# Research

CrossMark
click for updates

**Author for correspondence:**

Simon Tindemans

e-mail: s.tindemans@imperial.ac.uk

# Robust estimation of risks from small samples

## Simon H. Tindemans, Goran Strbac

Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, UK

Data-driven risk analysis involves the inference of probability distributions from measured or simulated data. In the case of a highly reliable system, such as the electricity grid, the amount of relevant data is often exceedingly limited, but the impact of estimation errors may be very large. This paper presents a robust nonparametric Bayesian method to infer possible underlying distributions. The method obtains rigorous error bounds even for small samples taken from ill-behaved distributions. The approach taken has a natural interpretation in terms of the intervals between ordered observations, where allocation of probability mass across intervals is well-specified, but the location of that mass within each interval is unconstrained. This formulation gives rise to a straightforward computational resampling method: Bayesian Interval Sampling. In a comparison with common alternative approaches, it is shown to satisfy strict error bounds even for ill-behaved distributions.

## 1. Motivation: risk analysis for grids

Managing a critical infrastructure such as the electricity grid requires constant vigilance on the part of its operator. The system operator identifies threats, determines the risks associated with these threats and, where necessary, takes action to mitigate them. In the case of the electricity grid, these threats include sudden outages of lines and generators, misprediction of renewable generating capacity, and long-term investment shortages.

Quantitative risk assessments are often based on data: In cases where directly relevant observations are available (e.g. historical wind variability), these can be used directly to estimate future performance (hindcasting). In other cases, risk analysis is based on models, for example in distribution network outage planning. Even so, for sufficiently complex models, risks are rarely computed directly. Instead, Monte Carlo simulations are used to generate 'virtual' observations, which are analysed statistically.

# THE ROYAL SOCIETY
PUBLISHING

A particular challenge is the management of high-impact low-probability events. By definition, such events are exceedingly rare, but misjudging their likelihood of occurrence can have significant consequences. Cascading outages that can lead to partial or full system blackouts are a prime example of such events. In recent years, their modelling and analysis has become a very active area of research. Models are typically evaluated using Monte Carlo simulations [1,2], although other innovative approaches are also being developed [3].

The statistical validity of simulation outcomes is often not addressed, and when it is, this is typically by means of the standard error with the implicit invocation of the central limit theorem (e.g. [1]). One recent exception is [4], which explicitly asked how many blackout occurrences were required to estimate event probability with sufficient accuracy. However, this analysis was limited to a single binary outcome (i.e. a Bernoulli random variable).

This paper contributes to closing this methodological gap, by introducing a method for the robust analysis of rare event data, which is applicable to generic real-valued observations. A simple example in Section 8 illustrates how the method can be applied to cascading outage simulations. Critically, the method satisfies strict accuracy requirements, which will become increasingly important as operators move away from deterministic security standards (e.g. '$N - 1$') to probabilistic standards that are enforced using Monte Carlo-based tools.

## 2. Introduction

We consider the problem of inferring properties of a random variable $X$ from independent observations $\{x_{1:n}\}$. For example, one may attempt to estimate its expectation $E[X]$ or distribution $F_X^*$. This problem can be found in idealised form in the analysis of Monte Carlo simulations, where independence of observations is often guaranteed.

A range of approaches exist for addressing this elementary inference problem, differing in their underlying assumptions and applicability [5]. On one end of the scale there are simple point estimates, such as the sample mean or - when a distribution is concerned - the empirical distribution. When a quantification of uncertainty is desired it is common practice to report a confidence interval, usually by means of an implicit invocation of the central limit theorem. However, this approach is not suitable when sample sizes are small or the distribution of $X$ is sufficiently heavy-tailed or skewed. In such cases the bootstrap method [6] is a well-established alternative, but it is limited to resampling of observed values, so there is no way to account for features of the distribution that have not been observed. This shortcoming is shared by its Bayesian variant, the Bayesian bootstrap [7]. It is particularly restrictive when sample sizes are very small, or for the analysis of rare events where a small fraction of the probability mass has a disproportionate effect on the parameter of interest.

Ferguson's Dirichlet process [8] may be used for inference beyond strictly observed values, but this requires the selection of a prior measure. Furthermore, the resulting inferred (random) distributions are ultimately discrete. A different approach is taken by the Nonparametric Predictive Inference (NPI) method [9], which makes use of imprecise probabilities to circumvent these issues. However, its definition as an incremental predictive process makes it less suitable for the inference of the underlying time-invariant distribution $F_X^*$ or for efficient computational implementations.

In this paper we introduce a robust method for inference from independent real-valued observations with minimal assumptions. The method relies only on the observations and the specification of a bounding interval $I$ for the random variable $X$ (which may be taken as the entire real line $\mathbb{R}$). It results in conservative posterior distributions for any quantities under consideration, but credible intervals for such quantities remain valid even for ill-behaved distributions. This is a desirable feature for applications where accuracy is paramount, such as risk assessment of critical systems. Specific contributions of this paper are as follows:

- The Dirichlet process is reformulated in a way that separates its expectation and random fluctuations (Section 4)
- A non-parametric robust posterior distribution is defined, making use of imprecise probabilities in the form of probability boxes (Section 5).
- Interval estimates for monotonic functions $q$ of $F_X^*$ (such as mean, median, etc.) are formulated (Section 6), accompanied by a straightforward resampling implementation (Section 7)
- Examples illustrate the robust nature of the proposed method, including an application to blackout risk estimation (Section 8).

## 3. Preliminaries

**Notation** The summary notation $v_{j:k}$ is used to indicate a sequence of values $v_i$ with consecutive indices $i = j, \ldots, k$. Capital symbols refer to random variables or non-decreasing functions on the real line (e.g. cumulative distribution functions); it will be clear which is intended from the context. Caligraphic letters (e.g. $\mathcal{F}$) indicate *random* non-decreasing functions, which are characterised by distributions-of-distributions.

## (a) Dirichlet distribution

Let us consider the problem of determining an unknown discrete (categorical) probability distribution $\Pr^*$ on a set of $k$ disjoint events $A_{1:k}$, i.e. $\sum_{i=1}^{k} \Pr^*(A_i) = 1$. To estimate $\Pr^*$ from independent realisations of the corresponding random variable we may use Bayesian inference with the Dirichlet distribution as a prior distribution. The Dirichlet distribution $\mathrm{Dir}[\alpha_{1:k}]$ is parametrised by $\alpha_{1:k}$ and its probability density is structured as

$$f(p_{1:k-1}; \alpha_{1:k}) \propto \prod_{i=1}^{k} p_i^{\alpha_i - 1}, \tag{3.1}$$

with the constraint $\forall i : p_i \geq 0$. Note that the density is formally only defined as a function of $p_{1:k-1}$, and the final component $p_k$ is computed as $p_k = (1 - \sum_{i=1}^{k-1} p_i)$. Nevertheless we use the notation $p_k$ on the right hand side of Eq. (3.1) to highlight the symmetry of the expression. Permissible values of $\{p_i\}_{i=1}^{k}$ are restricted to the unit $(k-1)$-simplex $\Delta^{k-1}$ that is defined as

$$\Delta^{k-1} = \left\{ \vec{p} \in \mathbb{R}^k : p_i \geq 0, \forall i; \sum_{i=1}^{k} p_i = 1 \right\}. \tag{3.2}$$

The values $p_i$ represent the probabilities associated with the events $A_i$. The Dirichlet distribution thus represents a continuous distribution on the space of $k$-dimensional discrete probability distributions. Let us define the random variables $\{P_{1:k}\}$ as

$$\{P_{1:k}\} \equiv \{P(A_1), \ldots, P(A_k)\} \sim \mathrm{Dir}[\alpha_{1:k}]. \tag{3.3}$$

