juliacon

ProbabilityBoundsAnalysis.jl: theory and algorithms

Ander Gray[1], Scott Ferson[1], and Edoardo Patelli[1, 2]

[1]Institute for Risk and Uncertainty, University of Liverpool
[2]Centre for Intelligent Infrastructure, University of Strathclyde

ABSTRACT

Probability bounds analysis combines interval arithmetic with probability theory, and provides a representation of sets of distributions in structures called probability boxes (p-boxes). These which generalise both distribution functions and intervals. P-boxes generally return interval bounds on all probabilistic quantities, for example samples, cdfs, and probability measures are all intervals. This framework also allows for the comprehensive propagation of probabilities through calculations in a rigorous way, in a similar fashion that interval arithmetic does for sets. As such, `ProbabilityBoundsAnalysis.jl` gives a rigorous arithmetic of random variables, where both marginal (univariate) distributions and dependency information can be known, partially known or missing completely. We describe the main theoretical elements of probability bounds analysis, and provide a simplified implementation of the method in code snippets which can be readily evaluated in the Julia command terminal.

Keywords

Julia, Probability, Uncertainty, Interval Arithmetic, Probabilistic Arithmetic

1. Introduction

An arithmetic of probability distributions has held a long interest among mathematicians and scientists:

> A question asked of Kolmogorov,
> answered for the sum by Makarov,
> partially answered by Sklar and Frank (for which the copula was invented),
> made algorithmically available by Williamson,
> and generalised by others.

Indeed it was Kolmogorov who originally asked about the result of an operation between two random variables without knowing their joint distribution. This was answered for the sum by Makarov [17], who showed the result was a set of distributions and was able to provide bounds on this function. Sklar, Schweizer and Frank generalised this result to other (positive) binary operations [11, 24]. In this pursuit they created copulas, a general way to encode probabilistic dependence independently from marginals. These are now an essential object used in probabilistic modelling. In his dissertation [29], Williamson described an algorithm for efficiently performing these arithmetic operations, which can give guaranteed

bounds on probability distributions in terms of an upper and lower cdf. He called his method *Probabilistic Arithmetic* and his sets of distributions *Dependency Bounds*. Since then the method has been generalised [10, 7, 9] to most of the base binary and unary operations that would be present in a programming language. Probability boxes (p-boxes) are the name now given to these structures

The goal of `ProbabilityBoundsAnalysis.jl`[1] is to **compute guaranteed bounds on functions of random variables given only partial knowledge of the input probability distributions and their dependencies**. That is to compute with partial knowledge about the input joint distribution. Ideally, all of the available information about random variables should be used, but no more than what actually is available.

The idea of bounding probability has a long tradition throughout the history of probability theory. George Boole [2, 13] used the notion of interval bounds on probability. Chebyshev [3] described bounds on a distribution when only the mean and variance of the variable are known, and Markov [18] found bounds on a positive variable when only the mean is known. Fréchet [12] discovered how to bound joint distributions solely from knowing the marginal distributions, without making independence assumptions. Bounding probabilities has continued to the present day, culminating into the modern theory of Imprecise Probabilities [27, 15, 26, 1].

Imprecise probability is effectively a generalisation of probability theory where uncertainty can be expressed about the probability measure. This is particularly relevant when information is scarce, unreliable, vague, conflicting or imprecise. In such cases defining a unique probability distribution is difficult. P-boxes are one of many ways to describe a set of distributions, others include: Dempster-Shafer structures [4, 25], random sets [20], possibility distributions [31, 5, 14] and lower previsions [26]. These structures were discovered independently, but are often synonymous and can be translated from one to another, with different degrees of generality. Imprecise probability links all these theories into one. For a comprehensive overview of the theory, and a for a formal description of uncertainty and information in terms of these structures, [15] is recommended.

2. Probability Boxes

A real-valued random variable is characterised by its distribution function F, which is a monotonically increasing function from the real numbers onto the interval $[0, 1]$ such that the value of the function at negative infinity is zero and the value of the function at positive infinity is one. A p-box consists of a pair of such functions that
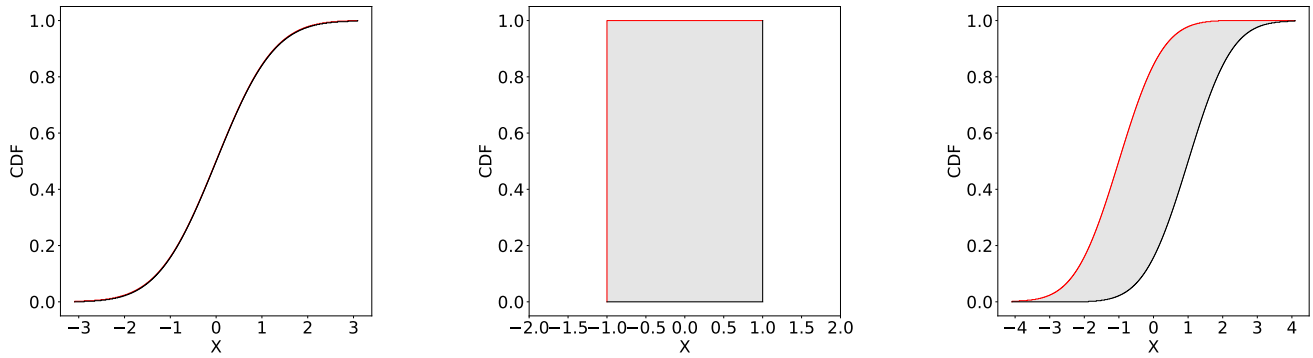
---

[1]https://github.com/AnderGray/ProbabilityBoundsAnalysis.jl

Fig. 1: A precise distribution, an interval and a probability box

are used to circumscribe an imprecisely known distribution function F. The p-box, consisting of the pair of bounding functions, identifies a set of probability distributions that lie entirely within these bounds. Additional information about the random variable may be available, such as bounds on its mean and variance and its distribution family, which may be used to further restrict the set of distributions. A p-box is thus defined by the following three constrains: (1) two bounding cumulative distribution functions (cdf), (2) interval bounds on the mean and variance, and (3) a collection of distribution families:

(1) $\underline{F}(x) \leq F(x) \leq \overline{F}(x)$

(2) $\mu \in [\underline{\mu}, \overline{\mu}]$

$\sigma^2 \in [\underline{\sigma}^2, \overline{\sigma}^2]$

(3) $F \in \mathbf{F}$

That is, a random variable is a member of a p-box if its cdf F falls within the cdf bounds of the p-box $F(x) \in [\underline{F}(x), \overline{F}(x)]$ for all x, its moments are inside the interval moments of the p-box, and it belongs to a family of distribution functions (e.g. normal, uniform) considered by the p-box. Some of the constraints may be missing. For example, if the distribution family is unknown then the set is defined solely from the cdf and moment bounds. Such p-boxes are sometimes called non-parametric, since its members do not belong to any particular class of distribution. Some constraints may also be inferred from others. For example, the interval moments may be bounded from the cdf bounds, and cdf may be bounded from moment information (explored further in section *2.2.2*). Figure[2] 1 shows an example of a distribution function, interval and a p-box. The grey shaded region in the interval and p-box bounds a (potentially infinite) collection of probability distributions.

