Prior elicitation and evaluation of imprecise judgements for Bayesian analysis of system reliability

Alireza Daneshkhah, Amin Hosseinian-Far, Tabassom Sedighi, Maryam Farsi

Abstract System reliability assessment is a critical task for design engineers. Identifying the least reliable components within a to-be system would immensely assist the engineers to improve designs. This represents a pertinent example of data informed decision making (DIDM). In this chapter, we have looked into the theoretical frameworks and the underlying structure of system reliability assessment using prior elicitation and analysis of imprecise judgements. We consider the issue of imprecision in the expert's probability assessments. We particularly examine how imprecise assessments would lead to uncertainty. It is crucial to investigate and assess this uncertainty. Such an assessment would lead to a more realistic representation of the expert's beliefs, and would avoid artificially-precise inferences. In other words, in many of the existing elicitation methods, it cannot be claimed that the resulting distribution perfectly describes the expert's beliefs. In this paper, we examine suitable ways of modelling the imprecision in the expert's probability assessments. We would also discuss the level of uncertainty that we might have about an expert's density function following an elicitation. Our method to elicit an expert's density function is nonparametric (using Gaussian Process emulators), as introduced by Oakley and O'Hagan [1]. We will modify this method by including the imprecision in any elicited probability judgement. It should be noticed that modelling imprecision does not have any impact on the expert's true density function, and it only affects the analyst's uncertainty about the unknown quantity of interest.

Alireza Daneshkhah
The Warwick Centre for Predictive Modelling, School of Engineering, The University of Warwick
Coventry, CV4 7AL, UK, e-mail: ardaneshkhah@gmail.com

Amin Hosseinian-Far
School of Computing, Creative Technologies & Engineering
Leeds Beckett University, Leeds, LS6 3QR, UK, e-mail: A.Hosseinian-Far@leedsbeckett.ac.uk

Tabasom Sedighi & Maryam Farsi
Cranfield School of Aerospace Transport & Manufacturing
Cranfield MK43 0AL, UK, e-mail: t.sedighi@cranfield.ac.uk; maryam.farsi@cranfield.ac.uk

We will compare our method with the method proposed in [2] using the 'roulette method'. We quantify the uncertainty of their density function, given the fact that the expert has only specified a limited number of probability judgements, and that these judgements are forced to be rounded. We will investigate the advantages of these methods against each other. Finally, we employ the proposed methods in this paper to examine the uncertainty about the prior density functions of the power law model's parameters elicited based on the imprecise judgements and how this uncertainty might affect our final inference.

1 Introduction

Assume the elicitation of a single expert's belief about some unknown continuous quantity denoted by θ. Identifying the underlying density function for θ based on the expert's beliefs, is the objective of the elicitation process. We denote this density function by $f(\theta)$. This function can be a member of a convenient parametric family. O'Hagan et al. [3] present a comprehensive review of the methods to elicit the density function chosen from a wide range of parametric families. Oakley and O'Hagan [1] criticised this way of eliciting expert's density function, and reported the following deficiencies as the main drawbacks of the so-called parametric elicitation methods: 1) The expert's beliefs are forced to fit the parametric family; 2) other possible distributions that might have fitted the elicited statements equally well have not been taken into account.

In order to tackle these drawbacks, they propose a nonparametric elicitation approach. In this approach, two parties are involved: an expert (female) and the analyst (male). The analyst receives the expert's elicited statements and will then make inference about the expert's density function, $f(\theta)$ based on her statements. In this approach, the analyst's fitted estimate of the expert's density function $f(\theta)$ is nonparametric, thereby avoiding forcing the expert's density into a parametric family.

This nonparametric elicitation approach can be seen as an exercise in Bayesian inference where the unknown quantity is $f(\theta)$. Moreover, the analyst's wishes are to formulate his prior beliefs about this unknown quantity. He then updates this prior in light of the data received from the expert in the form of probabilities assessments (e.g., quantiles, quartiles, mode, or mean) in order to obtain his posterior distribution for $f(\theta)$. The analyst's posterior mean serves as the 'best estimate' for the expert's density function $f(\theta)$, while the variance around this estimate describes the uncertainty in the elicitation process.

One crucial question that might arise in this non-parametric elicitation method is that, can the expert specify the probabilities assessments with absolute precision? This is evident that the facilitator cannot elicit her probabilities with absolute precision (or she cannot present her beliefs with the absolute precision). We are then interested to investigate what implications this might have on the uncertainty of the elicited prior distribution. There are several factors ([3]) that caused this issue. One of the most important challenges is that it is difficult for any expert to give precise numerical values for their probabilities; nevertheless, this is a requirement in most elicitation schemes. For instance, in many elicitation schemes, it is common to ask for percentiles or quantiles rather than probabilities. However, if the expert is unable to make precise probability assessments, then she will not be able to make precise percentile (or quantile) assessments either (see Section 3 and [3] for the differences between the 'precise' and 'imprecise' probabilities).

An obvious and natural source of imprecision comes from rounding done by the expert during probability assessments. We outline such imprecisions by the following explanation. Assuming expert's degree of belief probability $\theta < 0.1$, she may only consider probabilities rounded to the nearest, for instance, 0.05 (unless she wishes to state a probability close to 0 or 1). There may be an element of 'vagueness' in her imprecisely stated probabilities. She may believe that the event $\{\theta < 0.1\}$ is 'quite unlikely, 'yet she may find it difficult to attach a single numerical value to this belief. For example, she may find the question, "Is your probability of $\{\theta < 0.1\}$ less than 0.45?" straightforward to answer, but have difficulty in answering the question, "Is your probability of $\{\theta < 0.1\}$ closer to 0.1 or 0.2?", as she may have difficulty in deciding whether a probability of 0.1 describes her belief better than a probability of 0.2. The actual probability she reports may be chosen somewhat arbitrarily from the lower end of the probability scale, and all the subsequent probability judgements regarding θ will be dependent on this choice.

