Bayesian Analysis (2007) **2**, Number 4, pp. 693–718

Nonparametric elicitation for heavy-tailed prior distributions

John Paul Gosling* Jeremy E. Oakley† and Anthony O'Hagan‡

Abstract. In the context of statistical analysis, elicitation is the process of translating someone's beliefs about some uncertain quantities into a probability distribution. The person's judgements about the quantities are usually fitted to some member of a convenient parametric family. This approach does not allow for the possibility that any number of distributions could fit the same judgements.

In this paper, elicitation of an expert's beliefs is treated as any other inference problem: the facilitator of the elicitation exercise has prior beliefs about the form of the expert's density function, the facilitator elicits judgements about the density function, and the facilitator's beliefs about the expert's density function are updated in the light of these judgements. This paper investigates prior beliefs about an expert's density function and shows how many different types of judgement can be handled by this method.

This elicitation method begins with the belief that the expert's density will roughly have the shape of a t density. This belief is then updated through a Gaussian process model using judgements from the expert. The method gives a framework for quantifying the facilitator's uncertainty about a density given judgements about the mean and percentiles of the expert's distribution. A property of Gaussian processes can be manipulated to include judgements about the derivatives of the density, which allows the facilitator to incorporate mode judgements and judgements on the sign of the density at any given point. The benefit of including the second type of judgement is that substantial computational time can be saved.

Keywords: Expert elicitation, Gaussian process, heavy-tailed distribution, nonparametric density estimation.

1 Elicitation

In the context of statistical analysis, *elicitation* is the process of translating someone's beliefs about some uncertain quantities into a probability distribution. An elicited probability distribution can be used as a prior distribution in Bayesian analyses. Elicitation is an important subject: it has a part to play in every application where the data do not make the prior beliefs of the decision maker insignificant. Although Bernardo and Smith

*Department of Probability and Statistics, University of Sheffield, Sheffield, UK, mailto:j.p.gosling@sheffield.ac.uk
†Department of Probability and Statistics, University of Sheffield, mailto:j.oakley@sheffield.ac.uk
‡Department of Probability and Statistics, University of Sheffield, mailto:a.ohagan@sheffield.ac.uk

 ba0008

(1994) argue that the elicitation of prior beliefs is a mathematical idealization and sensitivity analysis is of greater importance, we take the view that prior beliefs are an important part of statistical analyses and that effort should be made to model them accurately.

Elicitation is far from being a precise science. It can be difficult for experts to articulate their beliefs. There are also other complications due to the possible biases of the experts and the biases created by the questioning process. The process of questioning people about their beliefs is certainly not a new subject: it has been the focus of many psychological studies. The psychological aspect of the elicitation process will not be formally considered in this paper. There is a vast amount of literature on the subject from a psychological perspective; two important reviews are Kahneman and Tversky (1973), and Hogarth (1975). Reviews of the psychological literature from a statistical viewpoint can be found in Kadane and Wolfson (1998) and Garthwaite et al. (2005).

There are many applications of elicitation techniques: they are not just used to obtain prior distributions when deriving posterior distributions. When making a decision, people often consider situations that have not happened before; this means that there is little or no relevant information available. Therefore, the expert's opinion is paramount in the decision making process. Gustafson et al. (2003) elicited expert opinions about organisational change; this was a situation where there were no data to use. Dominitz (1998) elicited opinion about future wages and applied a log-normal model; again, data were not available.

In risk assessment, elicited opinions are usually all a decision maker has to base a decision on when considering rare events. However, instead of eliciting distributions to model experts' beliefs, they often simply elicit point estimates. Ang and Buttery (1997) elicited assessments about the safety of nuclear power plants. Judgements about hazards in the workplace were elicited by Ramachandran et al. (2003); in this case, data were available, but the experts' opinions had a great impact on the results. Further examples of elicitation techniques can be found in Grisley and Kellogg (1983) and Smith and Mandac (1995) for agricultural applications; Cairns and Shackley (1999), Chaloner and Rhame (2001), and Must et al. (2002) for medical applications; and Coolen et al. (1992) and Sexsmith (1999) for engineering applications. O'Hagan et al. (2006) gives an extended overview of all these applications and more.

We consider the elicitation of a single expert's beliefs about some unknown continuous variable. For convenience of exposition, we consider a facilitator of the elicitation exercise and suppose that the facilitator wishes to make inferences about the expert's density function. (To avoid convoluted language, we let the expert be female and the facilitator be male.)

It is assumed in this paper that the expert cannot state a density function explicitly. She can only state certain summaries of the distribution such as the mean or various percentiles. However, these elicited summaries do not identify her distribution uniquely. Most of the applications of elicitation listed in the previous paragraphs employ *parametric* techniques where the facilitator fits elicited summaries to a distribution that is a member of some specified parametric family. The resulting distribution or summaries from that distribution can then be used to verify that the proposed distribution actually fits with her beliefs. It is conceivable that any number of distributions would be accepted as her distribution. We take the view that the fitting of a distribution to someone's beliefs is uncertain, and we should try to represent this uncertainty as accurately as we can. To do this, all the density functions that the facilitator believes are consistent with the expert's judgements should be considered. It is worth noting at this stage that we are only interested in his beliefs about $f(.)$ and not in his beliefs about the quantity of interest given her judgements.