The expectation of the event probabilities is given by

$$E[P_{1:k}] = \frac{1}{\sum_{i=1}^{k} \alpha_i} \alpha_{1:k}. \tag{3.4}$$

Because the Dirichlet distribution is a conjugate prior, the Bayesian posterior distribution is also a Dirichlet distribution. Specifically, let us consider a set of observation counts $n_{1:k}$ for each event, with $n_i \in \{0, 1, \ldots\}$ and a prior distribution $\mathrm{Dir}[\alpha_{i:k}]$. The posterior distribution is then given by $\mathrm{Dir}[\beta_{1:k}]$ with $\beta_i = \alpha_i + n_i$. This illustrates that the parameters $\alpha_i$ may be considered 'pseudo-observations' on the same footing as the actual observations $n_i$. As a result, the posterior

expection of $P_{1:k}$ conditional on the observations $n_{1:k}$ is

$$E[P_{1:k}|n_{1:k}] = \frac{1}{\sum_{i=1}^{k}(\alpha_i + n_i)}\{\alpha_i + n_i\}_{i=1}^{k}. \tag{3.5}$$

The probability associated with the union of two disjoint events $A_i$ and $A_j$ is given by the sum of their probabilities $p(A_i)$ and $p(A_j)$. In the case of a Dirichlet distribution, these probabilities are random variables $P_i$ and $P_j$. The Dirichlet distribution has the property that the resulting $(k-1)$-dimensional distribution is obtained simply by summing the relevant parameters $\alpha_i$ and $\alpha_j$. Assuming $i < j$ without loss of generality, we write

$$\{P_{1:i-1}, \boldsymbol{P_i} + \boldsymbol{P_j}, P_{i+1:j-1}, P_{j+1:k}\} \sim \mathrm{Dir}[\alpha_{1:i-1}, \boldsymbol{\alpha_i} + \boldsymbol{\alpha_j}, \alpha_{i+1:j-1}, \alpha_{j+1:k}]. \tag{3.6}$$

Repeated application of this process until only the events $A_i$ and its complement $A_i^c$ are left (e.g. 'heads' and 'tails') yields

$$\{P(A_i), P(A_i^c)\} \sim \mathrm{Dir}[\alpha_i, \sum_{j\neq i}\alpha_j]. \tag{3.7}$$

Because $P(A_i^c) = 1 - P(A_i)$ by definition, we may simply consider the marginal distribution of the random variable $P(A_i)$, which is the beta distribution with parameters $\alpha_i$ and $\sum_{j\neq i}\alpha_j$

$$P(A_i) \sim \beta[\alpha_i, \sum_{j\neq i}\alpha_j]. \tag{3.8}$$

## (b) Dirichlet process

Ferguson's Dirichlet process [8] provides a natural extension of the discrete Dirichlet distribution to an infinite set of events, and the continuous domain $\mathbb{R}$ in particular. The Dirichlet process $\mathrm{DP}[m]$ on $\mathbb{R}$ is a distribution of probability distributions (i.e. a random probability measure) that is parametrised by the measure $m : \mathcal{B}(\mathbb{R}) \to \mathbb{R}^+$, where $\mathcal{B}(\mathbb{R})$ is the Borel algebra on $\mathbb{R}$. The Dirichlet process is defined by the condition

$$\mathcal{P} \sim \mathrm{DP}[m] \implies \{\mathcal{P}(A_1), \ldots, \mathcal{P}(A_n)\} \sim \mathrm{Dir}[m(A_1), \ldots, m(A_n)], \tag{3.9}$$

where $\mathcal{P}$ is a random probability measure and $\{A_{1:n}\}$ is any arbitrary disjoint covering of $\mathbb{R}$. It deserves mention that even when the measure $m$ has a continuous density on $\mathbb{R}$, samples from the corresponding Dirichlet process distribution will have discrete features [8] (see also Figure 1).

Like the Dirichlet distribution for discrete distributions, the Dirichlet process on the real line is a conjugate prior for Bayesian inference [8]. Using $\mathcal{P}_0 \sim \mathrm{DP}[m_0]$ as a prior, and $n$ independent observations $\{x_{1:n}\}$, it follows from (3.9) that

$$\mathcal{P}_n \sim \mathrm{DP}\left[m_0 + \sum_{i=1}^{n}\delta_{x_i}\right]. \tag{3.10}$$

Here, $\delta_{x_i}$ are measures corresponding to a unit mass located at $x_i$, within the support set of $m_0$.

## (c) Probability box representation of imprecise probabilities

In probability theory as defined using Kolmogorov's axioms, each event $A$ has a specific probability $p_A$. When our state of knowledge does not permit us to make such precise statements one can resort to imprecise probabilities [10]: instead of the value $p_A$ we assign an interval probability consisting of a pair of lower and upper probability bounds $\underline{p}_A$ and $\overline{p}_A$.

In this paper we consider imprecise probabilities on the real line. The *probability box* [11] (also summarised as p-box), is a particular representation of imprecise probability, defined by a lower

probability distribution $\underline{F}(x)$ and an upper probability distribution $\overline{F}(x)$. We write

$$F^{\diamond} \equiv [\underline{F}, \overline{F}]. \tag{3.11}$$

The most intuitive interpretation of a p-box is that it represents the set of distribution functions enclosed by $\underline{F}$ and $\overline{F}$, without assigning a preference or probability:

$$F^{\diamond} = \{F : \forall x \in \mathbb{R} : \underline{F}(x) \leq F(x) \leq \overline{F}(x); \forall \varepsilon > 0 : F(x + \varepsilon) \geq F(x)\} \tag{3.12}$$

The boundary distributions $\underline{F}$ and $\overline{F}$ directly specify lower and upper probabilities for the elementary events $(-\infty, a]$ and $(b, +\infty)$. Together, these define a *probability interval* for events $X \in (a, b]$. The upper and lower bounds of this interval are given by:

$$\overline{P}_{(a,b]} = \overline{F}(b) - \underline{F}(a) \tag{3.13a}$$

$$\underline{P}_{(a,b]} = \max(\underline{F}(b) - \overline{F}(a), 0). \tag{3.13b}$$

The information contained in a probability box can also be expressed using Dempster-Shafer (DS) theory [11]. In DS theory, probability masses are associated with particular events (e.g. subsets of $\mathbb{R}$). The masses are used to define belief (lower probability) and plausibility (upper probability) functions for arbitrary events. Whereas every probability box can be expressed as a belief function (or, equivalently, a plausibility function), the converse is not true [11]. For example, a probability box cannot be used to represent the knowledge that a random variable $X$ has a known 'exclusion zone' between $a$ and $b$, whereas this feature could be embedded in the DS probability masses. Nevertheless, probability boxes are sufficiently flexible for the purpose of this paper. Furthermore, belief and plausibility functions in DS theory are equivalent to necessity and possibility functions in possibility theory [12]. Recent developments have focused on extending the intuitive appeal of p-boxes beyond the real line, to general preordered spaces [13] and bivariate distributions [14].

## 4. Inference using the Unit Dirichlet Process

In this section we reformulate the Dirichlet process to disentangle its expected distribution and the random fluctuations around that distribution, and we use this formulation to restate the Bayesian posterior distribution for the Dirichlet process.

**Definition** (Cumulative Dirichlet Process). For a given Dirichlet process $\mathcal{P} \sim \mathrm{DP}[m]$ on the real line, the corresponding Cumulative Dirichlet Process (CDP) is parametrised by the cumulative function

$$M(x) = m((-\infty, x]). \tag{4.1}$$

$\mathrm{CDP}[M]$ is defined such that $\mathcal{F} \sim \mathrm{CDP}[M]$ satisfies the cumulative identity

$$\mathcal{F}(x) \stackrel{d}{=} \mathcal{P}((-\infty, x]), \tag{4.2}$$

where $\stackrel{d}{=}$ denotes equality in distribution.