All of a p-box's probabilistic quantities are intervals. The cdf of a p-box may be found by:

$$[\underline{F}(x), \overline{F}(x)],$$

a sample of the p-box may be drawn using the inverse cdfs:

$$[\overline{F}^{-1}(\alpha), \underline{F}^{-1}(\alpha)]$$

where $\alpha \sim U(0, 1)$ is a sample from a uniform distribution, and the probability measure[3] on some interval $U = [a, b]$ is bounded as follows:

$$\underline{\mathbb{P}}(U) = \max(0, \underline{F}(b) - \overline{F}(a))$$
$$\overline{\mathbb{P}}(U) = \overline{F}(b) - \underline{F}(a),$$

where the max operator is required when $\underline{F}(b) < \overline{F}(a)$. Note the same can nearly be achieved by using the precise formula for the probability measure, $\mathbb{P}(U) = F(b) - F(a)$, and using interval arithmetic.

P-boxes generalise precise distributions and intervals in the following way. A distribution is a p-box with a precise cdf and moments, i.e. when:

$$\underline{F}(x) = \overline{F}(x),$$
$$\underline{\mu} = \overline{\mu},$$
$$\underline{\sigma}^2 = \overline{\sigma}^2$$

and an interval $[a, b]$ is a p-box whose bounds are step functions:

$$\underline{F}(x) = \epsilon_b(x)$$
$$\overline{F}(x) = \epsilon_a(x),$$

where ϵ_k is:

$$\epsilon_k(x) = \begin{cases} 0 & \text{when } x < k \\ 1 & \text{when } x \geq k \end{cases}$$

Moreover, theoretical bounds on the mean and variance of an interval can be found:

$$\mu \in [a, b]$$
$$\sigma^2 \in [0, (b - a)^{2/4}]$$

That is, it is not possible to find a distribution whose range is in $[a, b]$ and whose variance is greater than $(b - a)^{2/4}$. The lower

---

[2]Scripts for all figures can be found in /examples/JuliaCon

[3]The probability measure is a function which returns the probability that the random variable is in some set
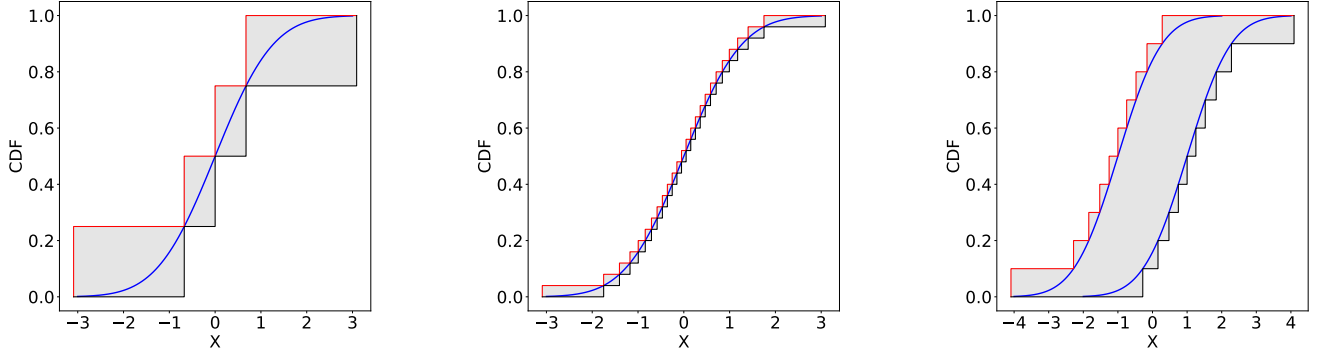
Fig. 2: Various outer representations of distribution functions, left 4 steps and centre 25 steps, and a p-box (right) with 10 steps

bound on the variance is zero, since scalars are also included in the interval. Under this definition of an interval, a random sample will always be the interval $[a, b]$, the cdf returns 0 when $x < a$, the interval $[0, 1]$ when $a <= x < b$, and 1 when $b <= x$. Further, the probability measure of the interval $X = [a, b]$ will be:

$$\mathbb{P}_X(U) = \begin{cases} 0 & \text{when } U \cap [a, b] = \emptyset \\ 1 & \text{when } [a, b] \subseteq U \\ [0, 1] & \text{otherwise} \end{cases}$$

That is, if the set U does not intersect the interval X, $\mathbb{P}_X = 0$, i.e. U certainly does not contain any of the random variables. If X is fully contained in U, $\mathbb{P}_X = 1$, i.e. U certainly contains all of the random variables. And finally if U intersects (but does not contain) X, the probability measure is the vacuous probability interval $\mathbb{P}_X = [0, 1]$, i.e. we are completely uncertain about the containment. Note that the $P_X = \{0, 1, [0, 1]\}$ is due to the interval bounds being step functions. Generally p-boxes may yield other probability intervals.

2.1 Outer representations of p-boxes

An important feature of probability bounds analysis is how distribution functions and p-boxes are represented. Generally, analytical solutions for cdfs are not readily available, for example the normal distribution's cdf can only be found by integrating the probability density, usually with quadrature. However, even if the functions were available analytically, when the variables are used in an arithmetic operation the output distribution will not necessarily belong to the same family. Therefore, we require a representation of these continuous functions which is not dependent on the distribution family (for example as used in polynomial chaos expansion), and ideally is robust in the sense that it can produce an interval error for the numerical representation. Such a representation of a probability is sometimes called an outer approximation, in the same fashion an interval is an outer approximation of a floating point number, which contracts to the exact value as more computational resources are used.

We follow the representation first introduced by Williamson and Downs [28], where a discrete upper and lower approximations of distribution functions are constructed using the inverse cdfs. Note that when considering inverse cdfs $\underline{F}^{-1}(p) \geq F^{-1}(p) \geq \overline{F}^{-1}(p)$, for $p \in [0, 1]$. An outer approximation using two finite vectors[4] u

[4]We begin vector indexing with 1 as is done in the Julia language.

and d of length N is constructed by evaluating the inverse cdfs for uniformly spaced probability levels $p_i = \frac{i-1}{N}$ for $i = 1, ..., N + 1$, e.g. $p_i = [0, 0.25, 0.5, 0.75, 1]$ for $N = 4$. The vectors u and d are defined as follows:

$$u[i] = \overline{F}^{-1}(p_i)$$
$$d[i] = \underline{F}^{-1}(p_{i+1}) \qquad \text{for } i = 1, ..., N$$

The vectors u and d are a collection of "real" or "physical" evaluations of a the random variable X, as opposed to cdf evaluations. Note that although the probability levels p are uniformly spaced, u and d are generally not. The vectors u and d can then serve as bounds on the monotonic functions $\overline{F}^{-1}$ and $\underline{F}^{-1}$:

$$u[i] \leq \overline{F}^{-1}(p) \qquad p_i \in [p_i, p_{i+1})$$
$$d[i] \geq \underline{F}^{-1}(p) \qquad p_i \in (p_i, p_{i+1}]$$

That is, in the region $[p_i, p_{i+1})$, $u[i]$ is a lower bound on $\overline{F}^{-1}$ and $d[i]$ is an upper bound on $\underline{F}^{-1}$ in $(p_i, p_{i+1}]$. In the case that the p-box characterises a precise distribution and $\overline{F}^{-1}(p) = \underline{F}^{-1}(p)$ gives: $u[i + 1] = d[i]$. Figure 2 shows examples of a distribution and a p-box with various outer representations.