In order to investigate the impact of the imprecise probabilities stated by the expert on the elicited prior distribution and uncertainty around it (due to the imprecision), the elicitation methods must be modified in a way to capture this imprecision.

The purpose of this paper is to modify a *nonparametric* elicitation approach that is originally proposed by Oakley and O'Hagan [1] when the stated probabilities by the experts are imprecise. In Section 2, we briefly introduce this elicitation method and its properties and flexibility in practice. Section 3 is devoted to the short literature review of some main papers regarding the modelling of the imprecise probabilities. We also introduce a simple way for modelling the imprecision which is more straightforward to combine with the nonparametric elicitation method with a view to report the expert's density function and uncertainty around it. Section 4 is dedicated to the modification of Oakley and O'Hagan's elicitation method where the stated probabilities are not precise. Instead of eliciting percentiles or quantiles, we prefer to use a simpler type of fixed interval elicitation method known as *trial roulette* which is originally introduced in [4]. This method is introduced in Section 5. Finally, in Section 6, we employ the proposed methods in this paper to examine the uncertainty about the prior density functions of the power law model's parameters elicited based on the imprecise judgements and how this uncertainty might affect our final inference.

2 Nonparametric elicitation

Let's consider eliciting an expert's beliefs about some unknown continuous variable θ. The elicitation is the process of translating someone's beliefs about some uncertain quantities into a probability distribution. The expert's judgements about the uncertain quantities are traditionally fitted to some member of a convenient parametric family. There are few deficiencies in this approach that can be tackled by the nonparametric approach proposed in [1] which will be briefly introduced in this section. In this approach, two parties are involved: expert and facilitator. We suppose that the elicitation is conducted by a (male) facilitator, who interviews the (female) expert and identifies a density function f that represents her beliefs about θ. He will help her as much as possible, for example by providing suitable training and discussing the various biases that can influence probability judgements.

The expert is usually only able to provide certain summaries of her distribution such as the mean or various percentiles. O'Hagan et al. [3] show that these information are not enough to determine $f(\theta)$, uniquely. In the traditional methods, a density function that fits those summaries, as closely as possible, is chosen from some convenient parametric family of distributions. Oakley and O'Hagan [1] reported two deficiencies regarding this approach: first, it forces the expert's beliefs to fit the parametric family; and second, it fails to acknowledge the fact that many other densities might have well fitted the same summaries equally. They then presented an approach that addresses both deficiencies together in a single framework.

This nonparametric approach allows expert's distribution to take any continuous form. The issue of fitting a density function to the summaries stated by the expert is considered as an exercise in Bayesian inference, where an unknown quantity of interest is the expert density function, denoted by $f(\theta)$. The facilitator formulates his prior beliefs about this quantity. He then updates his beliefs about the expert's true density function in light of the data that are obtained from the expert in the form of (probability) judgements. This is to obtain his posterior distribution for $f(\theta)$. The facilitator's posterior mean can then be offered as a 'best estimate' for $f(\theta)$, while his posterior distribution quantifies the remaining uncertainty around this estimate.

A very useful and flexible prior model for an unknown function (e.g., expert density function) is the Gaussian process. Thus, it is reasonable that the facilitator represents prior uncertainty about f using a Gaussian process: for any set $\{\theta_1,\dots,\theta_n\}$ of values of θ, his prior distribution for $\{f(\theta_1),\dots,f(\theta_n)\}$ is multivariate normal. The reason that the joint distribution of $f(.)$ at some finite points is multivariate normal comes from the formal definition of a Gaussian process over $f(\theta)$, and further details and examples can be seen in [3] and references therein. As $f(\theta)$ is a density function, two constraints are applied to the facilitator's prior: $\int_{-\infty}^{\infty} f(\theta)d\theta = 1$ and $f(\theta) \geq 0$ for all θ. The first constraint is applied as part of the data from the expert, and second constraint is applied (approximately) using simulation, which we discuss in later sections.

The facilitator specifies his mean and variance-covariance functions for f hierarchically, in terms of a vector α of hyperparameters. His prior expectation of $f(\theta)$ is some member $g(\theta \,|\, u)$ of a suitable parametric family with parameters u, contained within α, so

$$E(f(\theta) \mid \alpha) = g(\theta \mid u), \tag{1}$$

where $g(. \mid u)$ denotes the underlying density, in particular case, where $u = (m, v)$, is distributed as a normal density with mean m and variance v (t distribution can be considered as an alternative as suggested in [1, 6], but we do not consider this extension here), and

$$Cov\{f(\theta), f(\phi) \mid \alpha\} = \sigma^2 g(\theta \mid u) g(\phi \mid \phi) c(\theta, \phi \mid u, b), \tag{2}$$

where $c(.,. \mid u, b)$ is a correlation function that takes value 1 at $\theta = \phi$ and is a decreasing function of $|\theta - \phi|$, σ^2 specifies how close the true density function will be to its estimation, and so governs how well it approximates to the parametric family, and $\alpha = (m, v, b, \sigma^2)$. The following correlation function is used:

$$c(\theta, \phi \mid u, b) = \exp\{-\frac{1}{2vb}(\theta - \phi)^2\}, \tag{3}$$

where b is the smoothness parameter. The correlation function above makes $f(.)$ infinitely differentiable with probability 1 (see [6] for further details and exploitation of this property).

The prior distribution of $\alpha = (m, v, \sigma^2, b^*)$ have the following form

$$p(m, v, \sigma^2, b^*) \propto p(m, v) p(\sigma^2) p(b^*). \tag{4}$$

Oakley and O'Hagan [1] used an improper prior distribution for the Gaussian process variance σ^2. However, this prior makes the computation more feasible, but the analyst's posterior distribution of the expert density might be sensitive to the choice of prior of σ^2 (and b^* as well). We use the following proper prior distributions for σ^2 and b^*, respectively.

$$p(\sigma^2) = IG(a, d), \quad logb^* \sim N(0, 1) \quad \text{or} \quad N(0, 4), \tag{5}$$

where IG stands for Inverse Gamma.