**Lemma 4.1.** *Let* $\mathcal{F} \sim \mathrm{CDP}[M]$, *with* $lim_{x \to \infty} M(x) < \infty$. *We express* $M(x) = \alpha F(x)$ *in terms of a scalar concentration parameter* $\alpha = lim_{x \to \infty} M(x)$ *and a cumulative probability distribution* $F(x)$. *Then the following properties hold:*

$$E[\mathcal{F}] = F. \tag{4.3}$$

$$\lim_{\alpha \to \infty} \mathcal{F} = F. \tag{4.4}$$

*Proof.* $\mathcal{F}(x) = \mathcal{P}((-\infty, x])$, which, following (3.8), is distributed according to a beta distribution $\beta[\alpha F(x), \alpha(1 - F(x))]$. Therefore, $\forall x : E[\mathcal{F}(x)] = F(x)$ so that $E[\mathcal{F}] = F$. In addition, it follows

from the properties of the beta distribution that $\lim_{\alpha\to\infty}\mathrm{Var}(\mathcal{F}(x))=0$, so that $\lim_{\alpha\to\infty}\mathcal{F}=F$.
$\square$

**Definition** (Unit Dirichlet Process). Let the *Unit Dirichlet Process* be the CDP that is generated by the concentration parameter $\alpha$ and the identity function on the unit interval $[0,1]$:

$$\mathcal{U}_\alpha \sim \mathrm{CDP}[\alpha\,\mathrm{Id}_{[0,1]}]. \tag{4.5}$$

The unit Dirichlet process may be interpreted as a random non-decreasing map from the interval $[0,1]$ to $[0,1]$. From (4.3) and (4.4) it follows that $E[\mathcal{U}_\alpha]=\mathrm{Id}_{[0,1]}$ and $\lim_{\alpha\to\infty}\mathcal{U}_\alpha=\mathrm{Id}_{[0,1]}$. Figure 1 shows several realisations of the random distribution $\mathcal{U}_\alpha$. The case $\alpha=10$ clearly demonstrates the discrete nature of the Dirichlet process, whereas $\alpha=1000$ illustrates the approximation of the identity map for $\alpha\to\infty$. Finally, The Unit Dirichlet Process has a tidy expression in terms of the 'stick breaking process' definition of the Dirichlet process [15]:

$$\mathcal{U}_\alpha = \sum_{i=1}^{\infty} H_{U_i} B_i \prod_{j=1}^{i-1}(1-B_j). \tag{4.6}$$

Here $H_c$ is the unit step function at $x=c$, and the sequences of independent random variables $\{U_{1:\infty}\}$ and $\{B_{1:\infty}\}$ are defined as $U_i \sim \mathrm{uniform}[0,1]$ and $B_i \sim \beta[1,\alpha]$.
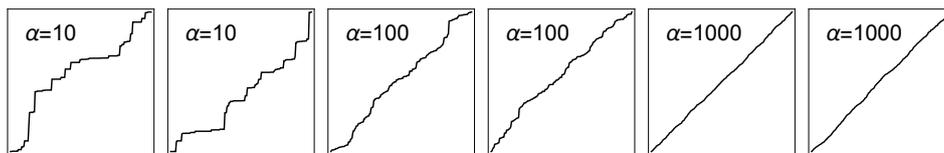


**Figure 1.** Realisations of the identity Dirichlet process $\mathcal{U}_\alpha$. The value of $\alpha$ used to generate the realisation is shown in each panel. Two realisations are shown for each value of $\alpha$, illustrating its intrinsic variability. The plots were generated by discretising the interval $[0,1]$ into 200 intervals and using the discrete Dirichlet distribution $\mathrm{Dir}[\{\alpha/200\}_{i=1}^{200}]$ to sample the associated weights.

**Theorem 4.1.** *Every Cumulative Dirichlet Process* $\mathcal{F} \sim \mathrm{CDP}[\alpha F]$ *can be expressed as*

$$\mathcal{F} = \mathcal{U}_\alpha \circ F \tag{4.7}$$

*where* $\mathcal{U}_\alpha$ *is a Unit Dirichlet Process and the* $\circ$ *symbol indicates composition of functions.*

*Proof.* Consider the cumulative Dirichlet process $\mathcal{F} \sim \mathrm{CDP}[\alpha F]$ and an arbitrary partitioning of the real line into half-open intervals $\{A_{1:n}\}$. This partitioning is defined by the ordered set of points $\{y_{0:n}\}$, with $y_0=-\infty$, $y_n=+\infty$, $A_i=(y_{i-1},y_i)$ for $i=1,\ldots,n-1$, and $A_n=(y_{n-1},y_n)$. Together, the set $\{y_{0:n}\}$ and $\mathcal{F}$ generate the non-decreasing sequence of random variables

$$\{Z_{0:n}\} = \{\mathcal{F}(y_i)\}_{i=0}^{n}. \tag{4.8}$$

The difference $Z_i - Z_{i-1}$ between two adjacent variables is the random probability $\mathcal{P}_\mathcal{F}(A_i)$ associated with interval $A_i$ under $\mathcal{F}$. It follows from (3.9) that

$$\{Z_i - Z_{i-1}\}_{i=1}^{n} = \{\mathcal{P}_\mathcal{F}(A_i)\}_{i=1}^{n} \sim \mathrm{Dir}[\{\alpha(F(y_i)-F(y_{i-1}))\}_{i=1}^{n}]. \tag{4.9}$$

We perform the same steps for $\alpha \circ F$ as above for $\mathcal{F}$, defining

$$\{Z'_{0:n}\} = \{(\mathcal{U}_\alpha \circ F)(y_i)\}_{i=0}^{n} = \{\mathcal{U}_\alpha(F(y_i))\}_{i=0}^{n}. \tag{4.10}$$

Making use of (4.5) and (3.9), it follows that

$$\{Z'_i - Z'_{i-1}\}_{i=1}^n = \{\mathcal{P}_{\mathcal{U} \circ F}(A_i)\}_{i=1}^n \sim \text{Dir}[\{\alpha(F(y_i) - F(y_{i-1}))\}_{i=1}^n]. \tag{4.11}$$

Comparing this result with (4.9) we find that $Z_{1:n} \overset{d}{=} Z'_{1:n}$. As this equality holds for arbitrary $y_{1:n-1}$, we conclude that (4.7) holds for the entire domain.

$\square$

Theorem 4.1 neatly splits the cumulative Dirichlet process $\mathcal{F}$ into the expected shape ($F = E[\mathcal{F}]$) of the distribution and its inherent randomness ($\mathcal{U}_\alpha$). It shows that $\mathcal{F}$ can be understood as a random distortion of the generating distribution $F$ with fluctuations that decrease in amplitude as $\alpha$ increases. The same procedure can be applied to the Bayesian posterior of the Dirichlet process.

**Corollary 4.1.** *Let $\mathcal{F}_0 \sim \text{CDP}[\alpha F_0]$ be a prior CDP. Consider observations $\{x_{1:n}\}$ with an empirical distribution function $\hat{F}_n = \frac{1}{n} \sum_{i=1}^n H_{x_i}$, where $H_c$ denotes the unit step function at $x = c$. Restating (3.10), the posterior CDP is given by*

$$\mathcal{F}_n \sim \text{CDP}\left[\alpha F_0 + \sum_{i=1}^n H_{x_i}\right]. \tag{4.12}$$

*Invoking Theorem 4.1, this can be restated in the form*

$$\mathcal{F}_n = \mathcal{U}_{\alpha+n} \circ F_n, \tag{4.13a}$$

*where*

$$F_n = \frac{\alpha}{\alpha + n} F_0 + \frac{n}{\alpha + n} \hat{F}_n. \tag{4.13b}$$

The implication of the decomposition (4.13) is that the (cumulative) posterior distribution $\mathcal{F}_n$ may be considered a random deformation ($\mathcal{U}_{\alpha+n}$) of a weighted combination of the prior ($F_0$) and empirical ($\hat{F}_n$) distributions.