In this paper, a *nonparametric* technique is used to reveal the facilitator's uncertainty about the final form of the expert's true probability distribution. This method is presented in Oakley and O'Hagan (2007). The idea is to treat the elicitation process as any other inference problem: the facilitator has prior beliefs about what form the expert's probability distribution will take, he obtains data in the form of judgements from her, and then his beliefs about her true probability distribution are updated in the light of these judgements. There are deficiencies with the method presented in Oakley and O'Hagan (2007): the facilitator's uncertainty about the tails of the expert's density can be understated and there are situations where the method implies that he knows her true density with no uncertainty. The method presented in Section 2 addresses these two problems. An application of the method to beliefs about river discharge in Netherlands is reported in Section 3.

2 Nonparametric elicitation

The facilitator's uncertainty about the expert's density function for some continuous quantity θ is to be modelled. Her density function for θ is $f(\theta)$; this function is assumed to be smooth and infinitely differentiable everywhere. When he elicits information from her about $f(.)$, his beliefs about the form of $f(.)$ are updated. She is not expected to accurately report the value of $f(\theta)$ for all possible values of θ. In fact, she should not be expected to be able to report $f(\theta)$ for any value of θ. Kadane and Wolfson (1998) suggest that quantiles or probabilities should be elicited.

First, consider the facilitator's prior beliefs about $f(.)$. The method does not assume a parametric form for $f(.)$. The Gaussian process model, which was introduced by Kimeldorf and Wahba (1970), Blight and Ott (1975), and O'Hagan (1978) as a tool for nonparametric curve fitting, is suggested as an appropriate representation of his prior beliefs. By using a Gaussian process model, his beliefs about $f(\theta)$ can be represented by a normal distribution for each θ. This leads to $f(\theta)$ having some probability of being negative, which is not desirable for a density function. To resolve this, we effectively condition on $f(\theta) > 0$ for all θ: the facilitator's posterior distribution is truncated at zero using simulation.

This Gaussian process model can be specified entirely through first- and second-order moments; hence, a parametric form must be specified for the facilitator's prior expectation of $f(\theta)$ and the prior covariance between $f(\theta)$ and $f(\theta')$. An appropriate structure is

$$E\left[f(\theta)\,\middle|\mathbf{w}, b^*, \sigma^2\right] = g(\theta|\mathbf{w}), \tag{1}$$

where $g(.)$, the *underlying* density, is a probability density function with parameters given in $\mathbf{w}$, and

$$Cov\left[f(\theta), f(\phi)\,\middle|\mathbf{w}, b^*, \sigma^2\right] = \sigma^2 g(\theta|\mathbf{w})g(\phi|\mathbf{w})c(\theta, \phi|\mathbf{w}, b^*). \tag{2}$$

A t density is used as the underlying density with location m, scale v and degrees of freedom d; this will allow for various tail behaviour. When a t density is employed as the underlying density, the facilitator must believe that $f(.)$ is similar to a t density.

In equation 2, $c(., .|\mathbf{w}, b^*)$ is a correlation function that takes the value 1 at $\theta = \phi$ and is a decreasing function of $|\theta - \phi|$. A Gaussian correlation function is used:

$$c(\theta, \phi|\mathbf{w}, b^*) = \exp\left\{-\frac{1}{2vb^*}(\theta - \phi)^2\right\}, \tag{3}$$

where b^* is the smoothness parameter. This correlation function makes $f(.)$ infinitely differentiable with probability 1, a property that is exploited later. The proof of this property of continuous stochastic processes is given in Belyaev (1959).

2.1 Prior beliefs about the Gaussian process hyperparameters

The facilitator's beliefs about the model hyperparameters must be specified before anything is elicited from the expert. It is important that his knowledge about θ does not influence the analysis: we are only interested in the opinion of the expert. Also, before he elicits information from the expert, he does not know the location or scale of her density function. Hence, he has vague prior beliefs about m and v such that

$$p(m, v) \propto v^{-1}. \tag{4}$$

His beliefs about the degrees of freedom parameter should be handled in a similar fashion: the facilitator does not know what value this parameter will take *a priori*. Hence, the prior distribution for d is set to be uniform over 0 to 40. This range is used because a t distribution with degrees of freedom greater than 40 is practically indistinguishable from a normal distribution. This uniform prior distribution for d is robust, and it has been found that a log-normal distribution for d produces indistinguishable results. The facilitator's prior distribution for all the underlying distribution's parameters is $p(m, v, d) \propto v^{-1}$ for $0 \leq d \leq 40$.

An improper prior distribution for the Gaussian process variance σ^2 can lead to the facilitator's posterior distribution for $f(.)$ showing that the facilitator knows $f(.)$ with no uncertainty with only a few judgements from the expert. The facilitator has prior beliefs about the smoothness of $f(.)$ and how far $f(.)$ is expected to deviate from $g(.)$, which imply beliefs about b^* and σ^2. The dependence between these two hyperparameters is illustrated by considering the form of functions that are allowed when using the Gaussian process model.

