Supplementary information
'Robust estimation of risks from small samples'

Simon H. Tindemans
Goran Strbac

Imperial College London
s.tindemans@imperial.ac.uk

Contents

S1 Bayesian interval sampling with multiplicity

The Bayesian Interval Sampling algorithm requires the generation of $n_{\text{resample}} \times (n+1)$ random weights from the Dirichlet distribution. When the sample size n is large the resampling procedure may require a significant amount of CPU time. This is true in particular for rare event simulations in which a large majority of observations does not contribute to the observable (e.g. the cost of system malfunctions in a very reliable system, which is 0 for the overwhelming majority of cases). For such cases the required sample size n can be in the millions.

However, the inference process can be made considerable more efficient if the data set contains duplicate observations. Let us consider the extended set of order statistics $\{x_{(0):(n+1)}\}$ with $x_{(0)} = x_L$ and $x_{(n+1)} = x_R$, and assume that $x_{(k)}$ to $x_{(k+m)}$ are $m+1$ identical observations. We define the reduced data set

$$\{z_{0:n-m+2}\} = \{x_{(0):(k)}, x_{(k+m):(n+1)}\} \tag{1}$$

in which the duplicate observation occurs only twice. We may then use the property (assuming $i < j$ without loss of generality)

$$\{P_{1:i-1}, \boldsymbol{P_i} + \boldsymbol{P_j}, P_{i+1:j-1}, P_{j+1:k}\} \sim \mathrm{Dir}[\alpha_{1:i-1}, \boldsymbol{\alpha_i} + \boldsymbol{\alpha_j}, \alpha_{i+1:j-1}, \alpha_{j+1:k}] \tag{2}$$

to merge overlapping point intervals in

$$\mathcal{F}_n^{\diamond} = \left[\sum_{i=1}^{n+1} W_i H_{x_{(i)}}, \sum_{i=1}^{n+1} W_i H_{x_{(i-1)}}\right] \tag{3}$$

with $\{W_{1:n+1}\} \sim \mathrm{Dir}\,[1, \ldots, 1]$. This results in

$$\mathcal{F}_n^{\diamond} = \left[\sum_{i=1}^{n-m+2} \tilde{W}_i H_{z_i}(x), \sum_{i=1}^{n-m+2} \tilde{W}_i H_{z_{i-1}}(x)\right] \tag{4}$$

with the adjusted weight vector

$$\{\tilde{W}_{1:n-m+2}\} \sim \mathrm{Dir}\,[1, \ldots, 1, m, 1, \ldots, 1]\,, \tag{5}$$

where a weight m has been assigned to the $k{+}1$ position. With this modification, only $(n - m + 2)$ random samples are required to sample a random realisation of $\mathcal{F}_n^{\diamond}$. When $n - m \ll n$ this results in a correspondingly large speedup.

S2 Further examples

This section presents the computation of c-credible intervals for two additional quantities of interest, both applied to data from the log-normal distribution with log-mean 0 and log-standard deviation 1.

We first consider the truncated mean, which discards one or both extremes of the inferred distribution. Figure S1 shows the probability box for a truncated mean $\mu_{99\%}$, discarding the largest 1% of sampled distribution for each resampling step. This can be considered the counterpart to the conditional value at risk ($\mathrm{CVaR}_{99\%}$), which reports the mean of the remaining tail contribution: $\mu = 0.99\mu_{99\%} + 0.01\mathrm{CVaR}_{99\%}$. Because the sample size $n = 15$ is insufficient to meaningfully compute a 99% quantile, we used 1000 data samples from the (0,1)-log-normal distribution instead. Note that the resulting upper and lower distributions are very similar, implying that the resulting uncertainty is predominantly probabilistic in nature. Estimates of the truncated mean are hardly affected by unknown features of the distribution.

Finally, we estimate the location for the 99% quantile $q_{99\%}$ (also known as the 0.99 value-at-risk), which is also the cutoff point for the truncated mean $\mu_{99\%}$. The resulting probability box and interval are shown in Fig. S2. Table S1 summarises all results, including those from Figure 3 in the main text.