# 5. Robust inference of distributions

The tools developed in the previous section can be applied to the central challenge in this paper: inferring the distribution $F_X^*$ from a set of $n$ independent real-valued observations $\{x_{1:n}\}$ of $X \sim F_X^*$. Throughout, we will assume that $X$ is restricted to an interval $I = [x_L, x_R] \subset \mathbb{R}$, but it should be noted that this is not a material constraint, as one can take the limits $x_L \to -\infty$ and $x_R \to \infty$ at any time. In practical applications, doing so is rarely required as physical bounds limit the range of $I$. For example, in the motivating example of blackouts, impacts are bounded from below by 0 (no blackout) and from above by the system size (total blackout).

This prior state of knowledge is said to be *vacuous* (subject to $I$): any distribution function on $I$ is considered a candidate for $X^*$. This is represented by the p-box

$$F_0^\diamond = [H_{x_R}, H_{x_L}]. \tag{5.1}$$

Recall that $H_c$ is the unit step function at $x = c$.

**Definition** (P-box transformation). Let $\mathfrak{F}$ be the space of all distribution functions on the interval $I$, and $\mathfrak{F}^\diamond$ the space of probability boxes on $I$. Let $\mathfrak{U}$ be the set of all non-decreasing unit maps $U : [0,1] \to [0,1]$ with $U(0) = 0$ and $U(1) = 1$. For every $U \in \mathfrak{U}$ we define (with a slight abuse of notation) an identically named map $U : \mathfrak{F}^\diamond \to \mathfrak{F}^\diamond$ so that

$$U \circ F^\diamond \equiv [U \circ \underline{F}, U \circ \overline{F}], \tag{5.2}$$

for any probability box $F^\diamond = [\underline{F}, \overline{F}]$, with $\underline{F}, \overline{F} \in \mathfrak{F}$. The map $U \in \mathfrak{U}$ is order-preserving, so the transformed upper and lower bound distributions are sufficient to define the transformed p-box. In the following, the appropriate domain of the map $U$ ([0,1] or $\mathfrak{F}^\diamond$) will be clear from the context.

The Unit Dirichlet Process $\mathcal{U}_\alpha \sim \mathrm{CDP}[\alpha \, \mathrm{Id}_{[0,1]}]$ is a random element of $\mathfrak{U}$. It therefore defines a p-box randomisation in an analogous manner:

$$\mathcal{U}_\alpha \circ F^\diamond = [\mathcal{U}_\alpha \circ \underline{F}, \mathcal{U}_\alpha \circ \overline{F}] \tag{5.3}$$

Note that the UDP acts identically on the lower and upper bounds (each realisation transforms both identically).

## (a) Robust posterior distributions

**Theorem 5.1.** *Define the s-robust posterior distribution as*

$$\mathcal{F}_n^{\diamond(s)} = \mathcal{U}_{n+s} \circ F_n^{\diamond(s)}, \qquad \text{with} \tag{5.4}$$

$$F_n^{\diamond(s)} = \frac{s}{n+s} F_0^\diamond + \frac{n}{n+s} \hat{F}_n. \tag{5.5}$$

*Then:*

(i) *$\mathcal{F}_n^{\diamond(s)}$ is a posterior distribution corresponding to the vacuous prior $F_0^\diamond$ (5.1) with concentration parameter $s$ and observations $x_{1:n}$.*
(ii) *$\mathcal{F}_n^{\diamond(s)}$ is a consistent estimator that converges to $F_X^*$ as $n \to \infty$.*

*Proof.* (i) The limiting distributions of the vacuous prior $F_0^\diamond$ are improper, with no probability mass outside the points $x_L$ and $x_R$. Formally, use of these priors prevents learning [10], but we consider them as envelopes for a dense set of proper priors. The posterior distribution $\mathcal{F}_n^{\diamond(s)}$ is itself the envelope of the posterior distributions corresponding to the family of proper priors with weight $s$. The derivation is contained in Section S4 of the Supplementary Material. (ii) Consistency follows from the fact that $\lim_{n\to\infty} F_n^{\diamond(s)} = \hat{F}_n = F_X^*$ and $\lim_{n\to\infty} \mathcal{U}_{n+s} = \mathrm{Id}_{[0,1]}$, thus $\lim_{n\to\infty} \mathcal{F}_n^\diamond = F_X^*$. $\qquad\square$

**Remark.** In the limit $s \downarrow 0$ the upper and lower distributions coalesce into the single random distribution

$$\mathcal{F}_n^{(s=0)} = \mathcal{U}_n \circ \hat{F}_n \sim \mathrm{CDP}\left[n\hat{F}_n\right]. \tag{5.6}$$

This is the Bayesian Bootstrap method introduced by Rubin [7] as a Bayesian analog of the bootstrap method. It assigns a (random) probability mass only to observed values $x_{1:n}$, so it has no ability to extrapolate beyond the lowest and highest observed values, and the posterior is always discrete.

**Remark.** Conceptually, the $s$-robust posterior distribution is closely related to the Robust Bayesian approach [16], but it differs in the way it embeds continuous functions in the posterior. Using a prior set $F_0 \in F_0^\diamond$ and concentration parameter $s$, the Robust Bayesian approach considers the set of posteriors given by $\mathcal{U}_{n+s} \circ F$, with $F \in F_n^{\diamond(s)}$. Although the span of these posteriors is identical to that of the $s$-robust posterior distribution, the discrete nature of the Dirichlet process means that members of the posterior distribution are smooth with probability zero. In contrast, the $s$-robust posterior distribution is a random p-box. Although its bounds have discrete features, continuous functions that fit within the p-box are equally part of the posterior distribution.

The parameter $s$ determines the relative weight of the prior and observations. It can therefore be considered the inverse of an effective 'learning rate': small values result in faster convergence, but picking a value that is too small leads to an underappreciation of parts of the distribution that have not been observed, as represented by the vacuous prior. Theorem 5.2 shows that choosing $s < 1$ imposes essential discontinuities on the inferred distribution. Hence, to maximise the learning rate without precluding continuous posteriors, $s = 1$ will henceforth be used.

**Theorem 5.2.** *Let $\mathcal{F}_n^{\diamond(s)}$ be an s-robust posterior distribution. The posterior is compatible with continuous distributions $F_X^*$ only if $s \geq 1$.*

*Proof.* If the original distribution $F_X^*$ is (locally) continuous, the observations $\{x_{1:n}\}$ are almost surely distinct. Take the $i$-th smallest observation $x_{(i)}$ and consider the posterior distributions at $x_{(i)} \pm \varepsilon$. The *upper* bound distribution at $x_{(i)} - \varepsilon$ is given by

$$U_i^{(s)} = \lim_{\varepsilon \downarrow 0} \overline{\mathcal{F}}_n^{(s)}(x_{(i)} - \varepsilon) = \mathcal{U}_{n+s}\left(\frac{s+i-1}{n+s}\right) \sim \beta[i-1+s, n-i+1], \tag{5.7}$$

where (3.8) was used to specify $U_i^{(s)}$ in terms of the beta distribution. Similarly, for the lower bound distribution at $x_{(i)} + \varepsilon$ we find

$$L_i^{(s)} = \lim_{\varepsilon \downarrow 0} \underline{\mathcal{F}}_n^{(s)}(x_{(i)} + \varepsilon) = \mathcal{U}_{n+s}\left(\frac{i}{n+s}\right) \sim \beta[i, n-i+s]. \tag{5.8}$$

To avoid biasing the posterior toward distributions with a discontinuity at $x_{(i)}$, it is necessary that $U_i^{(s)}$ (left upper bound) equals or exceeds $L_i^{(s)}$ (right lower bound), for all realisations of the random bounds, and for all $i$. This can only be the case if every quantile of $L_i^{(s)}$ is greater than or equal to the corresponding quantile of $U_i^{(s)}$ (first order stochastic dominance [17], denoted as $U_i^{(s)} \succeq L_i^{(s)}$). Inspection of the beta distributions in (5.7) and (5.8) results in the necessary condition $s \geq 1$.

