Eds.): IPMU 2014, Part III, CCIS 444, pp. 498–507, 2014.

In this paper, we propose and investigate a few statistical models for robust dependence learning from limited data. We state an imprecise version of Sklar's theorem when marginal distributions are fully known, separating precise marginals from the imprecise dependence structure. We propose a range of parametric models for the bivariate categorical case. Finally, we demonstrate our findings on a toy example: estimating the body mass index from height and mass data.

Section 2 explores a continuous model, focusing on the multivariate normal model, while section 3 provides a first exploration of a generic categorical model.

2 Robust Bayesian Correlation Learning for Bivariate Normal Sampling

We start with revising a simple and well-studied model: sampling from the bivariate normal distribution. We will derive some new results that are relevant to dependence learning. Our analysis starts from Quaeghebeur and De Cooman's robust Bayesian framework for sampling from the exponential family [8].

2.1 Inference with Known Mean and Unknown Covariance Matrix

Let $Z_i \coloneqq (Z_{i1}, \ldots, Z_{ik})$ be a multivariate normally distributed random variable with known mean—which we can assume to be zero without loss of generality through translation of the data—but unknown covariance matrix $\Sigma \in \mathbb{R}^{k\times k}$. A particular realisation of Z_i is denoted by a lower case letter $z_i \coloneqq (z_{i1}, \ldots, z_{ik}) \in \mathbb{R}^k$. The likelihood of an i.i.d. sample $z_1, \ldots, z_n$ is

$$f(z_1, \ldots, z_n \mid \Sigma) \propto |\Sigma|^{-n/2} \prod_{i=1}^{n} \exp\left(-\frac{1}{2} z_i^T \Sigma^{-1} z_i\right) \tag{1}$$

$$= |\Sigma|^{-n/2} \exp\left[-\frac{1}{2} \sum_{i=1}^{n} \operatorname{tr}\left(z_i z_i^T \Sigma^{-1}\right)\right], \tag{2}$$

where the data $z_i \in \mathbb{R}^k$ are organised as row vectors, so $z_i z_i^T$ is the matrix containing $z_{i\ell} z_{i\ell'}$ in row ℓ and column ℓ'.

A family of conjugate priors for this density is the family of inverse Wishart distributions with hyperparameters $\nu_0 > 0$ and $\Psi_0 \in \mathbb{R}^{k\times k}$ positive definite [2]:

$$f(\Sigma \mid \nu_0, \Psi_0) \propto |\Sigma|^{-\frac{\nu_0+k+1}{2}} \exp\left[-\frac{1}{2} \operatorname{tr}\left(\Psi_0 \Sigma^{-1}\right)\right]. \tag{3}$$

The posterior distribution is obtained by updating the hyperparameters through

$$\nu_n = \nu_0 + n, \qquad \Psi_n = \Psi_0 + \sum_{i=1}^{n} z_i z_i^T. \tag{4}$$

The prior expected covariance matrix is given by

$$\mathrm{E}\left(\Sigma \mid \nu_0, \Psi_0\right) = \frac{\Psi_0}{\nu_0 - k - 1} =: \Sigma_0 \tag{5}$$

and therefore, through conjugacy, the posterior expected covariance matrix is

$$\begin{aligned}\mathrm{E}\left(\Sigma \mid z_1, \ldots, z_n, \nu_0, \Psi_0\right) &= \mathrm{E}\left(\Sigma \mid \nu_n, \Psi_n\right) = \frac{\Psi_n}{\nu_n - k - 1} &(6)\\ &= \frac{\Psi_0 + \sum_{i=1}^n z_i z_i^T}{\nu_0 + n - k - 1} &(7)\\ &= \frac{(\nu_0 - k - 1)\Sigma_0 + \sum_{i=1}^n z_i z_i^T}{\nu_0 + n - k - 1} =: \Sigma_n. &(8)\end{aligned}$$

For robust Bayesian analysis aiming to learn about the dependence between two random variables, we now need to identify a reasonable set of prior distributions, or, in our conjugate setting, a reasonable set of hyperparameters ν_0 and Ψ_0.

The formula for the posterior expected covariance matrix shows that ν_0 determines our learning speed, that is, how many observations n we need before starting to move towards our data. So, ν_0 is similar to the s value in the imprecise Dirichlet model [14]. Here too, we will simply assume ν_0 to be fixed to whatever value is judged to lead to a reasonable learning speed. For fixed ν_0, any particular choice of Ψ_0 corresponds to a prior covariance matrix Σ_0.