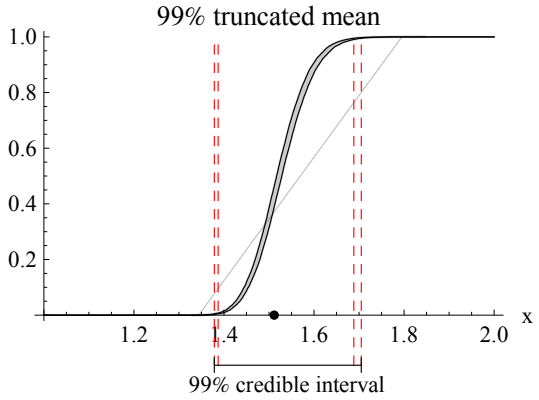

Figure S1: Probability box for the truncated mean $\mu_{99\%}$, considering only the 99% probability mass associated with the smallest values. Based on 1000 samples from a (0,1)-log-normal distribution and $n_{\text{resample}} = 10,000$. The black dot denotes the true value of the median of the originating distribution (≈ 1.51).

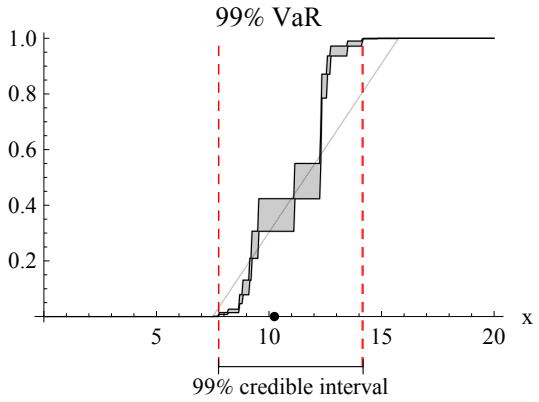

Figure S2: Probability box for the 99% quantile $q_{99\%}$. Based on 1000 samples from a (0,1)-log-normal distribution and $n_{\text{resample}} = 10,000$. The black dot denotes the true value of the median of the originating distribution (≈ 10.2).

Table S1: Interval estimates

N_{sample}	parameter	N_{resample}	credibility	result
15	median	1000	90 %	$[0.34, 3.60]$
15	mean	1000	90 %	$[1.21, \infty)$
1000	99% quantile	10,000	99 %	$[7.76, 14.2]$
1000	99% truncated mean	10,000	99 %	$[1.38, 1.70]$

S3 Connections to other methods

In the following we discuss various limiting cases of the robust posterior prior $\mathcal{F}_n^\diamond$ described in the paper, and we make connections to other inference methods described in the literature.

S3.1 Conditional distribution for x

It is illustrative to consider the restriction of the random imprecise distribution $\mathcal{F}_n^\diamond$ to a given position x, assuming $x \notin \{x_{1:n}\}$. Repeated application of (2) to (3) and $\{W_{1:n+1}\} \sim \mathrm{Dir}\,[1,\ldots,1]$ results in the random probability interval

$$\mathcal{F}_n^\diamond(x) = \left[W_x^-, 1 - W_x^+\right], \tag{6a}$$

with

$$\{W_x^-, W_x^1, W_x^+\} \sim \mathrm{Dir}[n^-(x), 1, n^+(x)]. \tag{6b}$$

Here $n^-(x)$ is the number of observations smaller than x and $n^+(x)$ is the number of observations larger than x.

For a Dirichlet distribution $\mathrm{Dir}[\alpha_{1:n}]$, the probability distribution of a compound event is the beta distribution $P(A) \sim \beta[\sum_{i\in A}\alpha_i, \sum_{j\notin A}\alpha_j]$. Therefore, the lower and upper bounds can be described as:

$$\underline{\mathcal{F}}_n(x) \sim \beta[n^-(x), n^+(x)+1], \tag{7}$$

$$\overline{\mathcal{F}}_n(x) \sim \beta[n^-(x)+1, n^+(x)]. \tag{8}$$

These random variables are inferred lower and upper bound distributions expressing our state of knowledge regarding the value $F_X^*(x) = P(X \leq x)$. The results are consistent with Bayesian estimators from Bernoulli process data, using the extreme priors $\lim_{\varepsilon\to 0}\beta[\varepsilon, 1-\varepsilon]$ ($F_0(x) \approx 0$) and $\lim_{\varepsilon\to 0}\beta[1-\varepsilon, \varepsilon]$ ($F_0(x) \approx 1$), respectively. In the language of Bayesian sensitivity analysis these priors may be considered bounds for the family of priors $\mathcal{P}_x = \{\beta[p, 1-p] : p \in [0,1]\}$.