A minimal[5] p-box structure is as follows:

```julia
using IntervalArithmetic

struct pbox{T <: Real}
  u :: Vector{T}     # left discrete inverse
  d :: Vector{T}     # right discrete inverse
  m :: Interval{T}   # interval mean
  v :: Interval{T}   # interval variance
  shape :: String    # shape
end
```

where the vectors u and d are the "up" and "down", or the "left" and "right", edges of the p-box[6]. The intervals m and v are the interval moments, and the $shape$ is a string containing any distribution fam-

[5]The code for the minimal structure can be found in examples/minimal

[6]There can be confusion since $\overline{F}$ is the upper bound on the cdf, but $\overline{F}^{-1}$ is the lower bound on the random variable, i.e. the left edge of the p-box. We therefore $\overline{F}$ the "up" or "left" edge, and $\underline{F}$ the "down" or the "right" edge.
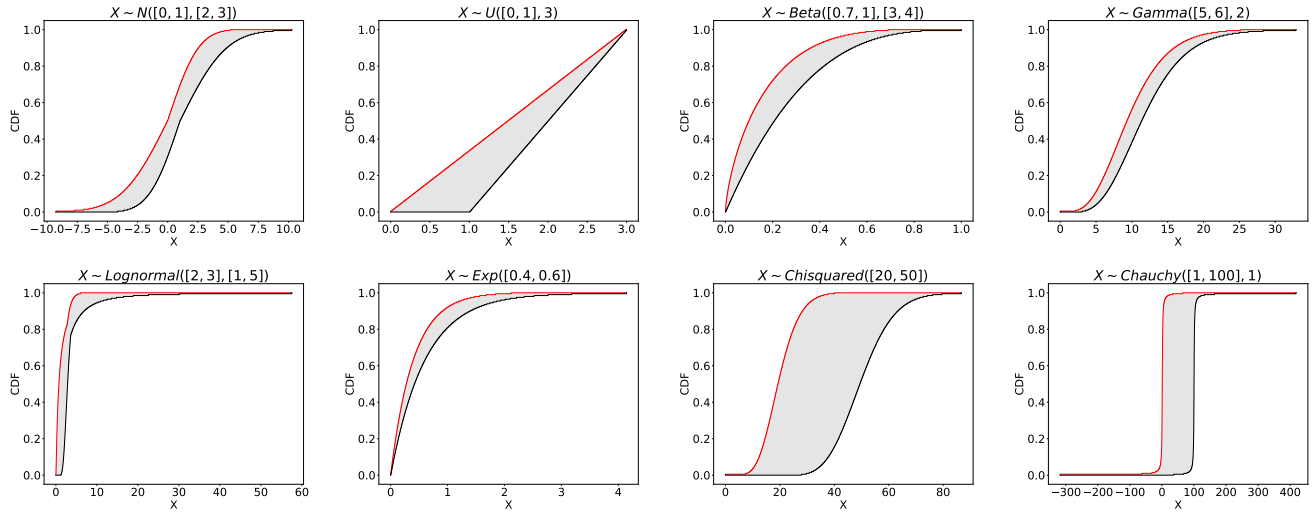
Fig. 3: Various parametrically constructed p-boxes with interval parameters

ily information. An example of a normal distribution constructor is as follows:

```julia
using Distributions
const steps = 200

function normal(m :: Interval, std :: Interval)

  ps = range(0, 1, length = steps+1)

  b1 = quantile.(Normal(m.lo, std.lo), ps)
  b2 = quantile.(Normal(m.lo, std.hi), ps)
  b3 = quantile.(Normal(m.lo, std.hi), ps)
  b4 = quantile.(Normal(m.hi, std.hi), ps)

  u = min.(b1, b2, b3, b4)[1:end-1]
  d = max.(b1, b2, b3, b4)[2:end]

  return pbox(u, d, m, std^2, "normal")
end
```

Other distributional p-boxes are constructed in a similar way. A evaluation of the cdf in terms of this representation is as follows:

```julia
function cdf(F :: pbox, x :: Real)

  u = F.u; d = F.d; n = length(u);

  cdf_u = 1 - sum(x .< u)/n;
  cdf_d = 1 - sum(x .<= d)/n;

  return interval(cdf_d, cdf_u)
end
```

A horizontal cut (an evaluation of the inverse cdf) and a sample of a p-box can be found as follows:

```julia
import Base: rand

function cut(F :: pbox, p :: Real)

  u = F.u; d = F.d; n = length(u);
  ps = range(0, 1, length = n+1);
```

```julia
  ind_u = findlast(ps[1:end-1] .<= p)
  ind_d = findfirst(ps[2:end] .>= p)

  return interval(u[ind_u], d[ind_d])
end

# a random sample is a random cut
rand(F :: pbox, N :: Int) = cut.([F], rand(N))
```

And finally the probability measure, which we call the *mass* for short:

```julia
function mass(F :: pbox, x :: Interval)

  cdfHi = cdf(F, x.hi)
  cdfLo = cdf(F, x.lo)   # returns intervals

  ub = cdfHi.hi - cdfLo.lo
  lb = max(0, cdfHi.lo - cdfLo.hi)

  return interval(lb, ub)
end
```

Other available functions on p-boxes which we do not describe here are:

—**env** or ∪: takes the envelope or hull of two p-boxes. A convex set-union of the probability sets

—**imp** or ∩: imposition or intersection of two p-boxes. Set-intersection of probability sets

—**subseteq** or ⊆: checks if a p-box is a subset or equal to another

—**in** or ∈: checks if a distribution is a member of a p-box

—**mixture**: a stochastic mixture of p-boxes.

2.2 Where do p-boxes come from?

In this section we discuss some situations where p-boxes arise naturally.
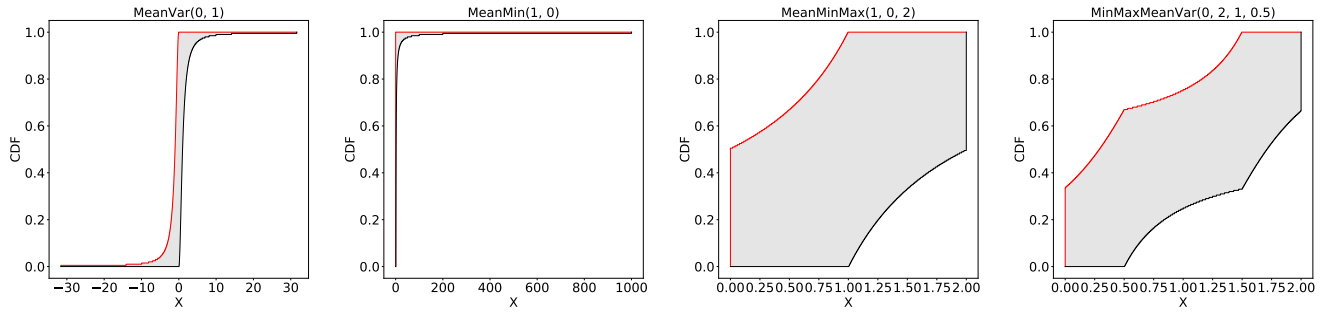
2.2.1 Partial distributional information.

Fig. 4: Examples of distribution-free p-boxes, constructed from various range and moment constrains

In situations where the parametric family of the uncertainty is known, for example it is known that X is normally distributed $X \sim N(\mu_X, \sigma_X^2)$ or that Y is a uniform $Y \sim U(a_Y, b_Y)$, but the precise parameters to these functions are unknown, then p-boxes are a natural representation. In the case that bounds are known on parameters, then bounds on the cdf can be found using the interval parameters and the inverse cdf of the parametric family. Figure 3 shows a variety of p-boxes constructed this way.