We will also investigate sensitivity of the expert density with respect to changes of the prior distributions for σ^2 and b^*. For instance, other reasonable distributions for b^*, used in [1], are $N(0, 1)$ or $N(0, 4)$. The result of this study will be presented in the next section.

To update the facilitator's prior beliefs about the expert's density, we need to elicit some summaries as the data from the expert. O'Hagan et al. [3] recommended that the expert should be asked about the probabilities, percentiles or quntiles. However, the nonparametric approach is being used in this paper can be applied to any of probabilities, percentiles or quntiles, we generate the data by an elicitation scheme called 'trial roulette' that will be introduced in the next section. This method is quite simple to implement by the expert and originally suggested by Gore [4] in a medical case study. Suppose that the vector of data, D, is

$$D^T = \left(\int_{a_0}^{a_1} f(\theta)d\theta, \ldots, \int_{a_{n-1}}^{a_n} f(\theta)d\theta \right), \tag{6}$$

where the end points $a_0, \ldots, a_n$ denote the possible value of θ, and $a_i > a_{i-1}$, $i = 1, \ldots, n$. It should be mentioned that D will also include this information that $\int_{-\infty}^{\infty} f(\theta)d\theta = 1$.

Because $f(\theta)$ has a normal distribution for all values of θ, then any data point also has a normal distribution and the joint distribution of D and any finite set of points on the function $f(\theta)$ is multivariate normal. The mean and variance of D are given by

$$H^T = E(D \mid (m,v))^T = \left(\int_{a_0}^{a_1} g(\theta \mid (m,v))d\theta, \ldots, \int_{a_{n-1}}^{a_n} g(\theta \mid (m,v))d\theta \right),$$

and

$$Cov(P_{A_i}, P_{A_j} \mid \alpha) = \sigma^2 \int_{A_j} \int_{A_i} g(\theta \mid (m,v)) g(\phi \mid (m,v)) c(\theta, \phi \mid b^*, v) d\theta d\phi = \sigma^2 A, \tag{7}$$

where $A_i = [a_{i-1}, a_i]$ and $A_j = [a_{j-1}, a_i]$ for all $i, j = 1, \ldots, n$. The details of the computation of H and A can be found in [5].

It then follows immediately from properties of the multivariate normal distribution that the updated facilitator's beliefs about $f(\theta)$ given D and α also has a normal distribution with

$$E(f(\theta) \mid D, \alpha) = g(\theta \mid m, v) + t(\theta \mid \alpha)^T A^{-1}(D - H), \tag{8}$$

and

$$Cov(f(\theta), f(\phi) \mid D, \alpha) = \sigma^2 \{ g(\theta \mid m,v) g(\phi \mid m,v) c(\theta, \phi \mid b^*, v) - t(\theta \mid \alpha)^T A^{-1} t(\theta \mid \alpha) \}, \tag{9}$$

where $t(\theta \mid \alpha)$ is given by

$$t(\theta \mid \alpha)^T = (Cov(f(\theta), P_{A_1} \mid \alpha), \ldots, Cov(f(\theta), P_{A_n} \mid \alpha)),$$

$A_i = [a_{i-1}, a_i]$, and $\alpha = (m, v, b^*, \sigma^2)$.

In fact, conditional on α and data, the analyst's posterior distribution of $f(\theta)$ is again a Gaussian process with equations (8) and (9) giving its mean and covariance structure, respectively.

The posterior distribution of α is given by

$$p(m, v, \sigma^2, b^* \mid D) \propto v^{-1} \sigma^{-n} |A|^{-\frac{1}{2}} \exp\{ -\frac{1}{2\sigma^2} (D - H)^T A^{-1} (D - H) \} p(b^*) p(\sigma^2) \tag{10}$$

where n denotes the number of elements in D.

The conditioning on σ^2 can be removed analytically to obtain

$$p(m, v, b^* \mid D) \propto v^{-1} |A|^{-\frac{1}{2}} (\hat{\sigma}^2)^{-\frac{n}{2}} p(b^*) \tag{11}$$

where

$$\hat{\sigma}^2 = \frac{(D-H)^T A^{-1}(D-H)}{n-2}.$$

The conditioning on the rest of the hyperparameters cannot be removed analytically. We then use MCMC to obtain a sample of values of $\{m, v, b^*\}$ from their joint posterior distribution. Given a set of values for the hyperparameters, a density function is sampled at a finite points of θ, $\{\theta_1, \ldots, \theta_k\}$, from the Gaussian process model. Repeating this many times, a sample of functions from $f(.) \mid D$ is obtained and the negative-valued functions are then removed. The remaining function are used to report estimates and pointwise credible bounds for the expert's density, $f(\theta)$.

3 Imprecise probabilities: modelling and modification issues

In this section, we briefly review the statistical literature of modelling of the imprecise subjective probabilities. We then present our approach to model the imprecision which exists in the expert's elicited probabilities. Finally, we modify the nonparametric elicitation approach presented in the previous section.

Imprecise probability models are needed in many applications of probabilistic and statistical reasoning. They have been used in the numerous scenarios such as:

- when there is little information for evaluating a probability (see Walley, 1991 and 1996);
- in robust Bayesian inference, to model uncertainty about a prior distribution (see Berger, 1984; Berger and Berliner, 1986; DeRobertis and Hartigan, 1981; Pericchi and Walley, 1991);
- in frequentist studies of robustness, to allow imprecision in a statistical sampling model, e.g., data from a normal distribution may be contaminated by a few outliers or errors that come from a completely unknown distribution (see Huber, 1981);
- to account for the ways in which people make decisions when they have indeterminate or ambiguous information (see Einhorn and Hogarth, 1985; Smithson, 1989).

Imprecise probability is used as a generic term to cover all mathematical models which measure chance or uncertainty without sharp numerical probabilities. It includes both qualitative and imprecise or nonadditive quantitative models. Most probability judgements in everyday life are qualitative and involve terms such as "probably" and "more likely than", rather than numerical values. There is an extensive amount of literature on this types of imprecise probability model.