S3.2 Purely probabilistic approach

We now consider the restriction of $\mathcal{F}_n^\diamond$ to a purely probabilistic description of uncertainty, starting from the definition

$$\mathcal{F}_n^\diamond = \mathcal{U}_{n+1} \circ F_n^\diamond, \qquad \text{with} \tag{9}$$

$$F_n^\diamond = \frac{1}{n+1}F_0^\diamond + \frac{n}{n+1}\hat{F}_n, \tag{10}$$

$$F_0^\diamond = [H_{x_R}, H_{x_L}]. \tag{11}$$

A possible uninformative choice is to replace the vacuous prior $F_0^\diamond$ by the average of its upper and lower distributions:

$$F_0^\diamond \to \frac{1}{2}H_{x_L} + \frac{1}{2}H_{x_R}, \tag{12}$$

which results in

$$\dot{\mathcal{F}}_n = \mathcal{U}_{n+1} \circ \left(\frac{\frac{1}{2} H_{x_L} + \frac{1}{2} H_{x_R} + \sum_{i=1}^{n} H_{x_i}}{n+1} \right). \tag{13}$$

The random variable $\dot{\mathcal{F}}_n(x)$ at $x \in I; x \neq x_i$ is

$$\dot{\mathcal{F}}_n(x) \sim \beta \left[n^-(x) + \frac{1}{2}, n^+(x) + \frac{1}{2} \right]. \tag{14}$$

As in the previous case, we can interpret this result as the Bayesian estimator for the probability $F_X^*(x) = P(X \leq x)$ given $n = n^+ + n^-$ independent measurements. This time, the prior is $\beta(\frac{1}{2}, \frac{1}{2})$, an uninformative prior which is both the Jeffreys prior and reference prior for this problem [1]. This correspondence retrospectively justifies the choice (13) as a minimally informative purely probabilistic contraction of $\mathcal{F}_n^\diamond$.

S3.3 Stochastic processes and NPI

Besides the intuitive interpretation as a distribution of distributions, the Dirichlet process may also be interpreted as a stochastic process in which each generated sample modifies the probability distribution for subsequent samples in a rich-get-richer scheme [2]. The observations in the resulting infinite sequence are dependent but exchangeable. The distribution and process viewpoints are linked by De Finetti's theorem, which states that distributions of the observations in (infinite) sequences generated in this way are distributed according to a measure representing a latent variable. In this case, that measure is the distribution-of-distributions representation of the Dirichlet process [3].

The robust posterior distribution $\mathcal{F}_n^\diamond$ has the property that, on average, it assigns a probability mass $1/(n+1)$ to the interval between measurements. Similarly, Hill [4] has proposed the 'A_n assumption' for a stepwise inference process, stating that "conditional upon the observations $X_1, \ldots, X_n$, the next observation X_{n+1} is equally likely to fall in any of the open intervals between successive order statistics of the given sample." The A_n assumption is then used to generate bounds on successive samples, each time adding the (possible) generated samples to the observations to be used in A_{n+1}.

In a 1993 paper [5], Hill has made use of the concept of adherent mass to assign the two halves of the probability associated with each interval to its lower and upper boundaries. In combination with limits at $-\infty$ and $+\infty$ it was shown that the resulting process is equivalent to the Dirichlet process defined in Eq. (13). The purely probabilistic projection $\dot{\mathcal{F}}_n$ of $\mathcal{F}_n^\diamond$ is therefore the De Finetti measure of an A_n process.

Hill's A_n assumption forms the basis of the Nonparametric Predictive Inference (NPI) method (see e.g. Coolen [6]). In this approach, successive predictive observations are generated using minimally informative interval probabilities (probability boxes) constrained by A_n. The first inference step of the NPI pro-

cess is given by the imprecise predictive distribution

$$F_{n+1}^{NPI} = \left[\sum_{i=1}^{n+1} \frac{H_{x_{(i)}}}{n+1}, \sum_{i=0}^{n} \frac{H_{x_{(i)}}}{n+1} \right], \tag{15}$$

where we make use of the extended set of order statistics $\{x_{(0):(n+1)}\}$. This imprecise distribution is equal to the imprecise input distribution $F_N^\diamond$ (10). This is consistent, because our best (non-random) estimate for the $(n+1)$-predictive distribution is the *expected* probability box $E[\mathcal{F}_n^\diamond] = F_n^\diamond$.