2.2.2 Missing family but partial moment information.

It is well known that generally a random variables distribution function cannot be uniquely determined solely from its moment information, even with an infinite number of moments, unless some assumptions are made about the family (i.e. it's normal). In probability theory this is known as the *moment problem*. However, in these situations it is possible to define a p-box which bounds all random variables with the known properties, without loss of generality or making unwarranted assumptions.
These p-boxes are constructed using the Chebyshev (mean and variance) [3], Markov (mean and one bound) [18] and Cantelli (mean and range) inequalities from probability theory. These inequalities allow for the cdf to be bounded from various range and moment information. For example, the Chebyshev inequality can be used to find p-box bounds from the random variables mean and variance:

$$\underline{F}_X(x) = \begin{cases} 1/(1 + (x - \mu_X)^2/\sigma_X^2), & \text{if } \mu_X < x \\ 1, & \text{if } x \le \mu_X \end{cases}$$

$$\overline{F}_X(x) = \begin{cases} 1/(1 + \sigma_X^2/(x - \mu_X)^2), & \text{if } x < \mu_X \\ 1, & \text{if } \mu_X \le x \end{cases}$$

A list of available distribution-free constructors are:

—**MeanVar (or MeanStd)**: using Chebyshev

—**MeanMin**: using Markov

—**MeanMax**: using Markov

—**MeanMinMax**: using Cantelli

—**MinMaxMeanVar (or MinMaxMeanStd)**: using a mix of inequalities

All distribution-free constructors may also take interval moments parameters.

2.2.3 Operations involving intervals and distributions.

An arithmetic operation involving an interval and a precise distribution will produce set of distributions, i.e. a unique distribution for every real number in the interval, whose p-box bounds can be found. The following function performs a binary operation between a p-box and an interval:

```
using MomentArithmetic

function binary(F :: pbox, X :: Interval, op)

  u_z = op.(Interval.(F.u), X)
  d_z = op.(Interval.(F.d), X)

  u_z_ = getfield.(u_z, :lo)   # get right edges
  d_z_ = getfield.(d_z, :hi)   # get left edges

  # for moments
  F_range = interval(F.u[1], F.d[end])
  F_moms = Moments(F.m, F.v, F_range);
  Z_moms = op(F_moms, X);

  return pbox(u_z_, d_z_, Z_moms.mean,
  Z_moms.var, F.shape)
end
```

How the moment transformations are performed are not described here, however they are available in `MomentArithmetic.jl`[7] [8]. Figure 5 shows an example of an operation between a precise distribution and an interval.
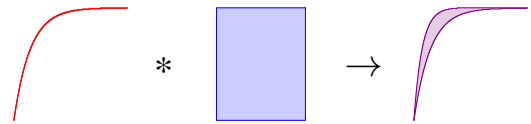

Fig. 5: Illustration of a p-box resulting from a binary operation between a distribution and an interval

2.2.4 When dependencies between distributions are unknown.

In the case where a binary operation is performed between two distributions but whose dependence or correlation is unknown, a single output distribution also cannot be defined. However a bounding p-box can be computed. This is the original question of Kolmogorov as described in the introduction. We call such operations

---

[7]https://github.com/AnderGray/MomentArithmetic.jl

between p-boxes where the marginals are fixed but no joint information is known *Fréchet* operations, after the mathematician Maurice R. Fréchet, who first discovered how to bound joint distributions with fixed marginals [12]. How operations between p-boxes with partial or no dependence information is known is described in section 3.2. Figure 6 illustrates a Fréchet operation between a uniform and a normal distribution.
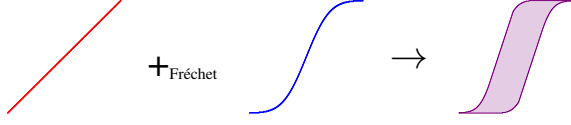

Fig. 6: Illustration of a p-box resulting from a Fréchet operation

2.2.5 *Outer approximations of precise distributions.*

Outer approximations of precise distributions described in section 2.1 also lead naturally to a p-box. All probabilistic quantities of discretised distribution functions will be intervals, which shrink to precise quantities as the discretisation increases. In that sense, both distributions functions and p-boxes are p-boxes in `ProbabilityBoundsAnalysis.jl`.

2.2.6 *Inferential methods when data is limited or bad.*

Probability distributions are almost always constructed from data or measurements, or otherwise are assumed as a model or derived from expert opinion. Usually to construct a probability distribution accurately, a large amount of high quality data is required. P-boxes arise naturally when this is not the case, for example if the data are intervals instead of points, or when data is sparse or not reliable, and an exact distribution function is difficult to prescribe.

An example is robust Bayesian analysis, where the analyst does not have enough information to reliably establish a single prior and likelihood function, but instead has a collection of priors and likelihoods, leading to a set of posteriors distributions, to which a p-box can be identified.

A method which is growing in popularity are inferential models [19]. These generalise classical inference methods to account for limited statistical knowledge, and give a statistical inference theory which is non-additive (as opposed to classical probability which is additive). These inference procedures, often based on random sets, give certain validity and performance guarantees. Inferential models can viewed from the perspective of bounding sets of distributions, and is closely linked to possibility theory [16]. A p-box can be prescribed to such structures, leading to structures like confidence boxes or c-boxes [6], which follow the same arithmetic rules as p-boxes.

2.3 Bivariate p-boxes

P-boxes have been extended to higher dimensions [21], however like with precise probabilities, when considering multivariate p-boxes the dependence information (like correlation) must be accounted for. Considering complex dependencies amongst p-boxes is important, as they play a key role in binary operations. That is, the result of an operation between two p-boxes not only depends on the input p-boxes, but also how the inputs are correlated.

Covariance, and subsequently the pearson correlation coefficient, is insufficient to fully determine the dependence between two random variables. Like for the univariate moments, a single multivariate distribution cannot be prescribed for a given value of the covariance without making distributional assumptions (like normality). A more descriptive (and general) model for dependence is required to exactly specify a multivariate distribution.

Copulas [22] provide a solution to this problem. A copula is a multivariate cdf with uniform marginals on the interval $[0, 1]$. Essentially a copula is a multivariate distribution which has all marginal (univariate) information stripped away, leaving only the dependency. All dependence information can therefore be encoded in a copula exactly and completely separately from its marginals. Often multivariate problems in probability can be reduced to an analysis of copulas.

A bivariate copula (2-copula) C is any function $C : [0, 1]^2 \rightarrow [0, 1]$ with the following properties:

(1) Grounded: $C(0, v) = C(u, 0) = 0$,
(2) Uniform margins: $C(u, 1) = u$; $C(1, v) = v$,
(3) 2-increasing:
$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0$
for all $0 \leq u_1 \leq u_2 \leq 1$ and $0 \leq v_1 \leq v_2 \leq 1$.

Three important 2-copulas are:

$$W(u, v) = \max(u + v - 1, 0),$$
$$\Pi(u, v) = uv,$$
$$M(u, v) = \min(u, v),$$

where the Π encodes stochastic independence, M encodes maximally positive correlation, and W encodes maximally negative correlation. Moreover, W and M are bounds on all 2-copulas: $W(u, v) \leq C(u, v) \leq M(u, v)$. A bivariate cdf F_{XY} can be constructed from two marginal cdfs F_X and F_Y and copula C using Sklar's theorem:

$$F_{XY}(x, y) = C(F_X(x), F_Y(y)),$$

where the constructed bivariate distribution will follow the dependence encoded in the copula and have the prescribed marginals. Sklar's theorem also works in reverse, where the dependence can be extracted from a multivariate distribution:

$$C(u, v) = F_{XY}(F_X^{-1}(u), F_Y^{-1}(v)).$$

Copulas also give a very convenient way to sample from F_{XY} in a generalisation of inverse transform sampling, where first the copula is sampled $(u, v) \sim C$, and then transformed through the inverses of the marginals $(F_X^{-1}(u), F_Y^{-1}(v))$.