That this condition is also sufficient follows by inserting $s = 1$ into (5.7) and (5.8). This results in the identity

$$U_i^{(1)} = L_i^{(s)} = \mathcal{U}_{n+1}\left(\frac{i}{n+1}\right), \tag{5.9}$$

which trivially satisfies $U_i^{(1)} \succeq L_i^{(s)}$. $\qquad \square$

**Definition.** The *robust posterior distribution* $\mathcal{F}_n^{\diamond}$ is defined as the $s$-robust posterior distribution with $s = 1$.

$$\mathcal{F}_n^{\diamond} = \mathcal{U}_{n+1} \circ F_n^{\diamond}, \qquad \text{with} \tag{5.10}$$

$$F_n^{\diamond} = \frac{1}{n+1}F_0^{\diamond} + \frac{n}{n+1}\hat{F}_n. \tag{5.11}$$

The 'input' p-box $F_n^{\diamond}$ can be considered the spanning distribution of empirical distributions generated by the observations $\{x_{1:n}\}$ and one unknown observation in the interval $I$. Figure 2 (left) depicts $F_n^{\diamond}$ for the observations in Table 1. The right panel shows four random realisations of $\mathcal{F}_n^{\diamond}$. The realisations visually resemble a 'chain of blocks' with the lower and upper bounds touching at each observation $x_i$. These upper and lower bound pairs force compatible distributions to traverse particular values at $x_i$ but leave them otherwise unconstrained, in line with theorem 5.2. This provides a further justification for p-box transformation (5.3): any other simultaneous transformation of the upper and lower bounds (with $s = 1$) would have necessarily implied a discontinuity in the distributions compatible with the posterior realisation. The choice $s = 1$ also ensures that an expected posterior probability of $1/(n+1)$ is assigned to each of the tail intervals $[x_L, x_{(1)}]$ and $[x_{(n)}, x_R]$. Further properties of $F_n^{\diamond}$, including comparisons with other methods, are discussed in the Supplementary Materials, Section S3.

## (b) Interval-associated posterior probabilities

The upper and lower bounds of the p-box $F_n^{\diamond}$ only take on $n+2$ distinct and equally spaced values. As a result, the (random) upper and lower values of the random probability box $\mathcal{U}_{n+1} \circ F_n^{\diamond}$ are given by $\mathcal{U}_{n+1}(v_i)$, with $v_i = i/(n+1)$ and $i = 0, \ldots, n+1$. This partitioning allows for

**Table 1.** Sample data. These 15 observations are drawn independently from the log-normal distribution $\ln \mathcal{N}(0, 1)$. For the robust analysis it is assumed that $I = [0, \infty)$.

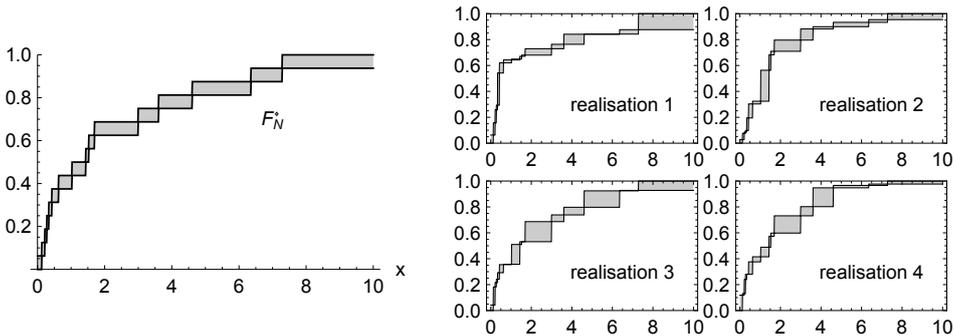| 1.435 | 0.276 | 3.603 | 0.211 | 2.996 |
|-------|-------|-------|-------|-------|
| 7.289 | 0.426 | 0.124 | 1.523 | 4.603 |
| 1.696 | 0.620 | 0.338 | 6.351 | 1.026 |



**Figure 2.** The probability box $F_n^{\diamond}$ (left) has been constructed from the sample data in Table 1 and the interval $[0, \infty)$. Also shown (right) are four realisations of $\mathcal{F}_n^{\diamond}$ resulting from the random mapping of $F_n^{\diamond}$ by $\mathcal{U}_{n+1}$.

the projection of (4.5) onto an $n + 1$-dimensional Dirichlet distribution. Making use of (3.9), we can specify $\mathcal{F}_n^{\diamond}$ directly as a random p-box:

$$\mathcal{F}_n^{\diamond} = \left[ \sum_{i=1}^{n+1} W_i H_{x_{(i)}}, \sum_{i=1}^{n+1} W_i H_{x_{(i-1)}} \right] \tag{5.12}$$

with $\{W_{1:n+1}\} \sim \mathrm{Dir}[1, \ldots, 1]$ and $\{x_{(1):(n)}\}$ is the ordered data set $\{x_{1:n}\}$, augmented by $x_{(0)} = x_L$ and $x_{(n+1)} = x_R$.

The fact that the robust posterior distribution $\mathcal{F}_n^{\diamond}$ can be expressed in this simple manner has significant implications. Eq. (5.12) provides an intuitive understanding of the method that has been developed. The interval $I$ is partitioned into $n + 1$ closed intervals with boundaries at the observed points $\{x_i\}_{i=1}^n$. Each of these *intervals* is assigned a random weight $W_i$, drawn from a Dirichlet distribution. Note that this Dirichlet distribution with $n + 1$ unit parameters is in fact the uniform distribution on the unit simplex $\Delta^n$, as can be seen from Eqs. (3.1) and (3.2).

Realisations of the posterior distribution $\mathcal{F}_n^{\diamond}$ (Figure 2 (right)) have a natural interpretation in the Dempster-Shafer theory of belief functions [11]: a probability mass $W_i$ is assigned to the elementary events $[x_{(i-1)}, x_{(i)}]$, from which a belief function on $I \subseteq \mathbb{R}$ can be constructed. This assignment is inherently 'imprecise' in the following sense. The elementary events have overlapping boundaries (the points $x_{(i)}$), so that there is no precise probability associated with the *open* intervals $(x_{(i-1)}, x_{(i)})$ or the points $x_{(i)}$. As a result, the set of probability events does not form a $\sigma$-algebra on $I$ (it is not closed under complementation), and one cannot define an underlying 'precise' probability model that could be extended to $I$ by means of inner/outer measure models [18].

Although a specific probability distribution governs the probability mass assigned to each interval, the method provides no guidance regarding the way this mass is distributed *within* each interval. Intuitively, this dichotomy can be interpreted as follows. The limited number of observations can only provide meaningful information about the large scale features (probability

associated with intervals) of the probability distribution $F_X^*$. Without additional observations or assumptions no substantiated statements can be made about the features of $F_X^*$ on a smaller scale.

# 6. Estimating population parameters

In practical applications the object of interest is often not the distribution $F_X^*$, but a particular real-valued function $q^* = q[F_X^*]$ of that distribution, such as the expectation, median or various measures of tail risk. In the context of this paper, we restrict ourselves to the case where $q$ is *monotonic*.

**Definition.** A function $q : \mathfrak{F} \to \mathbb{R}$ is monotonic if the following is true for any two random variables $Y \sim F_Y$ and $Z \sim F_Z$:

$$Y \succeq Z \Rightarrow q[F_Y] \geq q[F_Z]. \tag{6.1}$$

Where the partial ordering '$\succeq$' denotes first order stochastic dominance [17], which is equivalent to the statement $\forall x \in \mathbb{R} : F_Y(x) \leq F_Z(x)$, with a strict inequality holding for at least one point.