Furthermore, we can invoke Hill's adherent mass argument [5] to analyse the NPI process in the case of a monotonic population parameter q. In this case, we assign the adherent mass consistently to either the lower or upper bounds of each interval to estimate the lower and upper bounds of q, respectively. Analogous to Hill's result for $\dot{\mathcal{F}}_n$ we find

$$Q_{\text{NPI}}^{\min} = q \left[\text{CDP} \left[\frac{\sum_{i=1}^{n} H_{x_i} + H_{x_L}}{n+1} \right] \right], \tag{16}$$

$$Q_{\text{NPI}}^{\max} = q \left[\text{CDP} \left[\frac{\sum_{i=1}^{n} H_{x_i} + H_{x_R}}{n+1} \right] \right]. \tag{17}$$

This result is identical to the result $[Q^{\min}, Q^{\max}] = \left[q[\overline{\mathcal{F}}_n], q[\underline{\mathcal{F}}_n\right]$ from the main text. In other words, NPI *for infinite random sequences* produces bounds that are identical to those derived using the robust posterior distribution. Therefore we postulate that the robust posterior distribution may be considered the De Finetti measure of the NPI process.

Even though the results of the NPI process converge to those obtained using our method, they are conceptually quite different. NPI has a process-based definition that generates a series of predictive samples. The first sample is always drawn from $F_n^\diamond = E[\mathcal{F}_n^\diamond]$, and subsequent samples are generated from subsequent sample-dependent distributions. An infinite sequence of samples obtained in this way is equivalent to computing a *single* realisation of $\mathcal{F}_n^\diamond$. The direct formulation developed in this paper results in a significant computational advantage, especially for resampling approaches such as Bayesian Interval Sampling, which require many realisations of $\mathcal{F}_n^\diamond$.

S4 Posterior Dirichlet processes

In this section we demonstrate that $\mathcal{F}_n^{\diamond(s)} = \left[\underline{\mathcal{F}}_n^{(s)}, \overline{\mathcal{F}}_n^{(s)}\right]$ is a posterior distribution corresponding corresponding to the vacuous prior distribution $F_0^\diamond = [H_{x_R}, H_{x_L}]$ with concentration parameter s and observations $x_{1:n}$. This is done in two steps. First, a partial ordering is introduced on the space of cumulative Dirichlet processes (CDPs). Second, the degenerate CDPs are expressed as limits of non-degenerate CDPs. This enables us to provide a limiting definition of the posterior CDPs.

Definition. We define a partial ordering on CDPs $\mathcal{F}$ and $\mathcal{G}$ as follows:

$$\mathcal{F} \succeq \mathcal{G} : \mathcal{F}(x) \text{ equals or dominates } \mathcal{G}(x), \forall x \in \mathbb{R}, \tag{18}$$

where dominance is defined as first order stochastic dominance [7]. In other words, $\mathcal{F} \succeq \mathcal{G}$ if for all x each quantile of $\mathcal{F}(x)$ equals or exceeds the corresponding quantile of $\mathcal{G}(x)$.

Lemma S4.1. *Let $\mathcal{F}$ and $\mathcal{G}$ be CDPs with concentration parameter s:*

$$\mathcal{F}^{(s)} \sim \mathrm{CDP}[sF], \tag{19}$$

$$\mathcal{G}^{(s)} \sim \mathrm{CDP}[sG], \tag{20}$$

where F and G cumulative probability distributions. Then the following implication holds

$$(F(x) \geq G(x), \forall x \in \mathbb{R}) \Rightarrow \mathcal{F}^{(s)} \succeq \mathcal{G}^{(s)}. \tag{21}$$

Proof. It follows from the properties of the Dirichlet process that the random variable $\mathcal{F}(x)$ is distributed according to the beta distribution

$$\mathcal{F}^{(s)}(x) \sim \beta[sF(x), s(1 - F(x))]. \tag{22}$$

Therefore,

$$F(x) > G(x) \Rightarrow \mathcal{F}^{(s)}(x) \text{ dominates } \mathcal{G}^{(s)}(x). \tag{23}$$

This holds for all x, thus proving the statement (21). □

The implication of this lemma is that for a given value of the concentration parameter s the cumulative Dirichlet process preserves the partial ordering of the shape parameter (the 'input' distribution) in a probabilistic sense.