Let us now study the bivariate case ($k = 2$) in more detail. We will write X_i for Z_{i1} and Y_i for Z_{i2}. We would choose

$$\Psi_0 = \nu_0' \begin{bmatrix} \sigma_X^2 & \rho_0 \sigma_X \sigma_Y \\ \rho_0 \sigma_X \sigma_Y & \sigma_Y^2 \end{bmatrix} \tag{9}$$

where $\nu_0' = \nu_0 - k - 1 = \nu_0 - 3$, if we had prior standard deviations $\sigma_X > 0$ and $\sigma_Y > 0$ for the two components as well as the prior correlation coefficient $\rho_0 \in [-1, 1]$. For this paper focusing on dependence, we are mostly interested in cases where the marginals are well known, i.e. well known prior σ_X and σ_Y, but unknown prior correlation ρ_0. We will therefore study the set of priors with all parameters fixed, except for ρ_0, which we assume to be vacuous a priori. Without loss of generality, by rescaling, we can assume that $\sigma_X = \sigma_Y = 1$, leaving us with just two hyperparameters: $\nu_0 > 0$ and $\rho_0 \in [-1, 1]$.

The posterior covariance matrix becomes

$$\Sigma_n = \frac{1}{\nu_0' + n} \begin{bmatrix} \nu_0' + \sum_{i=1}^n x_i^2 & \nu_0' \rho_0 + \sum_{i=1}^n x_i y_i \\ \nu_0' \rho_0 + \sum_{i=1}^n x_i y_i & \nu_0' + \sum_{i=1}^n y_i^2 \end{bmatrix}. \tag{10}$$

Provided that the sample variance is approximately equal to the prior variance, i.e.

$$\sum_{i=1}^n x_i^2 \approx n\sigma_X^2 = n, \qquad \sum_{i=1}^n y_i^2 \approx n\sigma_Y^2 = n, \tag{11}$$

our expression for Σ_n becomes

$$\begin{bmatrix} 1 & \rho_n \\ \rho_n & 1 \end{bmatrix}, \tag{12}$$

where

$$\rho_n = \frac{\nu_0'\rho_0 + \sum_{i=1}^n x_i y_i}{\nu_0' + n}. \tag{13}$$

Equation (12) is the covariance matrix of a bivariate normal with unit marginal variances and correlation coefficient ρ_n. For vacuous prior correlation, $\rho_0 \in [-1, 1]$, we thus get the following posterior bounds on the correlation:

$$\underline{\rho}_n = \frac{-\nu_0' + \sum_{i=1}^n x_i y_i}{\nu_0' + n}, \qquad \overline{\rho}_n = \frac{\nu_0' + \sum_{i=1}^n x_i y_i}{\nu_0' + n}, \tag{14}$$

provided that our observations, and our prior, have unit variance and zero mean (which we can achieve by linear transformation without loss of generality).

This analysis generalises easily to cases with more than two variables, $k > 2$—we leave this to the reader. Essentially, we simply have to deal with more correlation parameters.

2.2 Application to the Body Mass Index Example

We now illustrate our model on the evaluation of the body mass index $R = X/Y^2$, where X is a person's weight in kilograms and Y is his or her height in meters. The body mass index is commonly used to detect under- and overweight. We aim (i) to assess the dependence between X and Y in a particular population, and (ii) to extract a robust inference about R in this population.

We consider 30 paired observations of heights and weights of girls aged 11 [5, p. 75]. The weight X has sample mean $\bar{x} = 36.2$, sample standard deviation $s_X = 7.7$, with no strong evidence against normality (p-value[1] 0.017). The height Y has sample mean $\bar{y} = 1.448$, sample standard deviation $s_Y = 0.077$, with no evidence against normality whatsoever (p-value 0.711). We will assume that X and Y have known means, equal to $\bar{x}$ and $\bar{y}$. We also assume that, a priori, $\sigma_X = 7.7$ and $\sigma_Y = 0.077$ in eq. (9), but we are vacuous about the prior correlation. For reference, it may be useful to note that the sample correlation between X and Y is in fact 0.742. For the sake of the example, we assume that the sample is drawn from a bivariate normal distribution, although there is reasonably strong evidence against joint normality (p-value 0.00388).