The facilitator knows that $f(.)$ is a density function; hence, he must believe that $f(.)$ is positive for all values of θ. By using the Gaussian process model of Section 2, any function simulated from the facilitator's distribution for $f(.)$ will be smooth and infinitely differentiable and a simulated function could at some point be negative. In addition to this, he believes that the expert's density will be roughly unimodal; this is reflected in the choice of a t distribution as the underlying distribution. Therefore, the facilitator believes that highly multimodal densities are unlikely. The values of b^* and σ^2 influence all of these properties of the simulated $f(.)$.

The derivation of a proper, joint distribution for b^* and σ^2 is given in the Appendix. The prior distribution we will employ is given by

$$\begin{aligned} p(\sigma^2, b^*) &= p(\sigma^2|b^*)p(b^*), \\ \log(b^*) &\sim N(0.56, 0.27^2), \\ \log(\sigma^2)|b^* &\sim N(M, S), \end{aligned} \tag{5}$$

where M and S, which are functions of b^*, are defined in the Appendix. This choice of prior distribution matches the facilitator's prior beliefs that the expert's density will not be highly multimodal and that the density will not be known for sure given just a few judgements.

2.2 Updating the facilitator's beliefs about $f(.)$

The facilitator's beliefs about $f(.)$ are updated using judgements from the expert about the percentiles of the distribution for θ. In this case, the vector of judgements D is

$$\begin{aligned} D^T &= \left(\int_{x_0}^{x_1} f(x)dx, \ldots, \int_{x_{n-1}}^{x_n} f(x)dx\right) \\ &= \left(P_{x_0,x_1}, \ldots, P_{x_{n-1},x_n}\right), \end{aligned} \tag{6}$$

where the *interval points*, $x_0 < x_1 < \cdots < x_n$, are possible values of θ. As $f(.)$ is a Gaussian process, then P_{x_i,x_j} follows a normal distribution. The expectation of D is given by

$$\begin{aligned} E[D|\mathbf{w}]^T &= \left(\int_{x_0}^{x_1} g(x|\mathbf{w})dx, \ldots, \int_{x_{n-1}}^{x_n} g(x|\mathbf{w})dx\right) \\ &= H^T. \end{aligned} \tag{7}$$

Equation 2 is extended to get

$$Cov(f(\theta), P_{x_i,x_j}|\mathbf{w}, b^*, \sigma^2) = \sigma^2 g(\theta|\mathbf{w}) \int_{x_i}^{x_j} g(x|\mathbf{w})c(\theta, x|b^*, v)dx, \tag{8}$$

and

$$Cov(P_{x_i,x_j}, P_{y_i,y_j}|\mathbf{w}, b^*, \sigma^2) = \sigma^2 \int_{y_i}^{y_j} \int_{x_i}^{x_j} g(x|\mathbf{w})g(y|\mathbf{w})c(x, y|b^*, v)dxdy. \tag{9}$$

This covariance structure is based on work by O'Hagan (1991) on Bayes-Hermite quadrature. The facilitator's beliefs about $f(.)$ are updated using the conditioning property of multivariate normal distributions.

It follows immediately from this property that the facilitator's beliefs about $f(.)$ conditional on $D, \mathbf{w}, b^*$ and σ^2 is also a Gaussian process with

$$E[f(\theta)|D, \mathbf{w}, b^*, \sigma^2] = g(\theta|\mathbf{w}) + \mathbf{t}(\theta|\mathbf{w}, b^*, \sigma^2)^T A^{-1}(D - H), \tag{10}$$

and

$$\begin{aligned} Cov(f(\theta), f(\theta')|D, \mathbf{w}, b^*, \sigma^2) \quad &= \sigma^2 \left(g(\theta|\mathbf{w})g(\theta'|\mathbf{w})c(\theta, \theta'|b^*, v) \right. \\ &\left. - \mathbf{t}(\theta|\mathbf{w}, b^*, \sigma^2)^T A^{-1}\mathbf{t}(\theta'|\mathbf{w}, b^*, \sigma^2) \right), \end{aligned} \tag{11}$$

where A is an $n \times n$ matrix with entries given by:

$$A_{ij}|\, \mathbf{w}, b^*, \sigma^2 = Cov\left(P_{x_i,x_{i+1}}, P_{x_j,x_{j+1}} \middle|\, \mathbf{w}, b^*, \sigma^2\right), \tag{12}$$

and $\mathbf{t}(\theta|\mathbf{w}, b^*, \sigma^2)$ is given by

$$\begin{aligned} \mathbf{t}(\theta|\mathbf{w}, b^*, \sigma^2)^T \quad &= \left(Cov\ \left(f(\theta), P_{x_0,x_1}|\, \mathbf{w}, b^*, \sigma^2\right), \ldots, \right. \\ &\left. Cov\ \left(f(\theta), P_{x_{n-1},x_n} \middle|\, \mathbf{w}, b^*, \sigma^2\right)\right). \end{aligned} \tag{13}$$

Conditional on the hyperparameters and the expert's judgements, $f(.)$ is a Gaussian process with equations 10 and 11 giving its mean and covariance structure respectively.