Definition. We define the family of ε-contaminated priors that are further parametrised by the concentration parameter s and a distribution function F on I.

$$\underline{\mathcal{F}}_0^{(s,\varepsilon)}[F] \sim \mathrm{CDP}[s\left((1-\varepsilon)H_{x_R} + \varepsilon F\right)] \tag{24a}$$

$$\overline{\mathcal{F}}_0^{(s,\varepsilon)}[F] \sim \mathrm{CDP}[s\left((1-\varepsilon)H_{x_L} + \varepsilon F\right)] \tag{24b}$$

When the implied support of F covers that of F_X^* (the target distribution), these Dirichlet processes have the desirable property that they almost surely have a nonzero probability mass associated with the local neighbourhood of the observed values $\{x_{1:n}\}$ and. Specifically, this is the case for the choices $F = \mathrm{uniform}(I)$ and $F = F_X^*$. The ε-contaminated priors also have the limits

$$\lim_{\varepsilon \downarrow 0} \underline{\mathcal{F}}_0^{(s,\varepsilon)}[F] = \underline{\mathcal{F}}_0^{(s)} \sim \mathrm{CDP}[sH_{x_R}] \tag{25a}$$

$$\lim_{\varepsilon \downarrow 0} \overline{\mathcal{F}}_0^{(s,\varepsilon)}[F] = \overline{\mathcal{F}}_0^{(s)} \sim \mathrm{CDP}[sH_{x_L}]. \tag{25b}$$

Property (21) proves that these limits exist for the Cumulative Dirichlet Process, by virtue of their existence in the regular distribution space. The p-box $[\underline{\mathcal{F}}_0^{(s)}, \overline{\mathcal{F}}_0^{(s)}]$ can therefore be considered the envelope of the set of proper priors on I with concentration parameter s.

For $\underline{\mathcal{F}}_0^{(s,\varepsilon)}[F]$ and $\overline{\mathcal{F}}_0^{(s,\varepsilon)}[F]$ with $\varepsilon > 0$ the Bayesian posterior distributions conditioned on the observations $\{x_{1:n}\}$ is given by

$$\underline{\mathcal{F}}_n^{(s,\varepsilon)}[F] \sim \text{CDP}\left[s\left((1-\varepsilon)H_{x_R} + \varepsilon F\right) + \sum_{i=1}^{N} H_{x_i}\right], \tag{26a}$$

$$\overline{\mathcal{F}}_n^{(s,\varepsilon)}[F] \sim \text{CDP}\left[s\left((1-\varepsilon)H_{x_L} + \varepsilon F\right) + \sum_{i=1}^{N} H_{x_i}\right]. \tag{26b}$$

Analogous to Eqs. (25) we can now define the posterior envelope distributions as

$$\underline{\mathcal{F}}_n^{(s)} = \lim_{\varepsilon \downarrow 0} \underline{\mathcal{F}}_n^{(s,\varepsilon)}[F] = \mathcal{U}_{n+s} \circ \left(\frac{s}{n+s} H_{x_R} + \frac{n}{n+s}\hat{F}_n\right) \tag{27a}$$

$$\overline{\mathcal{F}}_n^{(s)} = \lim_{\varepsilon \downarrow 0} \overline{\mathcal{F}}_n^{(s,\varepsilon)}[F] = \mathcal{U}_{n+s} \circ \left(\frac{s}{n+s} H_{x_L} + \frac{n}{n+s}\hat{F}_n\right). \tag{27b}$$

We conclude that $\mathcal{F}_n^{\diamond(s)} = \left[\underline{\mathcal{F}}_n^{(s)}, \overline{\mathcal{F}}_n^{(s)}\right]$ is indeed a tight envelope of the proper prior CDPs with concentration parameter s bounded by the vacuous prior $F_0^{\diamond} = [H_{x_R}, H_{x_L}]$.

References

[1] Bernardo JM. Reference Analysis. In: Handbook of Statistics, Volume 25: Bayesian Thinking Modeling and Computation. vol. 25 of Handbook of Statistics. Elsevier; 2005.

[2] Ferguson TS. A Bayesian Analysis of Some Nonparametric Problems. The Annals of Statistics. 1973;1(2):209–230.

[3] Lijoi A, Prünster I. Models beyond the Dirichlet process. In: Bayesian Nonparametrics. Cambridge; 2010. p. 80–135.

[4] Hill BM. Posterior Distribution of Percentiles: Bayes' Theorem for Sampling from a Population. Journal of the American Statistical Association. 1968;63(322):677–691.

[5] Hill BM. Parametric Models for An: Splitting Processes and Mixtures. Journal of the Royal Statistical Society Series B (Methodological). 1993;55(2):423–433.

[6] Coolen FPA. On Nonparametric Predictive Inference and Objective Bayesianism. Journal of Logic, Language and Information. 2006;15(1-2):21–47.

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