Figure 1 shows the bounds on the correlation of the posterior covariance matrix in eq. (10) with $\nu_0' = 2$ and $\rho_0 \in [-1, 1]$. The two values converge steadily with a final interval $[\underline{\rho}_{30}, \overline{\rho}_{30}] = [0.630, 0.759]$. The expectation of R is bounded by $\underline{E}(R) = 17.10$ and $\overline{E}(R) = 17.16$. Similarly, we may wonder about the probability of R to be in a "healthy" range, which is about $A = [14, 19.5]$ for girls aged 11. We obtain bounds $\underline{P}(R \in A) = 0.66$ and $\overline{P}(R \in A) = 0.71$. Note that bounds were obtained by straightforward numerical optimisation.

[1] Throughout, we test for normality using the Shapiro-Wilk test.

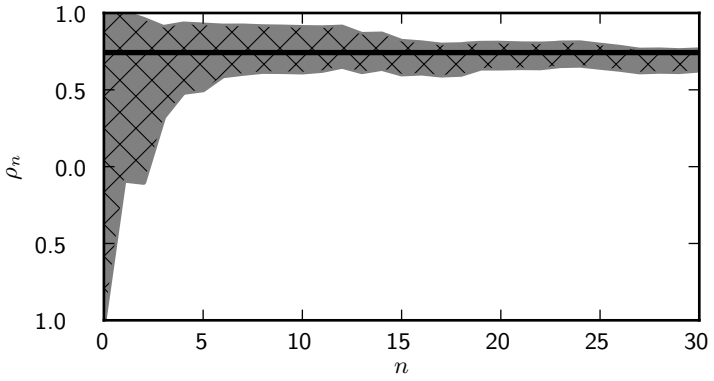

Fig. 1. Lower and upper correlation estimates $\underline{\rho}_n$ and $\overline{\rho}_n$ as a function of the sample size n, with $\nu_0' = 2$. The solid horizontal line denotes the sample correlation for $n = 30$.

3 Robust Bayesian Dependence Learning for Bivariate Categorical Data

3.1 The Model

Consider a bivariate categorical random quantity $Z := (X, Y)$ with X taking values in a finite set $\mathcal{X} = \{1, \ldots, m_X\}$, and Y taking values in a finite set $\mathcal{Y} = \{1, \ldots, m_Y\}$. The parameters θ_x and ϕ_y determine the marginal distributions:

$$p(x \mid \theta) = \theta_x, \qquad p(y \mid \phi) = \phi_y. \tag{15}$$

We assume that $m_X \geq 2$, $m_Y \geq 2$, $\theta_x > 0$ and $\phi_y > 0$.

We are interested in learning the dependence structure of X and Y. One very general way to express the full joint distribution of (X, Y) is by introducing parameters w_{xy} such that

$$p(x, y \mid \theta, \phi, w) = w_{xy}\theta_x\phi_y, \tag{16}$$

subject to the constraints

$$\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} w_{xy}\theta_x\phi_y = 1, \tag{17}$$

$$\sum_{x \in \mathcal{X}} w_{xy}\theta_x = 1, \tag{18}$$

$$\sum_{y \in \mathcal{Y}} w_{xy}\phi_y = 1. \tag{19}$$

Equations (18) and (19) simply follow from

$$\sum_{x\in\mathcal{X}} w_{xy}\theta_x\phi_y = \sum_{x\in\mathcal{X}} p\left(x, y \mid \theta, \phi, w\right) = p\left(y \mid \theta, \phi, w\right) = \phi_y, \tag{20}$$

$$\sum_{y\in\mathcal{Y}} w_{xy}\theta_x\phi_y = \sum_{y\in\mathcal{Y}} p\left(x, y \mid \theta, \phi, w\right) = p\left(x \mid \theta, \phi, w\right) = \theta_x, \tag{21}$$

respectively. This model specification is over-parametrized, but it allows us to model the marginal distributions and the dependence structure separately, where the matrix w plays precisely a similar role as a copula in general bivariate models for (usually) continuous random quantities. However, a key difference and major difficulty with the above model is that the constraints on w depend on θ and ϕ: the separation is thus not as complete as with copulas, where the dependence structure is parametrised independently of the marginals. For this reason, it seems most natural to consider a two-stage situation where we first learn about the marginal parameters θ and ϕ, followed by learning about the dependence structure w *conditional on what we learnt about* θ *and* ϕ.

3.2 Inference for Known Marginals

For this reason, as a stepping stone towards general inference about θ, ϕ, and w, here we consider a scenario where the marginal distributions are already fully known, and we only aim at inference about w. While this may appear somewhat restrictive, and perhaps even artificial, there are practical scenarios where one has very substantial information about the probability distributions for the random quantities X and Y separately, but relatively little information about their joint distribution.