Sklar's theorem has a straightforward imprecise extension [21], where the two bivariate bounds of a p-box $[\underline{F}_{XY}(x, y), \overline{F}_{XY}(x, y)]$ can expanded in terms of two marginal p-boxes and bounds on a copula $[\underline{C}(u, v), \overline{C}(u, v)]$:

$$\underline{F}_{XY}(x, y) = \underline{C}(\underline{F}_X(x), \underline{F}_Y(y)),$$
$$\overline{F}_{XY}(x, y) = \overline{C}(\overline{F}_X(x), \overline{F}_Y(y)).$$

Samples and the probability measure can also be found from a bivariate p-box. Like how a sample of a univariate p-box is an interval, a sample of a bivariate p-box will be an interval box, i.e. a two dimensional rectangular set. Multiple samples will therefore be a population of interval boxes, which together follow the dependency encoded in C. The bottom of Figure 7 shows 100 samples of a bivariate p-box, whose marginals are $X \sim \text{Beta}([4, 8], 3)$

and $Y \sim N([0, 2], 2)$, and with the Gaussian copula (multivariate normal's copula) with correlation coefficient $\rho_{XY} = -0.8$: $C = \Phi_{-0.8}$. The top of Figure 7 shows the cdf bounds. The probability measure from a bivariate p-box in the rectangular set $U = [x_1, x_2] \times [y_1, y_2]$ is[8]:

$$\underline{\mathbb{P}}(U) = \max(0, \underline{F}(x_2, y_2) - \overline{F}(x_1, y_2) - \overline{F}(x_2, y_1) + \underline{F}(x_1, y_1)),$$

$$\overline{\mathbb{P}}(U) = \min(1, \overline{F}(x_2, y_2) - \underline{F}(x_1, y_2) - \underline{F}(x_2, y_1) + \overline{F}(x_1, y_1)).$$

A minimal structure, cdf, random sample and probability measure of a bivariate p-box is [9]:

```
using BivariateCopulas
using BivariateCopulas: sample

struct BivPbox{T <: Real}

  X :: pbox{T}   # 1st marginal
  Y :: pbox{T}   # 2nd marginal
  C :: copula    # copula

end

function cdf(F :: BivPbox, x :: Real, y :: Real)

  cdf_X = cdf(F.X, x)   # returns interval
  cdf_Y = cdf(F.Y, y)

  F_cdf_lo = F.C(cdf_X.lo, cdf_Y.lo)
  F_cdf_hi = F.C(cdf_X.hi, cdf_Y.hi)

  return interval(F_cdf_lo, F_cdf_hi)
end

function rand(F :: BivPbox, N :: Int)

  C_samps = sample(F.C, N)

  X_samps = cut.([F.X], C_samps[:, 1])
  Y_samps = cut.([F.Y], C_samps[:, 2])

  return IntervalBox.(X_samps, Y_samps)
end

function mass(F :: BivPbox, box :: IntervalBox)

  x = box.v[1]; y = box.v[2];  # gets intervals

  F_1 = cdf(F, x.lo, y.lo)
  F_2 = cdf(F, x.hi, y.lo)
  F_3 = cdf(F, x.lo, y.hi)
  F_4 = cdf(F, x.hi, y.hi)

  # using intervals
  mass = F_4 - F_3 - F_2 + F_1;

  return max(0, min(1, mass))
end
```

3. P-box arithmetic

In this section we will discuss how p-boxes can be propagated through unary and binary operations in a rigorous way.

---

[8]Which is nearly a straightforward application of interval arithmetic to the precise calculation $F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1)$
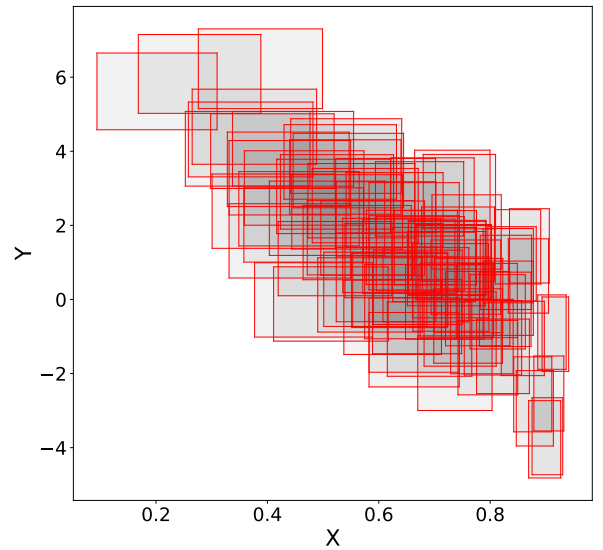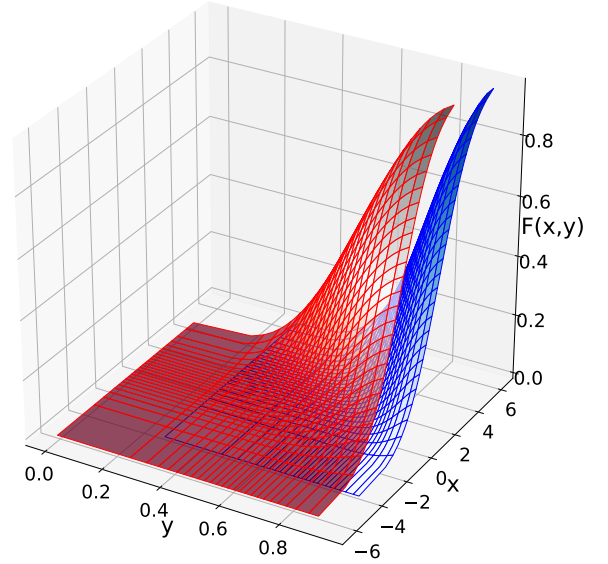[9]https://github.com/AnderGray/BivariateCopulas.jl

Fig. 7: Cdf bounds of a bivariate p-box (top) with $X \sim \text{Beta}([4, 8], 3)$, $Y \sim N([0, 2], 2)$, and $\rho_{XY} = -0.8$, and 100 samples (bottom)