The conditioning on the model hyperparameters cannot be removed analytically; instead, MCMC is used to obtain a sample of values from their joint posterior distribution. Given a set of values for the hyperparameters, a density function is sampled at a finite number of values of θ from the Gaussian process model. By repeating this many times, a sample of functions from $f(.)|D$ is obtained and the negative-valued functions are removed; the simulation process is described in Oakley and O'Hagan (2002). The remaining functions are used to report estimates and pointwise credible bounds for the expert's density.

2.3 Example 1(a)

This example shows the benefit of using a t distribution as the underlying distribution rather than a normal distribution. The expert has the following density for the parameter θ:

$$f(\theta) = \frac{0.7}{\pi}\frac{2}{4 + (\theta + 1)^2} + \frac{0.3}{\pi}\frac{1}{1 + \theta^2}. \tag{14}$$

This is a mixture of two Cauchy distributions. We assume that the expert can report P_{10}, P_{25}, P_{50}, P_{75} and P_{90} without error. In addition to these judgements, it is known that $P_{-\infty,\infty} = 1$.

MCMC is now used to obtain samples of the hyperparameters from their joint posterior distribution. A Metropolis-Hastings sampler is employed to propose values for m, v, d, b^* and σ^2 simultaneously. The following proposal distributions are used for both underlying distributions:

$$\begin{aligned} m_t|m_{t-1} &\sim N\left(m_{t-1}, 0.01\right), \\ \log v_t|v_{t-1}, m_{t-1}, m_t &\sim N\left(\log v_{t-1}, 0.1 + \left(|m_t - m_{t-1}|/2\right)\right), \\ \log b_t^*|b_{t-1}^* &\sim N\left(\log b_{t-1}^*, 0.01\right), \\ \log \sigma_t^2|\sigma_{t-1}^2 &\sim N\left(\log \sigma_{t-1}^2, 0.01\right). \end{aligned} \tag{15}$$

The proposal distribution for v_t is dependent on m_{t-1} and m_t so that large jumps in m_t are more likely to be accompanied by large jumps in v_t. Also, a proposal distribution for d is needed for the underlying t distribution:

$$\log d_t|d_{t-1} \sim N\left(\log d_{t-1}, 0.01\right). \tag{16}$$

The chain is run for 20,000 iterations and the first 10,000 runs are discarded to allow for the burn-in period. For each of the last 10,000 runs, a random density function is generated.

The generated density functions are used to plot a representation of the facilitator's beliefs about $f(.)$. Figure 1 is the plot of the pointwise median, 2.5th and 97.5th percentiles from the distribution of the function after the 34% of the random functions that were negative using an underlying normal distribution are discarded.

By considering the positive tail of the expert's distribution, the benefit of using the underlying t distribution is clear. Figures 3 and 4 show the pointwise median, 2.5th and 97.5th percentiles from the positive tail of $f(\theta)$. When the underlying distribution is a normal distribution, the credible intervals are not wide enough to allow for heavy-tailed densities. Figure 4 shows that the underlying t distribution does allow for this tail behaviour. Moreover, the credible intervals are wider towards the end of the range of θ values; this is representative of greater uncertainty about tail behaviour.

2.4 Making use of information about derivatives

It is not feasible to ask the expert to report the value of the derivative of $f(.)$ for different values of θ, but there are some simple judgements the facilitator can ask the expert to make to arrive at information about the derivatives of $f(.)$. In this section, the work given in O'Hagan (1992) on incorporating information about derivatives in a Gaussian process regression model is extended. The covariance between $f(\theta)$ and $df(\theta)/d\theta$ is calculated using

$$Cov\left(f(\theta), \frac{df(\theta')}{d\theta}\middle| \mathbf{w}, b^*, \sigma^2\right) = \sigma^2 g(\theta|\mathbf{w}) \frac{dg(\theta'|\mathbf{w})c(\theta,\theta'|b^*,v)}{d\theta'}, \tag{17}$$

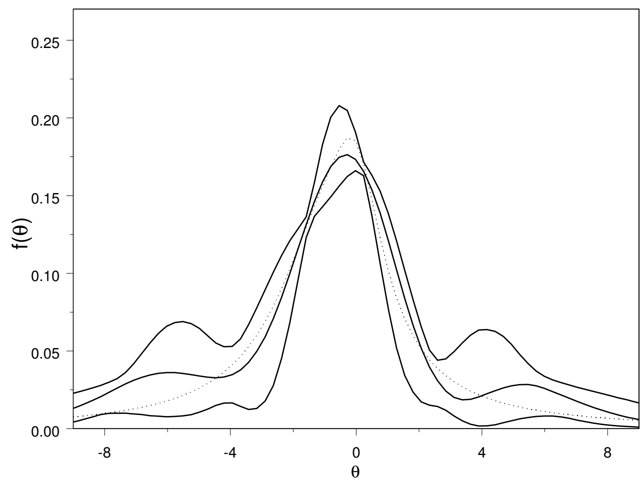

Figure 1: The median and pointwise 95% credible intervals for the expert's density function using an underlying normal distribution (solid lines) and the true density function (dotted line).