Our main attempt here is to represent the imprecise probability stated by an expert as precise as possible in numbers. The most relevant work has been reported by Walley (1991) who considered bounding a probability P with upper and lower probabilities, $\overline{P}$ and $\underline{P}$ respectively. Unfortunately, this still leaves the issue of how to specify $\overline{P}$ and $\underline{P}$ with absolute precision unresolved. Additionally, the expert may also feel that values in the center of the interval $[\underline{P}, \overline{P}]$ represent their uncertainty more appropriately than values towards the ends of the interval.

We express the stated probability $P^*(\theta \in A)$ as the expert's true probability plus an additive error which represents the imprecision in the stated probability as follows,

$$P^*(\theta \in A) = P(\theta \in A) + \varepsilon, \tag{12}$$

where $\varepsilon \sim N(0, \tau^2)$ for some appropriate values of τ^2.

It should be noticed that normality itself can be considered as a strong assumption, but we now no longer have absolute lower and upper limits for the true probability $P(\theta \in A)$. This is also more plausible for the facilitator to give greater probability to value $P(\theta \in A)$ closer to $P^*(\theta \in A)$. The variance parameter τ^2 now describes the imprecision in the probability assessment, and may vary for different probability judgements. These values could be appropriately chosen in consultation with the expert.

Now, suppose that the expert provides k probability statements, denoted by $p_1^*, \ldots, p_k^*$ that are imprecise and modelled as follows

$$p_i^* = p_i + \varepsilon_i, \;\; i = 1, \ldots, k, \tag{13}$$

where p_i's are the expert's true probabilities that cannot be precisely stated, and ε_i is considered as a noise which has a distribution centered around zero and variance that describes the imprecision in the probability assessment.

The noise components in (13) can be followed of either normal distribution or uniform distribution. Oakley and O'Hagan (2007) have briefly discussed the effect of including this noise with a specific variance for each noise on the elicited expert's density function.

In this section, the focus is to study the impact of a uniform noise on the expert's density function. The uniform noise is considered as $\varepsilon \sim U(-a, a)$, where the length of the interval $2a$, denotes the maximum possible error that would exist in the expert's stated probabilities.

We consider a type of fixed interval method known as trial roulette elicitation method originally proposed by Gore (1987). In this method, the expert is given n gaming chips and asked to distribute them amongst k bins. The proportion of chips allocated to a particular bin is representing her probability of θ lying in that bin, though clearly this probability is then subject to rounding error. This approach can be attractive to some experts, as they do not need to determine the probability statements directly, and the distributions of chips distributed between bins gives a histogram of their beliefs.

One element of this method is the lack of precision in the expert's stated probabilities. Suppose the expert distributes n given chips into k bins, then her stated probabilities must be multipliers of $1/n$. If the expert selects smaller n, then she is only able to make coarse probability assessments, so the precision will be lowered. Another issue to be considered is to locate the bins (particularly close to the endpoints). The choice of scale can have a substantial effect on an expert's judgements (Garthwaite, 2007) which is not further studied in this chapter.

We now modify the non-parametric elicitation method described above, by eliciting the imprecise probabilities using the trial roulette scheme. The k bins considered in this method are represented by the following intervals:
$[a_0,a_1),[a_1,a_2),\ldots,[a_{k-1},a_k)$ (with the possibility of $a_0=-\infty$ and $a_k=\infty$). The expert allocates n_i chips to the ith bin, $[a_{i-1},a_i)$, with $\sum_{i=1}^{k} n_i = n$. We define $p_i^* = n_i/n$ to be the expert's stated probability of $\theta \in [a_{i-1},a_i)$, with $p^* = (p_1^*,\ldots,p_k^*)$. These stated probabilities are usually subject to rounding errors which can be linked to the true probabilities through (13), where the noise components can be represented as $\varepsilon_i \sim U(-a,a)$, such that $\sum_{i=1}^{k} \varepsilon_i = 0$.

The facilitator's data consists of p^*, together with this knowledge that $\sum_{i=1}^{k} \varepsilon_i = 0$, and $f(\theta) \geq 0 \ \ \forall \theta$. The latter constraint implies $p_i \geq 0 \ \forall i$. We are now required to compute the facilitator's posterior distribution denoted by considering these conditions:
$(\sum_{i=1}^{k} \varepsilon_i = 0, f \geq 0)$.

If the expert was to provide her true probabilities p rather than her stated probabilities p^*, it would be straightforward to derive the distribution of $f|p,\alpha = (m,v,b^*),\sigma^2$. This was discussed and presented in Section 2. However, it is more challenging to derive the posterior distribution of $f(\theta)|p^*,\alpha=(m,v,b^*,\sigma^2),\sum_{i=1}^{k}\varepsilon_i = 0)$, as H and A cannot be computed as presented above. In this scenario, we consider $\varepsilon = (\varepsilon_1,\ldots,\varepsilon_k)$ as extra hyperparameters. Therefore, the following prior distribution for the new set of hyperparameters (α,ε) would be considered as follows:

$$p(\alpha,\varepsilon \mid \Sigma_{i=1}^{k}\varepsilon_i = 0) \propto v^{-1} p(\sigma^2) p(b^*) p(\varepsilon \mid \Sigma_{i=1}^{k}\varepsilon_i = 0), \tag{14}$$

where $p(\sigma^2)$, $p(b^*)$ are the same priors presented in (5).

The main issue is to generate k random variables from $U(-a,a)$ such that $\Sigma_{i=1}^{k}\varepsilon_i = 0$. We propose the following procedure to generate the noise components:

- Let $\xi = 1 + \frac{k-2}{k}$;
- Draw a random sample of size k from $y_i \sim U(0,1)$ and define $x_i = 2ay_i - a$. These randomly variables follow $U(-a,a)$;
- Let $\varepsilon_i = \frac{x_i - \bar{x}}{\xi}$, where $\bar{x} = \frac{1}{k}\Sigma_{i=1}^{k} x_i$.