3.1 Unary operations

Piggybacking off `IntervalArithmetic.jl`[10] [23], a p-box can be transformed through any unary operation $f : \mathbb{R} \to \mathbb{R}$ that has an interval extension, i.e. accepts intervals. For this, we first transform the p-box into a Dempster-Shafer structure [4, 25], propagate the structure through the operation with interval arithmetic, and transform it back to a p-box. A Dempster-Shafer structure (also called a belief function) is a finite collection of intervals,
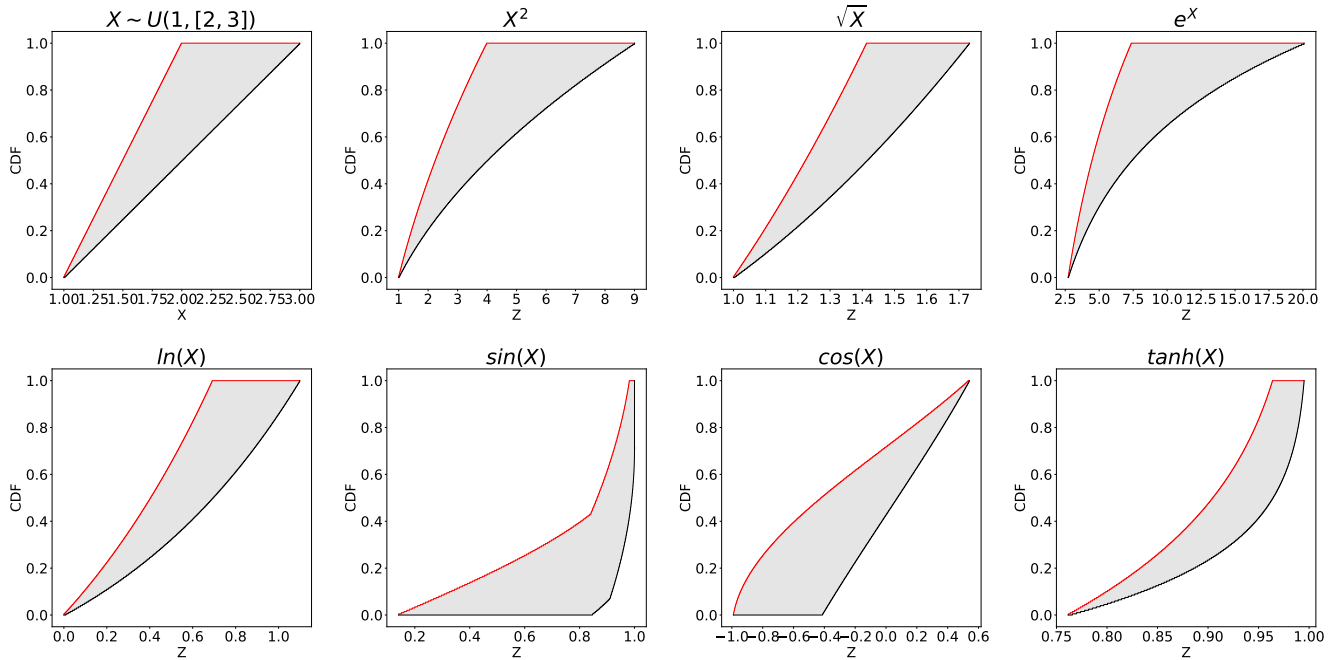
---

[10]https://github.com/JuliaIntervals/IntervalArithmetic.jl

Fig. 8: Top left shows an initial p-box, and others show results of unary operations on this p-box

called focal elements X, with probabilities attached to each interval, called basic mass assignments m, which must sum to 1. An example of a belief function with three focal elements is: $X = \{[0, 2], [0.5, 1.3], [1, 3]\}$ and $m = \{0.25, 0.5, 0.25\}$. Like a p-box, a Dempster-Shafer also bounds a set of probability distributions, using two set based functions called the *belief* and *plausibility*:

$$\text{bel}(U) = \sum_{X \subseteq U} m(X),$$

$$\text{pl}(U) = \sum_{X \cap U \neq \emptyset} m(X),$$

where the sum is taken over focal elements X. For example, the belief and plausibility of $U = [0, 2.5]$ are $\text{bel}([0, 2.5]) = 0.75$ ($\{[0, 2], [0.5, 1.3]\}$ are subsets) and $\text{pl}([0, 2.5]) = 1$ (all focal elements intersect). The belief and plausibility serve as lower and upper bounds on the probability measure:

$$\text{bel}(U) \leq \mathbb{P}(U) \leq \text{pl}(U).$$

A p-box is a special case of a belief function, where the focal elements are ordered by $\leq$. Looking at the leftmost p-box in Figure 2 (with 4 steps) it can be seen that the p-box could be represented with 4 intervals (i.e. each box in the plot is a focal element). The mass assignment in this case is the difference in cdf level of each box $p_{i+1} - p_i$, which gives an equal mass assignment since we discretised the p-box uniformly. Note that for a general belief function, the focal elements do not have to be ordered, as is in the case of the previous example: $X = \{[0, 2], [0.5, 1.3], [1, 3]\}$. The following function converts a p-box into a belief function:

```
function split(F :: pbox)
```

```
    n = length(F.u)

    focal_el = interval.(F.u, F.d)
    masses   = ones(n) ./n

    return focal_el, masses
end
```

A belief function can be converted to a p-box by accumulating the belief and plausibility (the lower and upper probability):

$$\underline{F}(x) = \text{bel}((-\infty, x]),$$

$$\overline{F}(x) = \text{pl}((-\infty, x]),$$

which numerically corresponds to sorting the bounds of the focal elements, accumulating the basic mass assignments, and finding the probability levels. Assuming that the basic mass assignments are equal, the following code converts a belief function into a p-box:

```
function makepbox(focal_el)

    x_lo = getfield.(focal_el, :lo);
    x_hi = getfield.(focal_el, :hi);

    u = sort(x_lo)
    d = sort(x_hi)

    return pbox(u, d)
end
```

The general function where the focal elements have non-equal mass is simple, but slightly lengthy, and will not be described here. A function to perform a unary operation on a p-box is thus:

```
function unary(F :: pbox, op)

  focal_el, _ = split(F)
  fe_z = op.(focal_el)

  return makebox(fe_z)
end
```

Figure 8 show examples of unary operations on a p-box.

3.2 Binary operations

Binary operations on p-boxes are based on *convolutions*, which are operations on two functions which produce a third. A convolution often used in signal processing is:

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(z-x)f_Y(x)dx.$$

It is well known that the result of above convolution f_Z corresponds to the probability density of the sum of two independent random variables with densities f_X and f_Y. P-box arithmetic is based on generalisations of such convolutions specifically designed for computing binary operations on distribution functions, which work with operations other than sum, and for copulas other than independence. These convolutions, which first appeared in the study of Probabilistic Metric Spaces [24], can be be used to compute binary operations between two p-boxes when their copula C is precisely known, only it's lower bound is known $\underline{C}$, or even missing completely. Further, Williamson and Downs [28] describe how robust outer approximations of the convolutions can be found, accounting for the representation error of the two input p-boxes, as well as the error from the integration.

