Bivariate P–Boxes

Enrique Miranda, Ignacio Montes
University of Oviedo, Spain
{mirandaenrique, imontes}@uniovi.es

Renato Pelessoni, Paolo Vicig
University of Trieste, Italy
{renato.pelessoni, paolo.vicig}@econ.units.it

Abstract

Given a random number X, a probability box or *p–box* $(\underline{F}_X, \overline{F}_X)$ is a couple of cumulative distribution functions (cdfs) s.t. $\underline{F}_X \leq \overline{F}_X$ [1, 4]. Here and in what follows, we impose no continuity property on any cdf, which is therefore a dF-coherent probability (a finitely, not necessarily σ-additive precise probability) on the monotone family of events $D_1 = \{A_x | x \in \mathbb{R}\} \cup \{\varnothing, \Omega\}, A_x = (X \leq x), \forall x \in \mathbb{R}$. A p–box therefore naturally extends to an imprecise probability framework the description of uncertainty about X by means of a cdf.

In this note we investigate properties of the generalisation of p–boxes, suited to describe couples (X, Y) of random numbers and to be called bivariate p–boxes. We focus on analogies between bivariate p–boxes and traditional joint distribution functions, and on how bivariate p-boxes may be obtained from marginal uncertainty judgements.

Definitions. Given (X, Y), let $A_{x,y} = (X \leq x \wedge Y \leq y)$. A map $F : D_2 = \{A_{x,y} : x, y \in \mathbb{R}\} \cup \{\varnothing, \Omega\} \to [0; 1]$ is *standardized* if F is non–negative, componentwise non–decreasing, $F(\varnothing) = 0$, $F(\Omega) = 1$. Later on, we shall also write $F(x, y)$ instead of $F(A_{x,y})$. $(\underline{F}, \overline{F})$ is a *bivariate p–box* if each of $\underline{F}$, $\overline{F}$ is standardized and $\underline{F} \leq \overline{F}$. $(\underline{F}, \overline{F})$ is a *coherent* p–box (a p–box that avoids sure loss (ASL)) iff, further, both $\underline{F}$ and $\overline{F}$ are *jointly* coherent (ASL) [5], lower and upper respectively, probabilities on D_2. We say that $\underline{F}$, $\overline{F}$ are jointly coherent (ASL) when the lower probability $\underline{P}$ defined as $\underline{P}(A_{x,y}) = \underline{F}(x, y)$ on $S = \{A_{x,y} | x, y \in \mathbb{R}\}$, $\underline{P}(A^c_{x,y}) = 1 - \overline{F}(x, y)$ on $S^- = \{A^c_{x,y} | x, y \in \mathbb{R}\}$ is coherent (ASL) on $S \cup S^-$.

A first major difference between coherent bivariate and univariate p–boxes is that $\underline{F}$, $\overline{F}$ need not be dF-coherent precise probabilities. This clearly depends on the structure of D_2, an only partially ordered set unlike D_1, but there are relationships with 2–monotonicity too:

Proposition 1 *Let $\underline{P}$ be a 2–monotone lower probability on some lattice $L \supset D_2$, and $\overline{P}$ its conjugate (hence, 2–alternating) upper probability.*

a) If $\underline{F}$ is the restriction of $\underline{P}$, $\underline{F}$ is dF-coherent [3].

b) If $\overline{F}$ is the restriction of $\overline{P}$, it is not necessarily dF-coherent, while its corresponding upper tail function is.

c) Conversely, if $(\underline{F}, \overline{F})$ is given and $\underline{F}$, $\overline{F}$ are jointly dF-coherent, the natural extension of $(\underline{F}, \overline{F})$ is not necessarily 2–monotone.

As well-known, a joint cdf F is characterised by some conditions, including a *rectangle inequality* $F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1) \geq 0, \forall x_1 \leq x_2, y_1 \leq y_2$. With a p-box $(\underline{F}, \overline{F})$, we have four rectangle inequalities:

[R1] $\underline{F}(x_2, y_2) - \underline{F}(x_1, y_2) - \underline{F}(x_2, y_1) + \overline{F}(x_1, y_1) \geq 0$

[R2] $\overline{F}(x_2, y_2) - \underline{F}(x_1, y_2) - \underline{F}(x_2, y_1) + \underline{F}(x_1, y_1) \geq 0$

[R3] $\overline{F}(x_2, y_2) - \underline{F}(x_1, y_2) - \overline{F}(x_2, y_1) + \overline{F}(x_1, y_1) \geq 0$

[R4] $\overline{F}(x_2, y_2) - \overline{F}(x_1, y_2) - \underline{F}(x_2, y_1) + \overline{F}(x_1, y_1) \geq 0.$