There are $(m_X - 1)(m_Y - 1)$ degrees of freedom for the components of w. In case data is limited, to enable sufficiently useful inference, it seems natural to assume a reduced-dimensional parametric form for w, which may naturally correspond to an (assumed) ordering of the categories, as we will illustrate in section 3.3. Let n_{xy} denote the number of observations of $(X, Y) = (x, y)$, with total number of observations $n = \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} n_{xy}$ and row and column totals denoted by $n_{x*} = \sum_{y\in\mathcal{Y}} n_{xy}$ and $n_{*y} = \sum_{x\in\mathcal{X}} n_{xy}$, respectively. So, there are n_{x*} observations of $X = x$ and n_{*y} observations of $Y = y$.

Without further restrictions on w, it seems tempting to fit the model to match the non-parametric maximum likelihood estimate

$$\hat{p}\left(x, y \mid w\right) = \frac{n_{xy}}{n} \tag{22}$$

by setting

$$\hat{w}_{xy} = \frac{n_{xy}}{n\theta_x\phi_y}. \tag{23}$$

A problem is that this estimate will usually violate eqs. (18) and (19). For instance,

$$\sum_{x\in\mathcal{X}} \hat{w}_{xy}\theta_x = \sum_{x\in\mathcal{X}} \frac{n_{xy}}{n\theta_x\phi_y}\theta_x = \frac{n_{*y}}{n\phi_y} \neq 1 \tag{24}$$

as soon as $\frac{n_{*y}}{n} \neq \phi_y$. A proper maximum likelihood estimate would maximize the likelihood subject to all constraints embodied by eqs. (17) to (19). Solving this optimisation problem poses an interesting challenge.

Bayesian inference for w will face a similar challenge: w lives in a convex subspace of $\mathbb{R}^{m_X \times m_Y}$ determined by eqs. (17) to (19). Application of Bayes's theorem requires numerical integration over this space. Nevertheless, the basic principles behind Bayesian inference for w are simple, and sensitivity analysis is similar to the imprecise Dirichlet model [14]. Specific dimension-reduced models, where we have a much better handle on the parameter space, will be illustrated in more detail in section 3.3.

The likelihood is given by

$$\prod_{x \in \mathcal{X}} \prod_{y \in \mathcal{Y}} (w_{xy} \theta_x \phi_y)^{n_{xy}}, \tag{25}$$

so as conjugate prior we can choose

$$f(w \mid \alpha_0) \propto g(w) \prod_{x \in \mathcal{X}} \prod_{y \in \mathcal{Y}} (w_{xy} \theta_x \phi_y)^{\alpha_{0xy}}, \tag{26}$$

where $\alpha_{0xy} > 0$ and g is some arbitrary non-negative function (as long as the right hand side integrates to a finite value). With $\nu_0 := \sum_{xy} \alpha_{0xy}$, this prior distribution can be interpreted as reflecting prior information equivalent to ν_0 observations of which α_{0xy} were $(X, Y) = (x, y)$. The corresponding posterior distribution is clearly $f(w \mid \alpha_n)$ with $\alpha_{nxy} = \alpha_{0xy} + n_{xy}$.

Sensitivity analysis on this model could then follow an approach similar to Walley's imprecise Dirichlet model [14], by taking the set of all prior distributions for a fixed value of ν_0. In case of an informative set of prior distributions, one may also allow the value of ν_0 to vary within a set to allow prior-data conflict modelling [15].

As already mentioned, the remaining key difficulty is to integrate the conjugate density over the parameter space. For this reason, in the next section, we consider a reduced model.

3.3 Reduced Model

As a first and basic example of a reduced parametric form, consider the case $\mathcal{X} = \mathcal{Y} = \{1, 2, 3\}$ with known $\theta_x = \phi_y = 1/3$ for all x and $y \in \{1, 2, 3\}$. If the categories are pairwise ordered in some natural manner, then it might be quite reasonable to specify

$$w = \begin{bmatrix} 1+2\alpha & 1-\alpha & 1-\alpha \\ 1-\alpha & 1+2\alpha & 1-\alpha \\ 1-\alpha & 1-\alpha & 1+2\alpha \end{bmatrix} \tag{27}$$

with $\alpha \in [0, 1]$. It is easily verified that this model satisfies eqs. (17) to (19): the full matrix sums to 9, and each of the rows and colums sum to 3.