3.2.1 Operations with known C.

A binary operation between two random variables X and Y with cdfs F_X and F_Y and copula C can be computed with the following Lebesgue-Stieltjes integral:

$$\sigma_{C,L}(z) = \int_{L\{z\}} dC(F_X(x), F_Y(y)),$$

where the integration domain is the set $L\{z\} = \{(x,y)|x,y \in \mathbb{R}, L(x,y) < x\}$ (the set of all x and y for which $L(x,y) < z$), and where L is some binary operation (e.g. $L(x,y) = x + y$). The above $\sigma_{C,L}$ convolution is written for precise distributions, however can be extended for p-boxes by performing the convolution with some combination of the p-boxes bounds in the style of interval arithmetic:

$$+ \begin{cases} \underline{F_Z} = \underline{F_X} + \underline{F_Y} \\ \overline{F_Z} = \overline{F_X} + \overline{F_Y} \end{cases}$$

$$- \begin{cases} \underline{F_Z} = \underline{F_X} - \overline{F_Y} \\ \overline{F_Z} = \overline{F_X} - \underline{F_Y} \end{cases}$$

$$\times \begin{cases} \underline{F_Z} = \underline{F_X} \times \underline{F_Y} \\ \overline{F_Z} = \overline{F_X} \times \overline{F_Y} \end{cases}$$

$$\div \begin{cases} \underline{F_Z} = \underline{F_X} \div \overline{F_Y} \\ \overline{F_Z} = \overline{F_X} \div \underline{F_Y} \end{cases}$$

Williamson and Downs provide an algorithm to compute rigorous bounds on the above convolution, but only under independence (when $C = \Pi$). Around the same time, Yager [30] described a similar style convolution on two belief functions under the assumption independence, which is very similar to that of Williamson and Downs'. We therefore provide a generalised version of Yager's algorithm where any copula C can be specified. Similar to a unary operation, a binary operation between two p-boxes can be performed by first transforming the bivariate p-box into a two dimensional belief function (which has interval boxes as focal elements), evaluating the operation on the focal elements with interval arithmetic, and transforming the resulting belief function (which will be univariate after the operation) back to a p-box. Converting a bivariate p-box into a 2D belief function is similar to that of the 1D case, however this time the mass assignment m will be determined by the copula C. For two univariate focal elements X and Y with probability widths $p_x = [p_{x1}, p_{x2}]$ and $p_y = [p_{y1}, p_{y2}]$ (i.e. their vertical width in Figure 2, with their 1D mass assignment being $p_{x2} - p_{x1}$), the 2D mass assignment of the interval box $X \times Y$ will be:

$$C(p_{x2}, p_{y2}) - C(p_{x1}, p_{y2}) - C(p_{x2}, p_{y1}) + C(p_{x1}, p_{y1})$$

The following function converts a bivariate p-box into a 2D belief function:

```
function split(F :: BivPbox)

  X = F.X; Y = F.Y; C = F.C

  nx = length(X.u); ny = length(Y.u);

  px = range(0, 1, length = nx + 1);
  py = range(0, 1, length = ny + 1);

  f_el_X = interval.(X.u, X.d);
  f_el_Y = interval.(Y.u, Y.d);

  masses = C(px[2:end],   py[2:end])
         - C(px[1:end-1], py[2:end])
         - C(px[2:end],   py[1:end-1])
         + C(px[1:end-1], py[1:end-1]);

  f_XY   = [IntervalBox(x,y) for x in f_el_X,
                                 y in f_el_Y];

  return f_XY[:], masses[:]
end
```

A robust σ convolution is simply thus:

```
function sigma(F :: BivPbox, op)

  nx = length(F.X.u); ny = length(F.Y.d);

  # number of steps of output
  num_steps = max(nx, ny);

  # Convert pbox to a belief function
  fe, masses = split(F)

  # Perform operation with interval arithmetic
  fe_z = [op(f.v[1],f.v[2]) for f in fe]

  # Convert back to p-box
  return makebox(fe_z, masses, num_steps)
end
```
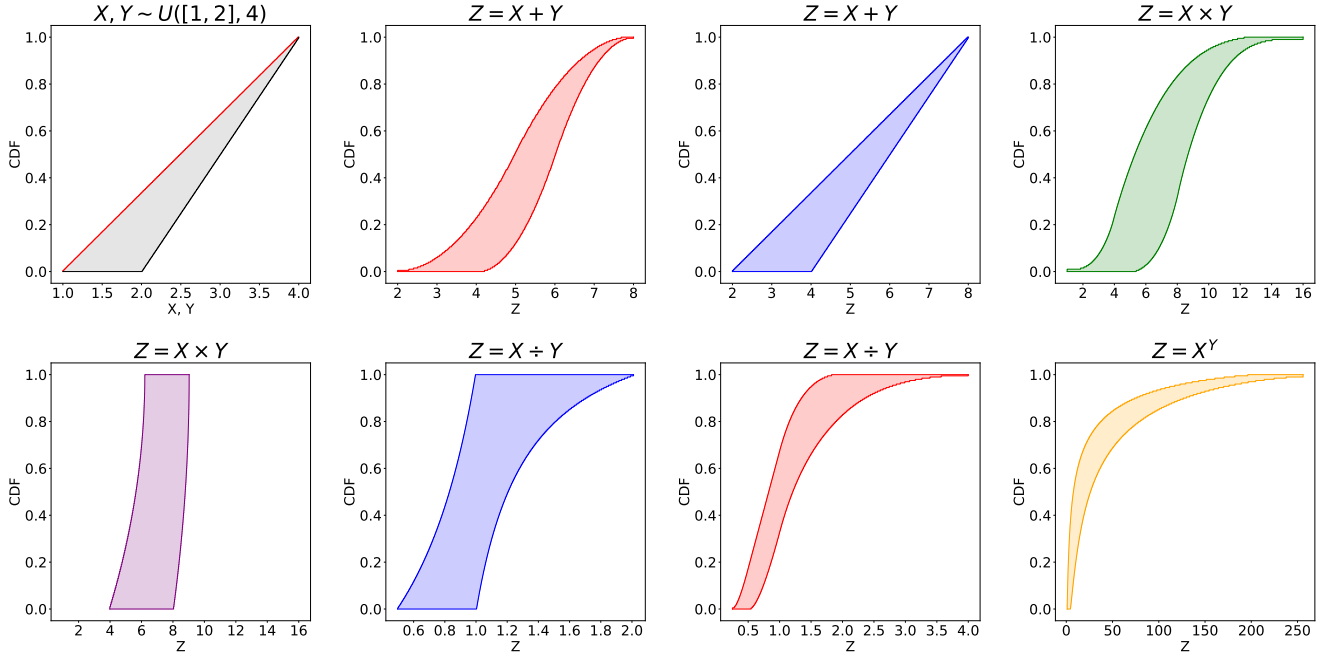
Fig. 9: Top left shows two initial p-boxes $X, Y \sim U([1, 2], 4)$, and others show results of binary operations on these two p-boxes with different dependencies. Each colour represents a different copula. The colour code is $C = \{\ W$ (opposite), $\Phi_{-0.5}$ (correlation -0.5), Π (independence), $\Phi_{0.5}$ (correlation 0.5), M(perfect) $\}$

The above σ convolution will work for any two input p-boxes, with any copula C, and binary operation *op* which has an interval extension. Again the general function for converting a belief function with non-equal masses is required. The 3rd argument *num_steps* is the desired discretisation of the output p-box, in a condensation step. The interval operation in the σ convolution performs the operation on all combinations of focal elements, so for example if each p-box has 200 steps, the resulting p-box will have $200 \times 200 = 40,000$ steps, which must be reduced in an outer approximation. Again, the general algorithm is simple but lengthy and will be omitted. Figure 9 shows examples of σ convolutions for different operators and copulas.