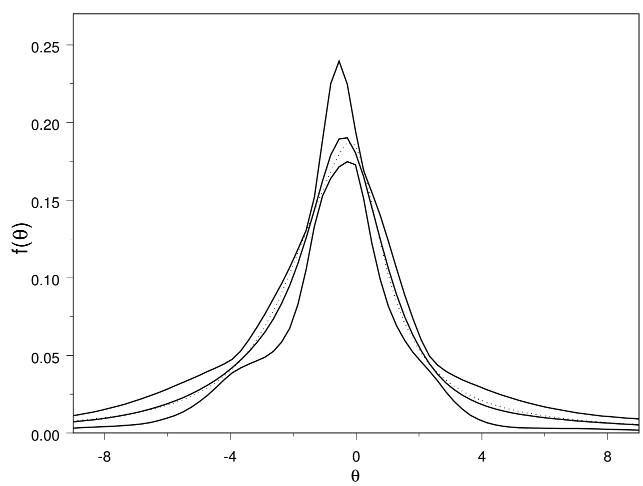

Figure 2: The median and pointwise 95% credible intervals for the expert's density function using an underlying t distribution (solid lines) and the true density function (dotted line).

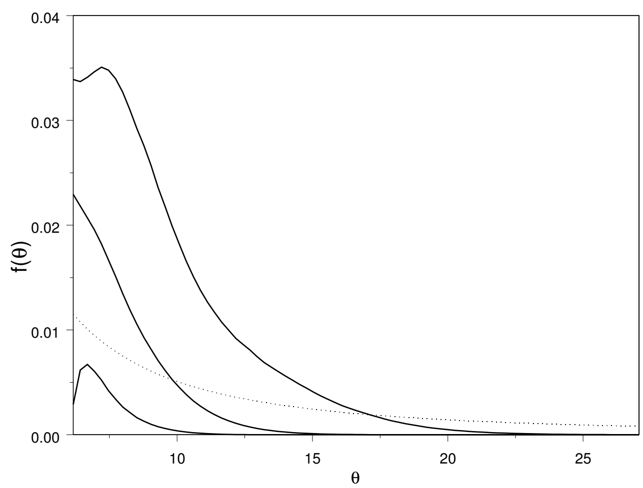

Figure 3: The median and pointwise 95% credible intervals for the expert's density function using an underlying normal distribution (solid lines) and the true density function (dotted line).

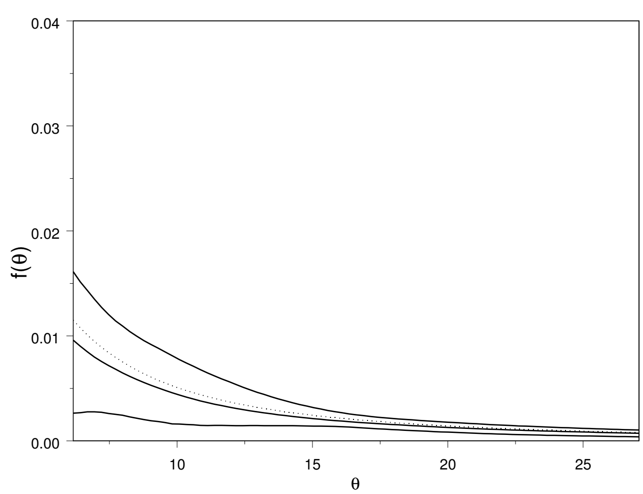

Figure 4: The median and pointwise 95% credible intervals for the expert's density function using an underlying t distribution (solid lines) and the true density function (dotted line).

and, for the covariance between two derivatives of $f(.)$,

$$Cov\left(\frac{df(\theta)}{d\theta}, \frac{df(\theta')}{d\theta}\middle| \mathbf{w}, b^*, \sigma^2\right) = \sigma^2 \frac{d^2 g(\theta|\mathbf{w})g(\theta'|\mathbf{w})c(\theta,\theta'|b^*,v)}{d\theta d\theta'}. \tag{18}$$

The corresponding covariance function to compare observations of the derivative of $f(.)$ and P_{x_i,x_j} must be derived. The theory follows through in an expected manner:

$$Cov\left(\frac{df(\theta)}{d\theta}, P_{x_i,x_j}\middle| \mathbf{w}, b^*, \sigma^2\right) = \sigma^2 \int_{x_i}^{x_j} g(x|\mathbf{w}) \frac{dg(\theta|\mathbf{w})c(\theta,x|b^*,v)}{d\theta} dx. \tag{19}$$

Equations 17 through to 19 specify all the covariance functions needed to determine the covariance matrix A and the vector $\mathbf{t}(.)$. Observations of the type $df(\theta)/d\theta = y$ can now be included.

The easiest property of the distribution for the expert to report is its mode. If the mode of the expert's density is denoted by M and there is the assumption that the density decays in both the negative and positive tails, it follows that $df(M)/d\theta = 0$, that is, a stationary point at $\theta = M$.

The data vector D of the expert's judgements is changed by adding an extra component:

$$D^T = \left(P_{x_0,x_1}, \ldots, P_{x_{n-1},x_n}, 0\right). \tag{20}$$

The added zero in this vector corresponds to the derivative of the density at the mode. It is worth noting that the added judgement does not necessarily give a mode at the desired point. As the Gaussian process model interpolates the expert's judgements exactly, $df(M)/d\theta = 0$. Hence, there will definitely be a stationary point of some kind at M in functions drawn from the facilitator's posterior distribution for $f(.)$, but this may not be a mode. In section 2.6, we add information about second derivatives to counteract this.