The strong condition of monotonicity holds for many basic population parameters, including the mean, quantiles (value-at-risk), and truncated means (conditional value-at-risk). More generally, it holds for the class of coherent risk measures [19], although a customary minus sign is usually incorporated into the definition of monotonicity.

Consider the set of permitted $q$-values for a given p-box $F^\diamond = [\underline{F}, \overline{F}]$, which we summarise by its minimum and maximum values. Due to the monotonicity of $q$, we have $[q^{\min}, q^{\max}] = \left[q[\overline{F}], q[\underline{F}]\right]$. For the *random* p-box $\mathcal{F}_n^\diamond$, this induces the random $q$-interval

$$[Q^{\min}, Q^{\max}] = \left[q[\overline{\mathcal{F}}_n], q[\underline{\mathcal{F}}_n]\right]. \tag{6.2}$$

The distributions of the bounding random variables $Q^{\min}$ and $Q^{\max}$ can be interpreted as the definition of a probability box for the population parameter $q^*$ (see also Figure 3):

$$F_Q^\diamond = [F_{Q^{\max}}, F_{Q^{\min}}] = [F_{q[\underline{\mathcal{F}}_n]}, F_{q[\overline{\mathcal{F}}_n]}]. \tag{6.3}$$

**Reporting intervals** In applications, it is usually preferable to summarise the estimate of a population parameter $q^* = q[F_x^*]$ in the form $[q_{\min}, q_{\max}]_c$, implicitly defined by

$$\Pr(q^* \in [q_{\min}, q_{\max}]_c) \geq c \tag{6.4}$$

for a given credibility level $c \leq 1$ (e.g. 95%). In keeping with our robust approach, this is achieved by taking the *span* of the $c$-credible values for the upper and lower bound processes, resulting in

$$[q_{\min}, q_{\max}]_c = \left[F_{Q^{\min}}^{-1}\left(\frac{1-c}{2}\right), F_{Q^{\max}}^{-1}\left(\frac{1+c}{2}\right)\right]. \tag{6.5}$$

The construction of the interval estimate is illustrated in Figure 3. Under the stated assumptions (independence of observations and the range of $X$ being confined to $I$) the credibility $c$ is a *lower bound* for the probability that the real value is contained in the interval estimate. To emphasise both the accuracy and the inherent conservativeness of the interval estimate, this can be expressed as a "$c+$ credible interval" (e.g. 95%+ credible interval).

# 7. Bayesian Interval Sampling

The posterior distribution $\mathcal{F}_n^\diamond$ as expressed in (5.12) associates random probabilities $W_{1:n+1} \sim \text{Uniform}(\Delta^n)$ with the intervals $[x_{(i-1)}, x_{(i)}]$, for $i = 1, \ldots, n+1$. This formulation suggests a computationally straightforward resampling method that can be used to probe the upper and lower bounds of $\mathcal{F}_n^\diamond$, and therefore (through (6.2)) the distributions of $Q^{\min}$ and $Q^{\max}$. The algorithm, *Bayesian Interval Sampling* (BIS), is described below.
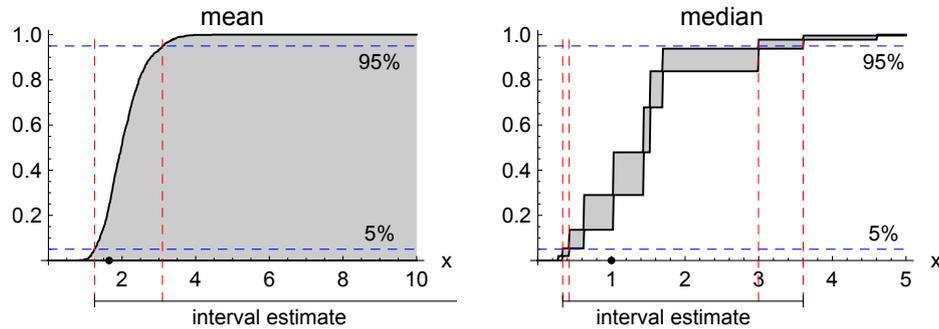
**Figure 3.** Probability boxes for the mean $\mu$ and median $m$ of $X$, determined using the sample data in Table 1, the bounding interval $I = [0, \infty)$ and $n_{\text{resample}} = 1000$. The construction of the 90% credibility interval ($c = 0.9$) is illustrated as follows: dashed blue lines (horizontal) indicate the 5% and 95% quantiles, and dashed red lines (vertical) the corresponding values of $\mu$ and $m$. In case of the median (right), the two sets of vertical lines correspond to values generated using the upper and lower bounds of the random p-box $\mathcal{F}_n^\diamond$. The black dot indicates the true mean and median of the originating distribution.

(i) For $i = 1$ to $n_{\text{resample}}$

    (a) Sample an $(n+1)$-dimensional weight vector $\{w_{1:n+1}^{(i)}\} \in \Delta^n$ from $\text{Dir}\,[1, \ldots, 1]$

    (b) Compute $q_i^{\min} = q\left[\frac{1}{n+1} \sum_{j=1}^{n+1} w_j^{(i)} H_{x_{(j-1)}}\right]$

    (c) Compute $q_i^{\max} = q\left[\frac{1}{n+1} \sum_{j=1}^{n+1} w_j^{(i)} H_{x_{(j)}}\right]$

(ii) Compute $\hat{F}_{Q^{\min}}$ as the empirical CDF of $\{q_{1:n_{\text{resample}}}^{\min}\}$

(iii) Compute $\hat{F}_{Q^{\max}}$ as the empirical CDF of $\{q_{1:n_{\text{resample}}}^{\max}\}$

(iv) Compute the $c$-credible interval estimate $[q_{\min}, q_{\max}]_c$ using (6.5)

As a rule of thumb we use $n_{\text{resample}} = 100/(1-c)$ to ensure that we have 100 observations in the tails of the empirical distributions for $Q^{\min}$ and $Q^{\max}$. This ensures reasonable statistical stability for the computation of the interval $[q_{\min}, q_{\max}]_c$. A modified algorithm for data sets with many identical observations is given in the Supplementary Materials, Section S1.

**Remark.** Bayesian Interval Sampling also has a natural interpretation in terms of the Bayesian bootstrap method [7], a Bayesian variation of the (frequentist) bootstrap method [6]. The connection between the algorithm above and the Bayesian bootstrap method is readily apparent through comparison of (5.6) and (6.3). The upper bound distribution $\hat{F}_{Q^{\min}}$ for $q$ is the Bayesian bootstrap posterior distribution for the dataset $x_{1:n}$ augmented with a 'pseudo-observation' of the lower domain boundary $x_L$. Similarly, $\hat{F}_{Q^{\min}}$ is the Bayesian bootstrap posterior distribution for $x_{1:n}$ augmented with the upper domain boundary $x_R$. In Bayesian Interval Sampling, both bounds are sampled simultaneously, guaranteeing sample-wise coherence ($q_i^{\min} \leq q_i^{\max}$).