3.2.2   *Operations with lower bound $\underline{C}$.*

ProbabilityBoundsAnalysis.jl also allows for binary operations to be performed between p-boxes when only the lower bound of the copula is known $\underline{C}$. That is, uncertainty can be placed on the dependence of two distributions, as well as the marginals. The resulting p-box of the binary operation will bound all distributions with copulas C which are pointwise greater than the specified copula $\underline{C} \leq C$. These operations are based on two more convolutions from Probabilistic Metric Spaces [24]. For a non-decreasing binary operator $L : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$, two random variables X and Y with distribution functions F_X and F_Y, and with copula lower bound $\underline{C}$:

$$\tau_{\underline{C}_{XY}, L}(z) = \sup_{L(x,y)=z} [\underline{C}(F_X(x), F_Y(y))],$$

$$\rho_{\underline{C}_{XY}, L}(z) = \inf_{L(x,y)=z} [\underline{C}^d(F_X(x), F_Y(y))],$$

where C^d is the copula dual of C, $C^d(u, v) = u + v - C(u, v)$. The τ and ρ convolutions each compute a cdf bound which together characterise the resulting p-box: $[\underline{F}_Z(z), \overline{F}_Z(z)] = [\tau(z), \rho(z)]$. The p-box resulting from a τ and ρ convolution will bound all σ convolutions which have copula C which is greater than $\underline{C}$. For example, setting $\underline{C} = W$ (the lower bound on all copulas) will propagate all copulas, which is complete uncertainty about the dependence, i.e. a *Fréchet* operation. Setting $\underline{C} = \Pi$ will propagate copulas more positive than independence, otherwise known as *positive quadrant dependence* [22]. Setting $\underline{C} = M$ (the upper bound on all copulas) will only propagate M, and will be equal to a σ convolution.

The τ and ρ convolutions can be efficiently computed in terms of the inverse representation using in *ProbabilityBoundsAnalysis*:

$$\tau_{\underline{C}_{XY}, L}^{-1}(p) = \sup_{\underline{C}(u,v)=p} [L(F_X^{-1}(u), F_Y^{-1}(v))],$$

$$\rho_{\underline{C}_{XY}, L}^{-1}(p) = \inf_{\underline{C}^d(u,v)=p} [L(F_X^{-1}(u), F_Y^{-1}(v))],$$

where the operation L is performed on the u and d vectors of the p-boxes, and $\sup_{\underline{C}(u,v)=p}$ and $\inf_{\underline{C}^d(u,v)=p}$ is solved for each probability level.

Note that L must be non-decreasing function (for example $+$ or $\times$), but may be extended to non-increasing functions (for example $-$ or $\div$) by first transforming one of the variables, and performing a non-decreasing operation. For example, $X - Y$ can be evaluated with sum: $X + (-Y)$. Note however, the copula C_{XY} must also be transformed to $C_{X,-Y}$. This transformation however is quite simple [22]:

$$C_{X,-Y}(u, v) = u - C_{XY}(u, 1 - v).$$

Note that when using the above copula transformation, the bound on $\underline{C}_{XY}$ will be flipped to $\overline{C}_{X,-Y}$, and $\tau_{M,-}$ and $\rho_{M,-}$ give the *Fréchet* convolution.

Acknowledgements

4.   References

[1] Thomas Augustin, Frank PA Coolen, Gert De Cooman, and Matthias CM Troffaes. *Introduction to imprecise probabilities*. John Wiley & Sons, 2014.

[2] George Boole. *An investigation of the laws of thought: on which are founded the mathematical theories of logic and probabilities*, volume 2. Walton and Maberly, 1854.

[3] Pafnutii L. Chebyshev [Tchebichef]. *Sur les valeurs limites des intégrales*. Imprimerie de Gauthier-Villars, 1874.

[4] Arthur P Dempster. Upper and lower probabilities induced by a multivalued mapping. In *Classic works of the Dempster-Shafer theory of belief functions*, pages 57–72. Springer, 2008.

[5] D Dubois and H Prade. possibility theory: Approach to computerized processing of uncertainty, plennm n. 4. 1988.

[6] Scott Ferson, Michael Balch, Kari Sentz, and Jack Siegrist. Computing with confidence. In *proceedings of the 8th International Symposium on Imprecise Probability: Theories and Applications*, 2013.

[7] Scott Ferson, Lev Ginzburg, and Resit Akçakaya. Whereof one cannot speak: when input distributions are unknown. *Risk Analysis*, 1996.

[8] Scott Ferson and Ander Gray. Distribution-free uncertainty propagation. In *proceedings of the 9th International Workshop on Reliable Engineering Computing (REC2021)*, 2021.

[9] Scott Ferson and Janos G Hajagos. Arithmetic with uncertain numbers: rigorous and (often) best possible answers. *Reliability Engineering & System Safety*, 85(1-3):135–152, 2004.

[10] Scott Ferson, Vladik Kreinovich, Lev Grinzburg, Davis Myers, and Kari Sentz. Constructing probability boxes and dempster-shafer structures. Technical report, Sandia National Lab.(SNL-NM), Albuquerque, NM (United States), 2015.

[11] Maurice J Frank, Roger B Nelsen, and Berthold Schweizer. Best-possible bounds for the distribution of a sum—a problem of kolmogorov. *Probability theory and related fields*, 74(2):199–211, 1987.

[12] Maurice Fréchet. Généralisation du théoreme des probabilités totales. *Fundamenta mathematicae*, 1(25):379–387, 1935.

[13] Theodore Hailperin. *Boole's logic and probability: a critical exposition from the standpoint of contemporary algebra, logic and probability theory*. Elsevier, 1986.

[14] Dominik Hose and Michael Hanss. Possibilistic calculus as a conservative counterpart to probabilistic calculus. *Mechanical Systems and Signal Processing*, 133:106290, 2019.

[15] George J Klir and Mark J Wierman. *Uncertainty-based information: elements of generalized information theory*, volume 15. Physica, 2013.

[16] Chuanhai Liu and Ryan Martin. Inferential models and possibility measures. *arXiv e-prints*, pages arXiv–2008, 2020.

[17] GD Makarov. Estimates for the distribution function of a sum of two random variables when the marginal distributions are fixed. *Theory of Probability & its Applications*, 26(4):803–806, 1982.

[18] A Markov [Markoff]. Sur une question de maximum et de minimum: Proposée par m. tchebycheff. *Acta Mathematica*, 9:57–70, 1900.

[19] Ryan Martin and Chuanhai Liu. *Inferential models: reasoning with uncertainty*, volume 145. CRC Press, 2015.

[20] Ilya Molchanov. *Theory of random sets*, volume 2. Springer, 2005.

[21] Ignacio Montes, Enrique Miranda, Renato Pelessoni, and Paolo Vicig. Sklar's theorem in an imprecise setting. *Fuzzy Sets and Systems*, 278:48–66, 2015.

[22] Roger B Nelsen. *An introduction to copulas*. Springer Science & Business Media, 2007.

[23] David P. Sanders, Luis Benet, and et. al. Juliaintervals/intervalarithmetic.jl: v0.19.2, September 2021. doi:10.5281/zenodo.5519761.

[24] Berthold Schweizer and Abe Sklar. *Probabilistic metric spaces*. Courier Corporation, 2011.

[25] Glenn Shafer. *A mathematical theory of evidence*, volume 42. Princeton university press, 1976.

[26] Matthias CM Troffaes and Gert De Cooman. *Lower previsions*. John Wiley & Sons, 2014.

[27] Peter Walley. *Statistical reasoning with imprecise probabilities*. Chapman and Hall, London, 1991.

[28] Robert C Williamson and Tom Downs. Probabilistic arithmetic. i. numerical methods for calculating convolutions and dependency bounds. *International journal of approximate reasoning*, 4(2):89–158, 1990.

[29] Robert Charles Williamson et al. *Probabilistic arithmetic*. PhD thesis, University of Queensland Australia, 1989.

[30] Ronald R Yager. Arithmetic and other operations on dempster-shafer structures. *International Journal of Man-Machine Studies*, 25(4):357–366, 1986.

[31] Lotfi Asker Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy sets and systems*, 1(1):3–28, 1978.