These inequalities interact variously with coherence or ASL of either a p–box $(\underline{F}, \overline{F})$ or its components $\underline{F}$, $\overline{F}$, taken separately:

Proposition 2 *a) [R1]÷[R4] are necessary for coherence of* $(\underline{F}, \overline{F})$.

b) Neither of them is, in general, necessary for ASL of $(\underline{F}, \overline{F})$*;* $\underline{F}$ *(being standardized) always avoids sure loss, while* $\overline{F}$ *avoids sure loss if [R2] holds.*

c) In the case that X, Y *are both two–valued, [R1]÷[R4] are also sufficient for coherence of* $(\underline{F}, \overline{F})$*, while [R1] is necessary and sufficient for* $(\underline{F}, \overline{F})$ *to be ASL.*

An important situation originating bivariate p–boxes is when marginal cdfs for X and Y are given, and there is uncertainty about the kind of interaction between X and Y. More generally, we may think that marginal p-boxes $(\underline{F}_X, \overline{F}_X)$, $(\underline{F}_Y, \overline{F}_Y)$ are assessed for X and Y. Then, under these assumptions,

Proposition 3 *Let* $\mathcal{C}$ *be a set of copulas. Define the bivariate p–box* $(\underline{F}, \overline{F})$ *as* $\underline{F}(x, y) = \inf_{C \in \mathcal{C}} C(\underline{F}_X(x), \underline{F}_Y(y)), \overline{F}(x, y) = \sup_{C \in \mathcal{C}} C(\overline{F}_X(x), \overline{F}_Y(y))$*. Then* $(\underline{F}, \overline{F})$ *is coherent.*

While the above proposition may be viewed as a sort of imprecise counterpart of Sklar's Theorem [2], in the part ensuring that a certain function (copula) of two univariate cdfs returns a joint distribution having the given cdfs as marginals, it has to be stated that the correspondence breaks down on the reverse side, when wishing to view any bivariate p–box as depending on its arguments through a function (not necessarily a copula or subcopula) of its marginals. This is in general not possible, outside some special cases.

Fréchet upper and lower bounds also play a very important role in obtaining joint p–boxes from marginal ones, even in the n–variate case. In fact,

Proposition 4 *a) Given* $F_1, F_2, \ldots, F_n$ *(marginal cdfs, for* $X_1, X_2, \ldots, X_n$ *respectively), the lower Fréchet bound* $\underline{F}^L(x_1, x_2, \ldots, x_n) = \max(F_1(x_1) + F_1(x_2) + \ldots + F_n(x_n)) - n + 1, 0)$ *is a coherent lower probability (also dF-coherent, as well–known [2], when* $n = 2$*).*

b) Given the n *marginal p-boxes* $(\underline{F}_1, \overline{F}_1), \ldots, (\underline{F}_n, \overline{F}_n)$*, their natural extension on* $D_n = \{X_1 \leq x_1 \wedge \ldots \wedge X_n \leq x_n | x_1, \ldots, x_n \in \mathbb{R}\} \cup \{\varnothing, \Omega\}$ *is the n-dimensional p–box* $(\underline{F}^L, \overline{F}^U)$*, where* $\underline{F}^L(x_1, x_2, \ldots, x_n) = \max(\underline{F}_1(x_1) + \underline{F}_2(x_2) + \ldots + \underline{F}_n(x_n)) - n + 1, 0)$*, while* $\overline{F}^U(x_1, x_2, \ldots, x_n) = \min(\overline{F}_1(x_1), \overline{F}_2(x_2), \ldots, \overline{F}_n(x_n))$ *is the Fréchet upper bound (which is dF–coherent,* $\forall n$*).*

Keywords. P–boxes, coherent lower/upper probabilities, rectangle inequalities, copulas, Fréchet bounds.

References

[1] S. Ferson, V. Kreinovich, L. Ginzburg, D.S. Myers and K. Sentz. *Constructing probability boxes and Dempster-Shafer structures.* Technical Report SAND2002-4015, Sandia National Laboratories, 2003.

[2] R. B. Nelsen. *An Introduction to Copulas.* Springer, 2006.

[3] M. Scarsini. Copulae of Capacities on Product Spaces. *Distributions with Fixed Marginals and Related Topics.* IMS Lecture Notes, Institute of Mathematical Statistics, 28:307–318, 1996.

[4] M. C. M. Troffaes and S.Destercke. Probability boxes on totally preordered spaces for multivariate modelling. *International Journal of Approximate Reasoning*, 52(6):767-791, 2011.

[5] P. Walley. *Statistical Reasoning with Imprecise Probabilities.* Chapman and Hall, London, 1991