# 8. Examples

Bayesian Interval Sampling was used to analyse the data in Table 1, using the bounds $I = [0, \infty)$. The population parameters considered were the median and the mean, resulting in the probability boxes in Figure 3. The probability box for the median (right) is well-behaved and results in a finite interval estimate. In contrast, the equivalent analysis for the mean (left) does not result in a finite interval estimate. This is caused by the lower bound distribution for the mean, which has a constant value of $F_{\mu^{\max}} = 0$. This perhaps surprising result is a consistent outcome of our

**Table 2.** Comparison of methods for truncated log-normal distribution, with and without additional extreme events. Methods were applied as follows: (i) Student $t$ distribution: 95% confidence interval based on the sample average and variance using Student's $t$ distribution with $n_{\text{sample}} - 1$ degrees of freedom. (ii) Bootstrap method: 95% confidence interval based on bootstrap resampling with $n_{\text{resample}} = 2000$. (iii) Bayesian Interval Sampling (BIS): 95% credible interval based on $n_{\text{resample}} = 2000$ and the bounding interval $[0, 50]$.

| distribution | method | accuracy (target: 95%) | median $q_{\min}$ | median $q_{\max}$ |
|---|---|---|---|---|
| Truncated log-normal (mean: 1.65) | Student t | 90.3% | 1.08 | 2.12 |
| | Bootstrap | 90.1% | 1.15 | 2.14 |
| | BIS | 98.7% | 1.17 | 5.10 |
| Truncated log-normal with extreme events (mean: 2.13) | Student t | 68.9% | 0.98 | 2.60 |
| | Bootstrap | 70.0% | 1.20 | 2.64 |
| | BIS | 98.8% | 1.26 | 5.42 |

robust approach. No assumption was made on the long tail behaviour of $F_X^*(x)$, other than a data-driven bound on the probability $Pr(X > x_{(n)})$. Therefore one cannot rule out the existence of a small probability mass at $x \to +\infty$, which makes an arbitrarily large contribution to the mean. This result also reflects the common understanding that the mean is not a robust statistic, whereas the median is. Further examples for the Value at Risk (VaR) and truncated mean are included in the Supplementary Materials, Section S2.

To illustrate the robustness of the Bayesian Interval Sampling method, its performance was analysed for the estimation of the mean in two variants of the (0,1)-log-normal distribution. The first variant truncated the distribution to the interval [0,50]. In realistic cases, such a truncation could reflect the maximum size of $X$ imposed by a finite system size, for example the highest possible cost of a system malfunction. The second variant took the truncated distribution and additionally included a 1% probability to observe the maximum value $x = 50$ (equal to the upper bound of the truncated distribution). The inference performance on these distributions was assessed using 10,000 repeated experiments. For each experiment, 50 random observations were generated and used to compute an interval estimate for the mean. This interval was compared with the true mean of the distribution (1.65 and 2.13, respectively), and a tally was kept to determine the overall accuracy of the method (i.e. proportion of correct predictions). Note that this is essentially a frequentist test, so both (frequentist) *confidence* and (Bayesian) *credible* intervals should satisfy (6.4). The accuracy of BIS was compared with the common Student $t$ and bootstrap approaches

The results are shown in Table 2. Both the Student t and bootstrap methods have success rates well below 95%, so neither method satisfies the objective stated in (6.4). This is a direct consequence of the long tail of the distribution, the impact of which is significantly exacerbated in the extreme event distribution. In contrast, the Bayesian interval sampling method has much lower misprediction rates of only 1.3% and 1.2%, well below the 5% bound imposed by the 95% credibility requirement. The rightmost columns of Table 2 show the median values of the lower and upper interval bounds (across 10,000 experiments). They illustrate that the improved accuracy of BIS is a direct consequence of more conservative interval estimates.

**Estimation of blackout risks** Finally, BIS was applied to the motivating example of estimating blackout risks in power systems. A simple minimal electrical network model was generated using a modified Barabási-Albert model with non-preferential random attachment of two edges (transmission lines) for each placed node ('model A' in [20]). A 1000-node network was grown, starting from a single node, resulting in 1998 lines. Customer loads were uniformly allocated to nodes, with a value of 1 (in arbitrary units) assigned to each node. Generating capacity equal to the total load was randomly distributed across the nodes according to a Dirichlet distribution $\text{Dir}[1/9, \ldots, 1/9]$. The power flows across the transmission lines were computed using the DC

**Table 3.** Computation of loss-of-load risk. Stratified sampling was used for $k = 2, \ldots, 5$, on the basis of 10,000 $k$-line outages each. Robust 99%+ estimates of the mean loss were computed using the Bayesian interval sampling method with $n_{\mathrm{resample}} = 10,000$.

| $k$ | $\Pr(k)$ | mean fraction of load lost (99%+ interval estimate) | risk contribution ($\Pr(k) \times$ loss of load) |
|---|---|---|---|
| 0, 1 | 0.98 | 0 (by design) | 0 |
| 2 | $1.6 \times 10^{-2}$ | $[1.5 \times 10^{-5}, 6.1 \times 10^{-4}]$ | $[2.5 \times 10^{-7}, 9.9 \times 10^{-6}]$ |
| 3 | $1.1 \times 10^{-3}$ | $[4.7 \times 10^{-5}, 5.9 \times 10^{-4}]$ | $[5.1 \times 10^{-8}, 6.4 \times 10^{-7}]$ |
| 4 | $5.4 \times 10^{-5}$ | $[2.0 \times 10^{-4}, 1.0 \times 10^{-3}]$ | $[1.1 \times 10^{-8}, 5.6 \times 10^{-8}]$ |
| 5 | $2.2 \times 10^{-6}$ | $[3.0 \times 10^{-4}, 1.3 \times 10^{-3}]$ | $[6.5 \times 10^{-10}, 2.7 \times 10^{-9}]$ |
| >5 | $7.4 \times 10^{-8}$ | [0,1] (no assumptions) | $[0, 7.4 \times 10^{-8}]$ |
| Total | 1 | | $[3.1 \times 10^{-7}, 1.1 \times 10^{-5}]$ |

power flow approximation, assuming identical reactances for all lines. Line capacities were initialised according to the common 'N-1' security criterion, enforcing that single line failures do not lead to overloads. An additional 10% margin was added to generator capacities and transmission lines.

For this random network, we investigated the risk (expected impact) of line outages that occur independently with a probability $\mathrm{P_{outage}} = 10^{-4}$ (in a given period of interest). The simultaneous occurrence of $k > 1$ line outages can result in overloads in other lines: overloaded lines are tripped one by one until no more overloads are present. Simultaneously overloaded lines are tripped one by one in random order. If the network breaks up into multiple components, supply and demand are balanced in each island individually. If insufficient local generating capacity is available, load is shed uniformly across the nodes in the island. The total fraction of load shed is used as an indicator for the severity of the cascading event.

The analysis was structured by conditioning on the number of simultaneous outages $k$. By design, $k = 0$ and $k = 1$ have no impact, and $k > 5$ is treated as a potential full blackout (load loss can take any value in $[0, 1]$), without further analysis. The impact of this assumption is limited by the small probability associated with $k > 5$ outages. For intermediate $k$, the mean loss of load was estimated using 10,000 simulated $k$-line outages. 99% credible intervals were computed using BIS. The intermediate results and the computation of the overall 99%+ risk estimate ($[3.1 \times 10^{-7}, 1.1 \times 10^{-5}]$) are shown in Table 3.

Current standards for power system reliability are based on deterministic computational procedures. As systems grow in complexity, the use of sampling-based procedures may become inevitable. For regulatory purposes, this should be accompanied by guaranteed probabilistic error bounds of the type (6.4). The Bayesian Interval Sampling method may provide an appropriate framework for such calculations, especially when there is reason to suspect the presence of undetected extreme events.

## 9. Discussion

We have described a robust nonparametric Bayesian method to estimate properties of a distribution function $F_X^*(x)$ from a set of independent observations $\{x_{1:n}\}$. It is formulated on a closed interval $I$, but the limit to an infinite interval can be taken to effectively remove this constraint.

The resulting posterior distribution $\mathcal{F}_n^\diamond$ (5.12) is expressed as a random probability box. It has the intriguing feature that posterior probabilities are associated with the *intervals* between observations, but the allocation of probability mass *within* each interval is left undetermined. Therefore, one can consider $\mathcal{F}_n^\diamond$ as a probabilistic estimate of only the low-frequency (coarse grained) content of $F_X^*(x)$. As expected, the resolution of this low-frequency estimate improves with the sample size $n$.