The expectation $E[D|\mathbf{w}]$ must also be extended to allow for the new derivative observation. By using the derivative of the underlying function at M, the facilitator's prior expectation is

$$E\left[\frac{df(M)}{d\theta}\middle| \mathbf{w}\right] = \frac{dg(M|\mathbf{w})}{d\theta}. \tag{21}$$

His prior expectation for D becomes

$$\begin{aligned} E[D^T|\mathbf{w}] &= H^T \\ &= \left(\int_{x_0}^{x_1} g(x|\mathbf{w})dx, \ldots, \int_{x_{n-1}}^{x_n} g(x|\mathbf{w})dx, \frac{dg(M|\mathbf{w})}{d\theta}\right), \end{aligned} \tag{22}$$

which is a simple extension of equation 7. With these changes, the method follows through in the way described earlier.

The mode of the expert's distribution for θ is not the only stationary point that the facilitator could elicit from the expert. If her density is multimodal, then he may expect her to be able to report the position of these local modal points. He could also ask her for any antimodes in $f(.)$. However, as in the mode case, the inclusion of a zero derivative at a point does not force the realisations from the facilitator's posterior distribution to have an antimode at that point.

2.5 Example 1(b)

In this example, the expert's density for θ is given by

$$f(\theta) = \frac{0.4}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(\theta+2)^2\right\} + \frac{0.6}{\sqrt{4\pi}} \exp\left\{-\frac{1}{4}(\theta-1)^2\right\}, \tag{23}$$

which is a mixture of two normal distributions. The expert can report $P_{-\infty,-3}$, $P_{-3,-1}$, $P_{-1,1}$ and $P_{1,3}$. Figure 5 is the plot of the pointwise median, 2.5th and 97.5th percentiles from the facilitator's posterior distribution for the function after applying the method to these judgements. In addition to these judgements, the expert can also report what they believe is the most likely value for θ, which is denoted by M. The modal point is approximately -1.78 for this distribution. Figure 6 is the plot of the pointwise median, 2.5th and 97.5th percentiles from the facilitator's posterior distribution for the function.

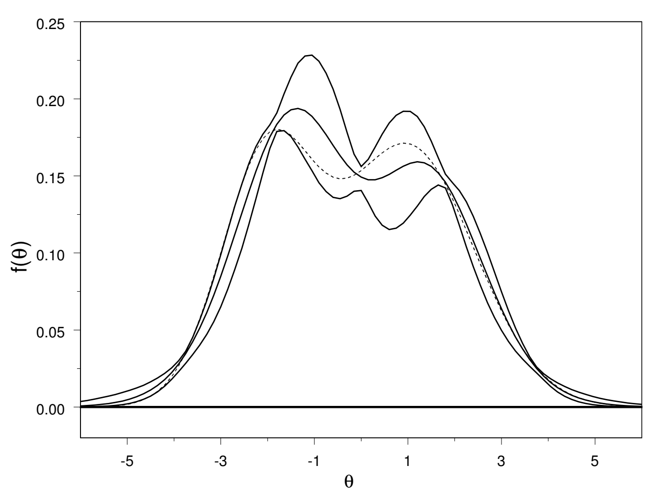

Figure 5: The median and pointwise 95% credible intervals for the expert's density function (solid lines), and the true density function (dotted line).

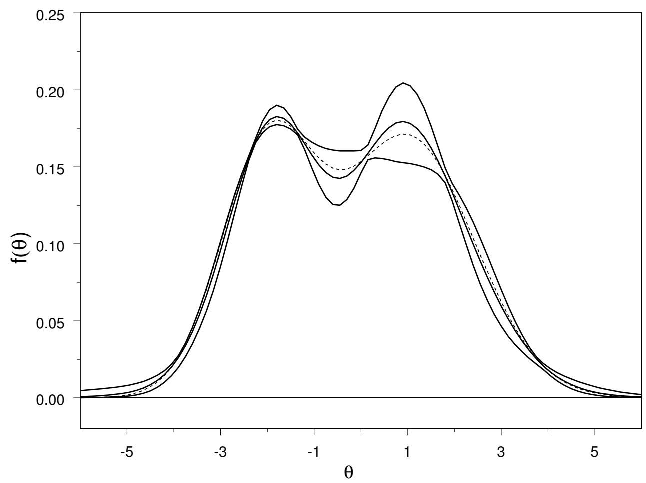

Figure 6: The median and pointwise 95% credible intervals for the expert's density function (solid lines), and the true density function (dotted line) after adding the mode judgement.

It is clear from Figure 6 that the modal point is a local maximum in the facilitator's posterior distribution for $f(.)$. However, for some of the functions drawn from the facilitator's posterior distribution for $f(.)$, a higher maximum occurs at about $\theta = 1.2$. The mode observation added in to the model only forces all functions drawn from the facilitator's posterior distribution to have a stationary point at M. It is possible to discard all functions where $f(M)$ is not the overall mode. In this case, this would lead to about 40% of the simulated functions being discarded. If this is coupled with the removal of all functions that become negative at some point, approximately 50% of functions will be discarded. This increases the overall time needed for the computational part of the method.