The posterior distribution of (α,ε), after integrating out σ^2, is then given by

$$p(m,v,b^*,\varepsilon \mid D) \propto v^{-1} |A|^{-\frac{1}{2}} (\hat{\sigma}^2)^{-\frac{n+d}{2}} p(b^*) p(\varepsilon) \tag{15}$$

where $\hat{\sigma}^2 = \frac{1}{n+d-2}[a + (D+\varepsilon-H)^T A^{-1} (D+\varepsilon-H)]$.

To derive the expert's density function, we are similarly required to use MCMC to obtain samples of the hyperparameters, $\alpha = (m,v,b^*)$ from their joint posterior distribution. Weak priors are assumed for m and v as in (4). The following proposal distributions are chosen for the Metropolis-Hasting sampler:

$$m_t \mid m_{t-1} \ \sim \ N(m_{t-1}, 0.01)$$

$$\log(v_t) \mid m_{t-1}, m_t, v_{t-1} \ \sim \ N(\log(v_{t-1}), 0.1 \times (1 + 5 \times |m_t - m_{t-1}|))$$

$$\log(b_t^*) \mid b_{t-1}^* \sim N(\log(b_{t-1}^*), 0.01).$$

The chain is run for 20,000 iterations and the first 10,000 runs are discarded to allow for the burn-in period. A random density function is generated for each of the last 10,000 runs. Given a set of values for the hyperparameters, a density function is sampled at k points of θ, $\{\theta_1, \ldots, \theta_k\}$, from the Gaussian process model. Repeating this will give us a sample of functions from the posterior $p\{f(.) \mid D\}$, and estimated and pointwise credible bounds for $f(\theta)$ can then be reported.

4 A Gaussian process model for imprecise probabilities

In the trial roulette elicitation scheme described above, we only addressed the possible imprecision due to rounding errors. The more fundamental problem is that the expert does normally have the challenge of representing a feeling of uncertainty with a numerical value. The expert usually first states a small number of probability assessments about θ. She may then wish to draw her cumulative distribution or density function between these points. However, this would completely specify her distribution for θ over the chosen range, but it does not seem reasonable that the facilitator should know her distribution in this range with no uncertainty. In particular, if the expert's original probabilities were imprecisely given, she might then believe that other probabilities would represent her beliefs equally well, which would have led to different curves being drawn.

This motivates the development of a model for imprecision in which the facilitator will still have some uncertainty about the expert's true density function regardless of how many probabilities the expert provides. Therefore, there is a need to consider correlation in the expert's imprecision. In the case that the stated probability departs from her true probability by some amount at a particular value of θ, θ_0 say, a similar deviation at values of θ close to θ_0 would be then expected. In this section, we present an alternative way to model the imprecise probabilities, in which the error terms have a different distribution and correlation structure.

Suppose that $f^*(\theta)$ illustrates the expert's reported imprecise density function. The relationship between the $f^*(\theta)$ and her true distribution $f(\theta)$ can be given by

$$f^*(\theta) = f(\theta) + q(\theta), \tag{16}$$

where $q(\theta)$ is an *error* indicating the imprecision in the expert's probabilities; $q(\theta)$ is a distortion of the expert's true, precise probabilities that result in the imprecise reported probabilities.

Suppose that, the stated probabilities reported by the expert are presented as follows:

$$\mathscr{D} = (P_{B_1}^* = \int_{B_1} f^*(\theta)d\theta, \ldots, P_{B_n}^* = \int_{B_n} f^*(\theta)d\theta, \int_{\Theta} q(\theta)d\theta = 0), \tag{17}$$

where $P^*_{B_i}$ is the expert's stated probability for the i^{th} interval, denoted by B_i, and $\Theta = \bigcup_{i=1}^{n} B_i$.

We believe both the expert's true distribution and reported distribution to be smooth functions, and we also specify a Gaussian process distribution of $q(\theta)$ that ensures that $q(\theta)$ will also be smooth. The variance of $q(\theta)$ is constructed in such a way that it will decrease when $f(\theta)$ approaches 0 or 1, so that the absolute magnitude of any imprecision error will decrease appropriately. Additionally, we condition on this assumption that the expert's imprecise probabilities still sum to 1. Therefore, the error term can be modelled by a Gaussian process with zero mean and the following covariance matrix,

$$Cov(q(\theta), q(\phi)) = \sigma_2^2 g(\theta) g(\phi) c_2(\theta, \phi).$$

As shown above, the facilitator's prior distribution of $f(\theta)$ can be also expressed by a Gaussian process:

$$f(\theta) \sim GP(g(\theta), \sigma_1^2 g(\theta) g(\phi) c_1(\theta, \phi)),$$

where $g(\theta) = N(m, v)$.

It can be shown that $f^*(\theta) \mid \mathscr{D}, m, v, b_1^*, b_2^*, \sigma_1^2, \sigma_2^2$ has a normal distribution with

$$E(f^*(\theta) \mid \alpha, \mathscr{D}) = g(\theta) + t(\theta)^T A^{-1}\{D - H\}$$

$$Cov\{f^*(\theta), f^*(\phi) \mid \alpha, \mathscr{D}\} =$$

$$\sigma_1^2\{g(\theta)g(\phi)c(\theta,\phi) - t(\theta)^T A_1^{-1} t(\phi)\} + \sigma_2^2\{g(\theta)g(\phi)c(\theta,\phi) - t(\theta)^T A_2^{-1} t(\phi)\}$$

where, $A = \sigma^1 A_1 + \sigma_2 A_2$, $b_i^* = \frac{b_i}{v}$, and details of calculating H , A_1, A_2 and t are available in [5].