Note that there is no logical requirement to avoid $\alpha \in [-1/2, 0)$, we just use this small example as an illustration. In this model, $\alpha = 0$ corresponds to full independence between X and Y, whereas $\alpha = 1$ corresponds to perfect correlation $X = Y$. Therefore, this corresponds to a scenario where we may suspect positive correlation between X and Y, but we are unsure about the strength of correlation. Note that the actual model reduction is achieved by additionally assuming that $X = Y = x$ has the same probability for all $x \in \{1, 2, 3\}$, and similar for $X = x \cap Y = y$ for all $x \neq y$.

With these assumptions, statistical inference is concerned with learning about the parameter $\alpha \in [0, 1]$. The likelihood function is

$$(1 + 2\alpha)^t (1 - \alpha)^{n-t} \tag{28}$$

with $t = n_{11} + n_{22} + n_{33}$ and $n = \sum_{xy} n_{xy}$ as before. The maximum likelihood estimate is

$$\hat{\alpha} = \begin{cases} \frac{3t-n}{2n} & \text{if } 3t \geq n, \\ 0 & \text{otherwise.} \end{cases} \tag{29}$$

For a Bayesian approach to inference for this model, we can define a conjugate prior

$$f(\alpha \mid \nu_0, \tau_0) \propto (1 + 2\alpha)^{\tau_0} (1 - \alpha)^{\nu_0 - \tau_0} \tag{30}$$

with $\tau_0 \in [0, \nu_0]$, with the possible interpretation that it reflects prior information which is equivalent to ν_0 observations of which τ_0 have $X = Y$.

The posterior distribution is simply $f(\alpha \mid \nu_0 + n, \tau_0 + t)$. Sensitivity analysis is again straightforward by taking the set of prior distributions for $\tau_0 \in [0, \nu_0]$ and a fixed ν_0. For instance, we would get the following robust estimate for the posterior mode of α:

$$\hat{\alpha}_n = \left[\frac{3t - \nu_0 - n}{2(\nu_0 + n)}, \frac{3t + 2\nu_0 - n}{2(\nu_0 + n)} \right] \tag{31}$$

when $3t \geq \nu_0 + n$, with similar formulas when $3t < \nu_0 + n$ (truncating negative values to zero).

3.4 Application to the Body Mass Index Example

To apply the categorical model to our data, we must first discretize them, with the ordering of the categories following the ordering of natural numbers. To obtain three categories with uniform marginals, we simply discretized the 99% prediction intervals of each Gaussian marginals of section 2.2, obtaining $\mathcal{X} = \{[17, 32], [32, 39], [39, 56]\}$ and $\mathcal{Y} = \{[1.24, 1.41], [1.41, 1.47], [1.47, 1.64]\}$.

Figure 2 shows the bounds on the posterior mode $\hat{\alpha}_n$ in eq. (31), with $\nu_0 = 2$. The results are similar to those obtained in section 2.2, showing that even this very simple discretized model can capture the correlation between X and Y, with the bounds on $\hat{\alpha}_{30}$ being $[0.56, 0.66]$. From these values and the bounds of the categories, we can easily obtain bounds on the expectation of R: $\underline{E}(R) = 12.9$

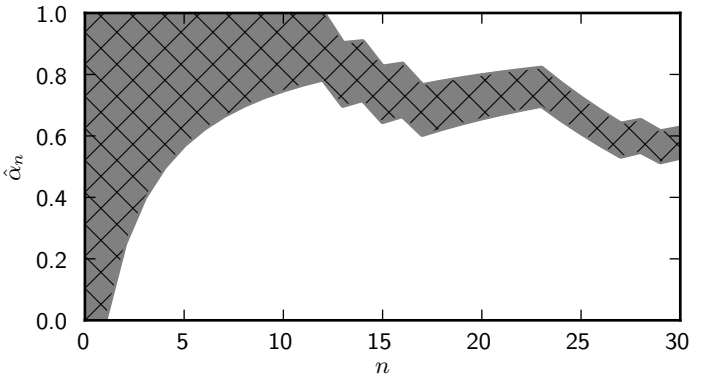

Fig. 2. Bounds on $\hat{\alpha}_n$ as a function of the sample size n, with $\nu_0 = 2$

and $\overline{E}(R) = 22.4$, which gives a much wider interval than in section 2.2. This is due to the very rough discretization: a finer discretization would likely provide a narrower interval. The lower and upper probabilities of $A = [14, 19.5]$ are this time $\underline{P}(R \in A) = 0$ and $\overline{P}(R \in A) = 0.96$, which are almost vacuous and again show that to have meaningful inferences, a finer discretization is needed.