2.6 Incorporating sign information

More alterations to the method are required to prevent the reported modal points from being treated as antimodes by the model. For there to be a mode at $\theta = M$, two conditions must be met:

$$\frac{d\,f(M)}{d\theta} = 0 \tag{24}$$

$$\text{and} \quad \frac{d^2\,f(M)}{d\theta^2} < 0. \tag{25}$$

The model has already been developed to include the information given in equation 24 and will be developed further to include the information given in equation 25 in this section.

The covariance between the function and its second derivative is given by

$$Cov\left(\frac{d^2\,f(\theta)}{d\theta^2}, f(\phi)\middle|\,\mathbf{w}, b^*, \sigma^2\right) = \sigma^2 g(\phi|\mathbf{w})\frac{d^2\left(g(\theta|\mathbf{w})c(\theta, y|b^*, v)\right)}{d\theta^2}. \tag{26}$$

Similarly,

$$Cov\left(P_i, \frac{d^2\,f(\theta)}{d\theta^2}\middle|\,\mathbf{w}, b^*, \sigma^2\right) = \sigma^2 \int_{-\infty}^{x_i} g(\phi|\mathbf{w})\frac{d^2\left(g(\theta|\mathbf{w})c(x, \phi|b^*, v)\right)}{d\theta^2}d\phi, \tag{27}$$

for the covariance between a second derivative observation.

The covariance between a first derivative observation and a second derivative observation is

$$Cov\left(\frac{d^2\,f(\theta)}{d\theta^2}, \frac{d\,f(\phi)}{d\phi}\middle|\,\mathbf{w}, b^*, \sigma^2\right) = \sigma^2\frac{d^3\left(g(\theta|\mathbf{w})g(\phi|\mathbf{w})c(\theta, \phi|b^*, v)\right)}{d\theta^2 d\phi}, \tag{28}$$

and the covariance between two second derivative observations is

$$Cov\left(\frac{d^2\,f(\theta)}{d\theta^2}, \frac{d^2\,f(\phi)}{d\phi^2}\middle|\,\mathbf{w}, b^*, \sigma^2\right) = \sigma^2\frac{d^4\left(g(\theta|\mathbf{w})g(\phi|\mathbf{w})c(\theta, \phi|b^*, v)\right)}{d\theta^2 d\phi^2}. \tag{29}$$

It should also be noted that the facilitator's prior expectation is

$$E\left[\frac{d^2\,f(\theta)}{d\theta^2}\middle|\,\mathbf{w}\right] = \frac{d^2\,g(\theta|\mathbf{w})}{d\theta^2}. \tag{30}$$

If the expert reports a mode at M and the second derivative at this point is to be negative, the condition that $d^2\,f(M)/d\theta^2 < 0$ must be included when functions are drawn from the facilitator's posterior distribution for $f(.)$. The new judgement type changes the calculation of the facilitator's posterior distribution for $f(.)$, and the theory underpinning the method has to be verified.

All of the judgements about $f(.)$ that the facilitator has elicited from the expert have followed normal distributions so far. This fact has allowed properties of multivariate normal distributions to be utilised. However, if a condition is placed on the sign of the derivative of $f(.)$ at a point, then the Gaussian process model leads to a truncated normal distribution. The partition property of multivariate normal distributions, which was mentioned in Section 2.2, holds for the partially truncated multivariate normal distributions despite the marginal of a partially truncated multivariate normal distribution being a partially truncated multivariate skew normal distribution as given in O'Hagan and Leonard (1976).

Consider $\mathbf{z}$, which is an $n \times 1$ vector that has a truncated multivariate normal distribution with mean $\boldsymbol{\mu}$ and variance-covariance matrix Σ. For the ith element of $\mathbf{z}$, z_i, there is a_i and b_i that define the lower and upper truncation points of z_i respectively. If there is no upper (or lower) truncation point in the ith dimension, then $b_i = \infty$ (or $a_i = -\infty$). The density for $\mathbf{z}$ is given by

$$p(\mathbf{z}) = \frac{\exp\left\{-\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{z}-\boldsymbol{\mu})\right\}}{\int_Z \exp\left\{-\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{z}-\boldsymbol{\mu})\right\} d\mathbf{z}},$$

where Z defines the region over which z can exist, that is, inside the truncation points. If $\mathbf{z}$ is partitioned into two vectors $\mathbf{z}_1$ and $\mathbf{z}_2$, it can then be proved that both $\mathbf{z}_1|\mathbf{z}_2$ and $\mathbf{z}_2|\mathbf{z}_1$ also have truncated multivariate normal distributions where $\mathbf{z}_1$ and $\mathbf{z}_2$ preserve the truncation points given for $\mathbf{z}$. The proof of this using characteristic functions is given in Horrace (2005).