The posterior distribution of $\alpha = (m, v, b_1^*, b_2^*, \sigma_1^2, \sigma_2^2)$ is easily found from the multivariate normal likelihood for $\mathscr{D}$ given α:

$$p(\alpha \mid \mathscr{D}) \propto v^{-1} |A|^{-\frac{1}{2}} \times \exp\{-\frac{1}{2}(D-H)^T A^{-1}(D-H)\} p(\sigma_1^2) p(\sigma_2^2) p(b_1^*) p(b_2^*), \tag{18}$$

where $p(\sigma_i^2)$ and $p(b_i^*)$ are the same priors presented in (5).

5 Applications in Reliability Analysis

In this section, we are going to investigate how allowing imprecision in the expert's probability assessments could influence the expert's density function regardless of its skewness to the right or left through some applications in reliability analysis. We focus on two main prior distributions that are widely used in Bayesian analysis of failure data: Beta distribution and Gamma distribution. The former one has been commonly used as a prior distribution to study success/failure data, while the latter one is the most popular conjugate prior for failure count data and failure time data.

5.1 Example: Beta prior distribution

In order to conduct a reliability assessment for a single item or for a system, analysts might require to obtain success/failure data. Such data is comprised of details surrounding the component's or system's success or failure while performing with a view to complete its intended function. For instance, testers may try an emergency diesel generator to see if it will start on demand, and record whether this generator starts or not. The record of whether the launched missile system completes its mission successfully or not is another example of such a dataset. This data can be modelled using binomial distribution for a fixed number of tests. For this model, the unknown parameter is the success probability, denoted by θ, that must be estimated based on the test data and experts' knowledge.

The conjugate prior distribution for success/failure data is the beta distribution, denoted by $\mathscr{B}(\alpha,\beta)$ and is defined as follows:

$$f(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)+\Gamma(\beta)}\theta^{\alpha-1}(1-\beta)^{\beta-1}, \quad \theta\in[0,1], \;\; \alpha>0, \;\; \beta>0,$$

where α is usually interpreted as the prior number of successful component tests and β as the prior number of failed component tests; and $\Gamma(.)$ stands for the Gamma function.

In this example, we focus on eliciting prior distribution for θ using the non-parametric prior elicitation described above and modelling the imprecision when the expert is asked to present her probability statements with trial roulette elicitation scheme. Without loss of generality, we suppose that the expert's true distribution is $\mathscr{B}(4,2)$.

The trail roulette scheme is chosen to elicit the expert's probabilities. The method is explained to the expert and she is then asked to choose the number of bins and chips to express her belief. Assuming that the expert has allocated her chips precisely, as shown in Table 1, the facilitator has two sources of uncertainty about the expert's distribution. Firstly, the expert has only provided four probabilities, corresponding to the four bins. This is illustrated in Figures 1 and 2 which show the facilitator's pointwise 95% intervals for the expert's density and distribution functions (based on a sample of 4000 functions), respectively. This issue is also investigated in Oakley et al. (2010) in details. The second source of uncertainty, which is of main interest in this paper, is due to the rounding in the expert's judgements. The effect of accounting for this as discussed above is also shown in Figures 1 and 2, where the dashed lines show the facilitator's pointwise 95% intervals for the expert's density/distribution function, the solid lines show the expert true density/distribution function, and dot-dashed $(.-.-)$ line illustrate the expert's

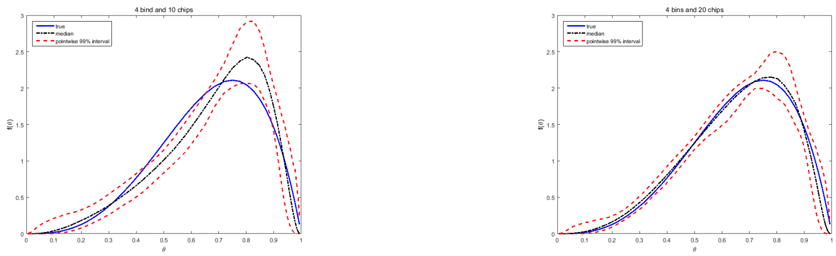

Fig. 1 The mean (dashed-point line) and pointwise 95% intervals (dashed lines) for the expert's density, and the true density (solid line) with uniform noise when the imprecise probabilities are included and rounded to nearest 5% (10 chips) and 2.5% (20 chips).

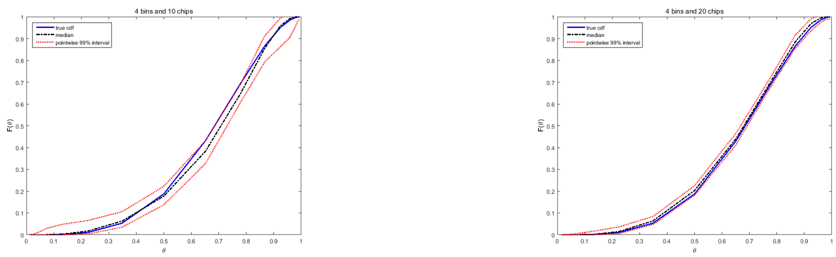

Fig. 2 The mean (dashed-point line) and pointwise 95% intervals (dashed lines) for the expert's CDF, and the true CDF (solid line) with uniform noise when the imprecise probabilities are included and rounded to nearest 5% (10 chips) and 2.5% (20 chips)

desnity/distribution function estimation based on the stated probabilities. Accounting for this imprecision has resulted in a wider range of distributions that might be consistent with the expert's beliefs.

num. of chips	$P(\theta \leq 0.4)$	$P(0.4 \leq \theta \leq 0.6)$	$P(0.6 \leq \theta \leq 0.8)$	$P(\theta \geq 0.8)$
10	0.10	0.20	0.40	0.30
20	0.10	0.25	0.40	0.25

Table 1 Stated expert's probabilities given 20 chips and 4 bins. True probabilities are generated from a $\mathscr{B}(4,2)$ distribution.