In applications, one often aims to infer real-valued monotonic properties $q$ of $F_X^*(x)$ (e.g. mean, median, any coherent risk measure). We have shown how the posterior distribution $\mathcal{F}_n^\diamond$ induces a probability box for the estimation of $q* = q[F_X^*]$. In turn, this probability box can be used to define a robust credible interval for $q^*$. Due to the robust construction, this interval may be interpreted as the range of values for $q$ that is *not ruled out* by the observations. The Bayesian Interval Sampling (BIS) method described in Section 7 provides a straightforward resampling algorithm to perform this estimation.

The imprecise nonparametric Bayesian method (and the BIS implementation) is uniquely suitable for applications that require strict bounds on the accuracy of confidence or credible intervals, where (i) the set of (relevant) data points is small (roughly 100 or less) or (ii) unobserved events can have a disproportionate effect on the quantity that is being estimated.

When only a small number of data points is available, the set of observations $\{x_{1:n}\}$ may be significantly skewed compared to the generating distribution $F_X^*(x)$. The nonparametric Bayesian approach results in an inferred distribution for $F_X^*(x)$ that correctly accounts for such small sample variations. It should be noted that this same argument also applies when the total number of samples is large, but the number of *relevant* samples is small. This occurs when one characterises rare failures in reliable systems, for example electrical blackouts in power systems.

Furthermore, the interval probability approach provides a robust framework to deal with unobserved events. Such events may be so-called high-impact low-probability events, which have a disproportionate impact on the population parameter $q^*$. In applications where it is necessary to analyse largely unobserved tails of a distribution it is generally advisable to use Extreme Value Theory (EVT) (see e.g. Coles [21]). However, the application of EVT assumes that the long-tail behaviour of the process is sufficiently well-described by the data set. This may not be the case if the data set is very small, or when the underlying process is prone to exhibit very rare events that are qualitatively different from other large events. Events of the latter type are sometimes referred to as 'Dragon kings' [22].

The method described in this paper is conservative by design, and the resulting uncertainty bands may be dishearteningly large (see e.g. Figure 3). Conventional methods will usually provide estimates with tighter bounds, and their use may well be justified if the underlying distribution $F_X^*(x)$ is well-behaved. However, the robust posterior distribution (and BIS implementation) can still be used for comparison purposes. Significant differences in results will highlight the impact of explicit and implicit assumptions made in the analysis, pointing out the need to confirm their validity in a particular application.

**Multivariate applications** A natural extension of this work is to consider the case of multivariate samples. In its full generality, this is far from trivial, because a number of convenient properties that accompany working on the real line no longer apply. In particular, the total ordering of real numbers, the convenient cumulative form of the Dirichlet Process and the fact that upper and lower bounds of p-boxes are each valid cumulative distribution functions.

The definition and properties of multivariate p-boxes are a topic of ongoing research [13]. A general account of the bivariate case has recently been given by Pelessoni *et al.* [14]. Retracing the construction of the robust posterior distribution $\mathcal{F}_n^\diamond$ for the bivariate case will require addressing a number of challenges: 1) defining a random posterior p-box along the lines of (5.3) and Theorem 5.1; 2) maximising the learning rate (minimising the impact of the prior), while ensuring that the p-box is compatible with continuous distributions (see Theorem 5.2 for the one-dimensional case); 3) defining a practical numerical procedure for sampling from the posterior distribution and determining the impact on quantities of interest.

However, results of Troffaes and Destercke [13] suggest that the extension to multivariate spaces is significantly simpler when a total preorder is defined on elements of the space. A prominent example is the analysis of Monte Carlo samples when importance sampling has been used to reduce sample variance. Instead of unbiased random values $V^u$, the process generates biased value-weight pairs $(V^b, W) \in \mathcal{V} \times \mathcal{W}$. The quantity of interest in this case is $E[V^b W]$,

which equals $E[V^u]$ by design. The magnitude of the product $V^bW$ induces a preorder on $\mathcal{V} \times \mathcal{W}$, which can be used to define effective univariate p-boxes. In practice, this means a posterior distribution for $V^bW$ can be constructed on the basis of product observations $v_i^bw_i$.

**Authors' Contributions.** ST conceived of and performed the mathematical aspects of the study, and drafted the manuscript. Both authors developed the research questions and applications to electricity networks, and read and approved the manuscript.

**Data Accessibility.** All data required to reproduce the results in this paper is either included in the main text or procedurally generated using methodology described in the text.

**Competing Interests.** The authors declare that they have no competing interests.

# References

1. Alizadeh Mousavi O, Cherkaoui R, Bozorg M.
   Blackouts risk evaluation by Monte Carlo Simulation regarding cascading outages and system frequency deviation.
   Electric Power Systems Research. 2012 aug;89:157–164.
2. Henneaux P, Labeau PE, Maun JC.
   A level-1 probabilistic risk assessment to blackout hazard in transmission power systems.
   Reliability Engineering and System Safety. 2012;102:41–52.
3. Rezaei P, Hines PDH, Eppstein MJ.
   Estimating Cascading Failure Risk With Random Chemistry.
   IEEE Transactions on Power Systems. 2015 sep;30(5):2726–2735.
4. Dobson I, Carreras BA, Newman DE.
   How Many Occurrences of Rare Blackout Events Are Needed to Estimate Event Probability?
   IEEE Transactions on Power Systems. 2013 aug;28(3):3509–3510.
5. Robert CP, Casella G.
   Monte Carlo Statistical Methods.
   Springer-Verlag GmbH; 1999.
6. Efron B.
   Bootstrap Methods: Another Look at the Jackknife.
   The Annals of Statistics. 1979 jan;7(1):1–26.
7. Rubin DB.
   The Bayesian Bootstrap.
   The Annals of Statistics. 1981 jan;9(1):130–134.
8. Ferguson TS.
   A Bayesian Analysis of Some Nonparametric Problems.
   The Annals of Statistics. 1973 mar;1(2):209–230.
9. Coolen FPA.
   On Nonparametric Predictive Inference and Objective Bayesianism.
   Journal of Logic, Language and Information. 2006 apr;15(1-2):21–47.
10. Walley P.
    Statistical Reasoning with Imprecise Probabilities.
    Chapman and Hall/CRC; 1991.
11. Ferson S, Kreinovich V, Ginzburg L, Myers DS, Sentz K.
    Constructing Probability Boxes and Dempster-Shafer Structures.
    Sandia National Laboratories; 2003. January.
12. Flage R, Aven T, Baraldi P, Zio E.
    An imprecision importance measure for uncertainty representations interpreted as lower and upper probabilities, with special emphasis on possibility theory.
    Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability. 2012 dec;226(6):656–665.

13. Troffaes M, Destercke S.
    Probability boxes on totally preordered spaces for multivariate modelling.
    International Journal of Approximate Reasoning. 2011;52(6):767–791.

14. Pelessoni R, Vicig P, Montes I, Miranda E.
    Bivariate p-boxes.
    International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems. 2016 apr;24(02):229–263.

15. Sethuraman J.
    A Constructive Definition of Dirichlet Priors.
    Statistica Sinica. 1994;4:639–650.

16. Berger JO.
    An overview of robust Bayesian analysis.
    Test. 1994 jun;3(1):5–124.

17. Hadar J, Russell WR.
    Rules for Ordering Uncertain Prospects.
    The American Economic Review. 1969;59(1):25–34.

18. Fagin R, Halpern JY.
    Uncertainty, belief, and probability'.
    Computational Intelligence. 1991;6:160–173.

19. Artzner P, Delbaen F, Eber JM, Heath D.
    Coherent Measures of Risk.
    Mathematical Finance. 1999 jul;9(3):203–228.

20. Barabási AL, Albert R, Jeong H.
    Mean-field theory for scale-free random networks.
    Physica A: Statistical Mechanics and its Applications. 1999;272(1):173–187.

21. Coles S.
    An Introduction to Statistical Modeling of Extreme Values.
    Springer; 2001.

22. Sornette D, Ouillon G.
    Dragon-kings: Mechanisms, statistical methods and empirical evidence.
    The European Physical Journal Special Topics. 2012 may;205(1):1–26.