4 Conclusion

In this paper, we have introduced two preliminary models—a continuous one and a discrete one—to model dependence when joint data is limited, but assuming that the marginals are precisely known. The continuous model focused on the very special multivariate normal case. However, already in our simple example, we have seen that joint normality is rarely satisfied in practice. A major challenge is to provide methods for dependence modelling, that are both flexible and computationally tractable, whilst still producing useful inferences.

Even though the models and example studied are very preliminary, we feel that extensions of the discrete model could provide more flexibility, whilst still being easy to learn and to compute with. We see it as a promising path to learn dependency structures with imprecise probabilistic models. In particular, it can be seen as a way to approximate a continuous model, as we did in the example. In the future we plan to work on such extensions and on the identification of parametric matrices of weights more flexible than the reduced one presented here.

Finally, an obvious extension to the present work would be to relax the assumption that marginals are precisely identified, and to work with sets of marginals instead. However, this raises challenging theoretical issues, as defining a well-founded extension or equivalent formulation of Sklar's theorem for imprecise models is far from trivial.

References

1. Augustin, T., Coolen, F.P.A., de Cooman, G., Troffaes, M.C.M.: Introduction to imprecise probabilities. Edited book (submitted to publisher)
2. Bernado, J.M., Smith, A.F.M.: Bayesian Theory. John Wiley and Sons (1994)
3. Boole, G.: An investigation of the laws of thought on which are founded the mathematical theories of logic and probabilities. Walton and Maberly, London (1854)
4. Ferson, S., Kreinovich, V., Ginzburg, L., Myers, D.S., Sentz, K.: Constructing probability boxes and Dempster-Shafer structures. Technical Report SAND2002–4015, Sandia National Laboratories (January 2003)
5. Hand, D.J., Daly, F., McConway, K., Lunn, D., Ostrowski, E.: A handbook of small data sets. CRC Press (1993)
6. Kass, R.E., Wasserman, L.: The selection of prior distributions by formal rules. Journal of the American Statistical Association 91(435), 1343–1370 (1996)
7. Nelsen, R.B.: An introduction to copulas. Springer (1999)
8. Quaeghebeur, E., de Cooman, G.: Imprecise probability models for inference in exponential families. In: Cozman, F.G., Nau, R., Seidenfeld, T. (eds.) ISIPTA 2005: Proceedings of the Fourth International Symposium on Imprecise Probabilities and Their Applications, Pittsburgh, USA, pp. 287–296 (July 2005)
9. Pelessoni, R., Vicig, P., Montes, I., Miranda, E.: Imprecise copulas and bivariate stochastic orders. In: De Baets, B., Fodor, J., Montes, S. (eds.) Proceedings of Eurofuse 2013 Workshop, pp. 217–225 (2013)
10. Sklar, A.: Fonctions de répartition à n dimensions et leurs marges. Publ. Inst. Statist. Univ. Paris 8, 229–231 (1959)
11. Troffaes, M.C.M., Blake, S.: A robust data driven approach to quantifying common-cause failure in power networks. In: Cozman, F., Denœux, T., Destercke, S., Seidenfeld, T. (eds.) ISIPTA 2013: Proceedings of the Eighth International Symposium on Imprecise Probability: Theories and Applications, Compiègne, France, pp. 311–317. SIPTA (July 2013)
12. Troffaes, M.C.M., Destercke, S.: Probability boxes on totally preordered spaces for multivariate modelling. International Journal of Approximate Reasoning 52(6), 767–791 (2011)
13. Walley, P.: Statistical Reasoning with Imprecise Probabilities. Chapman and Hall, London (1991)
14. Walley, P.: Inferences from multinomial data: Learning about a bag of marbles. Journal of the Royal Statistical Society, Series B 58(1), 3–34 (1996)
15. Walter, G., Augustin, T.: Imprecision and prior-data conflict in generalized Bayesian inference. Journal of Statistical Theory and Practice 3, 255–271 (2009)
16. Williamson, R.C., Downs, T.: Probabilistic arithmetic I: Numerical methods for calculating convolutions and dependency bounds. International Journal of Approximate Reasoning 4, 89–158 (1990)