The facilitator's uncertainty about the expert's PDF (probability density function) (or CDF (cumulative distribution function) when she uses 4 bins and 20 chips (i.e., the imprecisely stated probabilities are rounded to nearest 2.5%) to express her beliefs is smaller than the uncertainty of the corresponding elicited PDF (CDF), calculated based on 10 chips (where the probabilities stated by the expert are rounded to nearest 5%). Therefore, a combination of a small number of bins and large number of chips can result in fairly little uncertainty about the expert's density. This is to be expected: the Gaussian process model for f implies that the CDF is also a smooth function, so observing a small number of points on the CDF should reduce uncertainty substantially. In addition, it is trivial to conclude that including the noise in the data would effect the expert's density (see Oakley and O'Hagan (2007) for another example and relevant details).

In other words, the approach proposed in this paper to modelling the imprecision would help the facilitator to measure the level of his uncertainty about the expert's density by decreasing/ increasing the number of chips and/or bins. The elicitation task might become easier for the expert by choosing the suitable number of chips and bins. In the following table, we have shown that if 20 chips (instead of 10 chips) are used by the expert, then the uncertainty of the expert's density will be decreased by more than 50%. This rate will be more than 300%, if we use 40 chips against 10 chips. The corresponding results are given in Table 2. In this table, n stands for the number of chips used by the expert, and E_m and V_m are the mean and variance of the expert density function obtained by fitting a Gaussian process to her stated beliefs using the trial roulette scheme, respectively. In addition, E_L and E_U are the endpoints of 95% bound around the posterior mean of the fitted Gaussian process to the experts probabilities, and V_L and V_U are the endpoints of 95% bounds around the corresponding posterior variance.

n	E_L	E_m	E_U	V_L	V_m	V_U	$E_U - E_L$	$V_U - V_L$
10	0.6938	0.7142	0.7335	0.0245	0.0310	0.0407	0.0397	0.0153
20	06836	0.6951	0.7062	0.0258	0.0302	0.0354	0.0226	0.0096
40	0.7047	0.7111	0.7173	0.0246	0.0267	0.0292	0.0126	0.0046

Table 2 The posterior mean, variance and their corresponding 95% pointwise bounds of the expert's density function using the trial roulette elicitation scheme with different number of chips.

5.2 Example: Gamma prior distribution

A commonly used prior distribution for the mean number of failures per unit time, and for the scale and shape parameters of the fitted distributions (e.g., exponential, Gamma, Weibull) to model failure time data (the most commonly used data to assess component reliability) is the gamma distribution, which is denoted by $\mathscr{G}(\alpha,\beta)$ and presented as:

$$f(\theta \mid \alpha,\beta) = \frac{1}{\beta^{\alpha}\Gamma\alpha}\theta^{\alpha-1}e^{-\frac{\theta}{\beta}}, \quad \theta > 0, \ \alpha > 0, \ \beta > 0.$$

Similar to the beta distribution example, we are interested in eliciting prior distribution for θ using the above elicitation method and modelling the imprecision, when the expert is asked to present her probability statements with trial roulette elicitation scheme. We assume the expert's true prior is $\mathscr{G}(5,1)$. The expert allocates 10 and

20 into 4 bins as shown in Table 3. The effect of rounding the expert's probability assessments for this prior is illustrated in Figure **??**, where the dotted lines show the facilitator's pointwise 95% intervals for the expert's distribution function. Similar to the beta distribution example, accounting for this imprecision has resulted in a wider range of distributions that might be consistent with the expert's beliefs.

num. of chips	$P_{(0,2)}$	$P_{(2,4)}$	$P_{(4,6)}$	$P_{(6,\infty)}$
10	0.10	0.30	0.30	0.30
20	0.05	0.30	0.35	0.30

Table 3 Stated probabilities given the combinations of chips and bins. True probabilities are generated from a G(5,1) distribution.

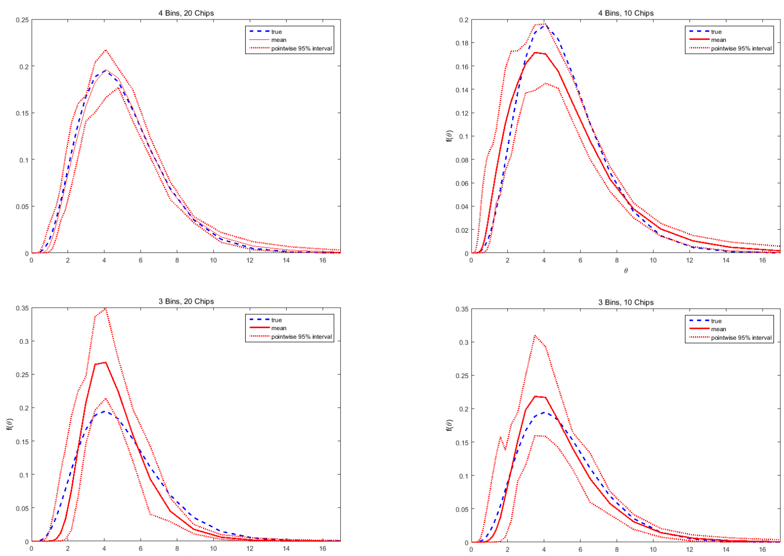

Fig. 3 Pointwise 95% intervals for the expert's density function (dotted lines), and the G(5,1) density function (dashed lines). Solid lines show the expert's estimated density using the non-parametric method.

We also study what such intervals would be developed, when the expert state her true probability statements using different number of bins and chips. From Figure 3, it can be concluded that the facilitator's uncertainty about the expert's PDF will grow by decreasing the number of chips and bins. This is simply because the expert is forced to round her true probabilities to the nearest number using smaller number of chips. Additionally, when we provide small number of bins, she cannot freely decide to allocate the right number of chips in a large bin. In Oakley et al. (2010), we explored a close link between the number of chips and bins (used by expert to express her beliefs) and the uncertainty induced by the imprecision of experts' probabilities assessments.

6 Conclusions and Future Work

Acknowledging the imprecision in the expert's probabilities increases the uncertainty around the expert's true density. We have considered two ways to model imprecise probabilities: a uniform errors model; a Gaussian process model, as presented in (13).

We have used a very straightforward approach to implement the elicitation scheme so called trial roulette. However, it is crucial to investigate whether or not the trial roulette in itself is a good method for elicitation, and how it might work for unbounded parameters, in particular, when the stated probabilities are very imprecise.

References

1. Oakley, J. E. and O'Hagan, A. (2007). Uncertainty in prior elicitations: a nonparametric approach, *Biometrika*, **94**: 427–441.
2. Oakley, J. E., Daneshkhah, A., and O'Hagan, A. (2012). Nonparametric Prior Elicitation using the Roulette, submitted to Bayesian Analysis.
3. O'Hagan, Buck, Daneshkhah, Eiser, Garthwaite, Jenkinson, Oakley and Rakow]O'Hagan:2006 O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D., Oakley, J. E. and Rakow, T. (2006). *Uncertain Judgements: Eliciting Experts' Probabilities*, Chichester: Wiley.
4. Gore, S. M. (1987). Biostatistics and the Medical Research council, *Medical Research Council News*.
5. Daneshkhah, A. and Sedighi, T. (2016) Nonparametric prior elicitation with imprecisely assessed probabilities (Working Paper)
6. Gosling, J. P., Oakley, J. E. and O'Hagan, A. (2007). Nonparametric elicitation for heavy-tailed prior distributions, *Bayesian Analysis*, **2**: 693–718.
7. Abbas, A. E., Budescu, D. V., Yu, H.-T. and Haggerty, R. (2008). A comparison of two probability encoding methods: fixed probability vs. fixed variable values, *Decision Analysis*, **5**: 190–202.
8. Abrams, K., Ashby, D. and Errington, D. (1994). Simple Bayesian analysis in clinical trials - a tutorial, *Controlled Clinical Trials*, **15**: 349–59.
9. Abrams, K. R. and Dunn, J. A. (1998). Discussion on the papers on elicitation, *Journal Of The Royal Statistical Society Series D*, **47**: 60–61.
10. Alpert, M. and Raiffa, H. (1982). A progress report on the training of probability assessors, in *Judgement and Uncertainty: Heuristics and Biases*, edited by Kahneman, D., Slovic, P. and Tversky, A., Cambridge: Cambridge University Press.
11. Berger, J. O. (1990). Robust Bayesian analysis: sensitivity to the prior, *Journal of Statistical Planning and Inference*, **25**: 303–328.
12. Bowman, A. W. and Crawford, E. (2008). *R package `rpanel`: simple control panels (version 1.0-5)*, University of Glasgow, UK.

13. Daneshkhah, A. and Oakley, J. E. (2010). Eliciting multivariate probability distributions, in *Rethinking Risk Measurement and Reporting: Volume I*, edited by Böcker, K., Risk Books, London.
14. French, S. (1980). Updating of belief in the light of someone else's opinion, *Journal of the Royal Statistical Society, Series A*, **143**: 43–48.
15. Garthwaite, P. H., Jenkinson, D. J., Rakow, T. and Wang, D. D. (2007). Comparison of fixed and variable interval methods for eliciting subjective probability distributions, Tech. rep., University of New South Wales.
16. Garthwaite, P. H., Kadane, J. B. and O'Hagan, A. (2005). Statistical methods for eliciting probability distributions, *J. Am. Statist. Assoc.*, **100**: 680–701.
17. Good, I. J. (1980). Some history of the hierarchical Bayesian methodology (with discussion), in *Bayesian Statistics*, edited by Bernardo, J. M., Groot, M. H. D., Lindley, D. V. and Smith, A. F. M., pp. 481–519, Valencia: University Press.
18. Hughes, M. D. (1991). Practical reporting of bayesian analyses of clinical trials, *Drug information journal*, **25**: 381–93.
19. Johnson, S. R., Tomlinson, G. A., Hawker, G. A., Granton, J. T., Grosbein, H. A. and m. Feldman, B. (2010). A valid and reliable belief elicitation method for bayesian priors, *Journal of Clinical Epidemiology*, **63**: 370–383.
20. Lindley, D. V., Tversky, A. and Brown, R. V. (1979). On the reconciliation of probability assessments (with discussion), *J. R. Statist. Soc A*, **142**: 146–180.
21. Moala, F. M. and O'Hagan, A. (2010). Elicitation of multivariate prior distributions: a nonparametric Bayesian approach, *J. Statist. Plan. and Infer.*, **140**: 1635–1655.
22. Morgan, M. G. and Henrion, M. (1990). *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*, Cambridge: Cambridge University Press.
23. Murphy, A. H. and Winkler, R. L. (1974). Credible interval temperature forecasting: some experimental results, *Monthly Weather Review*, **102**: 784–794.
24. Oakley, J. E. (2010). Eliciting univariate probability distributions, in *Rethinking Risk Measurement and Reporting: Volume I*, edited by Böcker, K., Risk Books, London.
25. Oakley, J. E. and O'Hagan, A. (2010). *SHELF: the Sheffield Elicitation Framework (version 2.0)*, School of Mathematics and Statistics, University of Sheffield, http://tonyohagan.co.uk/shelf.
26. Parmar, M. K. B., Spiegelhalter, D. J. and Freedman, L. S. (1994). The CHART trials: Bayesian design and monitoring in practice, *Statistics in Medicine*, **13**: 1297–312.
27. Parmar, M. K. B., Ungerleider, R. S. and Simon, R. (1996). Assessing whether to perform a confirmatory randomised clinical trial, *Journal of the National Cancer Institute*, **88**: 1645–51.
28. R Development Core Team (2010). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.
29. Spiegelhalter, D. J., Abrams, K. R. and Myles, J. P. (2004). *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*, Chichester: Wiley.
30. Tan, S.-B., Chung, Y.-F., Tai, B.-C., Cheung, Y.-B. and Machin, D. (2003). Elicitation of prior distributions for a phase III randomized controlled trial of adjuvant therapy with surgery for hepatocellular carcinoma, *Controlled Clinical Trials*, **24**: 110–121.
31. Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities*, London: Chapman and Hall.