Most of the extra computation time when using this type of judgement is taken up by the evaluation of the distribution of $\mathbf{w}, b, \sigma^2|D$; the posterior distribution is given by

$$\begin{aligned} p(\mathbf{w}, b^*, \sigma^2|D) \quad &\propto \quad p(\mathbf{w}, b^*, \sigma^2) \\ &\times \quad \int_0^\infty \exp\left(-\frac{1}{2}\left((D-H)^T A^{-1}(D-H)\right)\right) df''(M), \end{aligned} \qquad (31)$$

where the $df''(M)$ corresponds with the point that is being conditioned on. In this case, the integral is of singular dimension; hence, it is relatively simple to evaluate this distribution. However, if more than one sign observation is included, the integral in equation 31 becomes multidimensional.

When drawing random functions from the facilitator's posterior distribution for $f(.)$, the fact that the second derivative is positive at $\theta = M$ must be taken into account. Hence, when simulating the functions, the extra information that

$$\frac{d^2 f(M)}{d\theta^2} \quad < \quad 0 \qquad (32)$$

is included. $d^2 f(M)/d\theta^2$ follows a normal distribution if the information given in equation 32 is not conditioned on. To condition on the information, the distribution that is drawn from must be truncated at zero. When a derivative is to be negative at a point, a sample must be taken from a negative distribution that has been truncated from its positive part and *vice versa*. In order to do this, a rejection sampling technique using an exponential density as the envelope function is employed.

2.7 Example 1(c)

Reconsider the example where the expert's density is a mixture of two normal distributions. Again, the expert can report $P_{-\infty,-3}$, $P_{-3,-1}$, $P_{-1,1}$ and $P_{1,3}$. There are two modes and one antimode that the expert can report. They are $M_1 \approx -1.78$, $M_2 \approx 0.92$ and $L \approx -0.42$ for the expert's distribution. The facilitator can now fix

$$\frac{d^2 f(M_1)}{d\theta^2} < 0, \quad \frac{d^2 f(M_2)}{d\theta^2} < 0, \quad \frac{d^2 f(L)}{d\theta^2} > 0, \tag{33}$$

using the theory presented in this section.

Figure 7 is the plot of the pointwise median, 2.5th and 97.5th percentiles from the facilitator's posterior distribution for the function. There has been a sizeable reduction in uncertainty from that shown in example 1(b). By adding the extra information, density functions that do not match the expert's beliefs are disallowed. Hence, the facilitator's uncertainty about $f(.)$ is modelled more accurately by using all the information that is available.

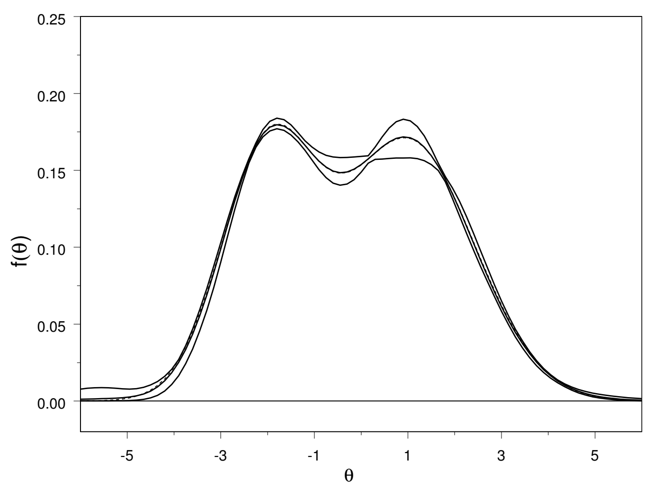

Figure 7: The median and pointwise 95% credible intervals for the expert's density function (solid lines), and the true density function (dotted line).

2.8 Information about the sign of $f(.)$

By using the same theoretical basis as for the second derivative of $f(.)$, sign judgements for any order of derivative of $f(.)$ and sign judgements made directly on $f(.)$ can be included.

A deficiency of a Gaussian process model for a density function is the significant probability of the function being negative in the facilitator's posterior distribution. If the information that $f(\theta)$ is positive at the design points is used when simulating from $f(.)|D$, the chances of the drawn function being negative can be restricted. Areas of the θ-axis where there are greater probabilities of $f(\theta)$ becoming negative can also be targeted. To do this, a sign judgement is added stating $f(\theta) \geq 0$ for a θ where a high probability of $f(\theta)$ becoming negative is expected.

The following example will show how functions drawn from the facilitator's posterior distribution can be prevented from being negative too often, that is, reducing the number of functions discarded due to them being negative. The judgements for this example are

$$P_{-\infty,-3} = 0.05, \quad P_{-3,-1} = 0.4, \quad P_{-1,1} = 0.1 \quad \text{and} \quad P_{1,3} = 0.4.$$

These probability judgements come from a symmetric bimodal distribution that has a relatively low valued antimode at zero.

The first graph of Figure 8 is the plot of the pointwise median, 2.5th and 97.5th percentiles from the facilitator's posterior distribution for the function. Notice that this plot reveals a large probability of the function being negative. In order to draw 10,000 functions that are positive for all θ in the area of interest, approximately 500,000 functions must be drawn. This whole process takes too long for the method to be useful.

To reduce the number of functions being drawn from the facilitator's posterior distribution for $f(.)$, information is added about the sign of the function at a point. The first graph of Figure 8 shows that there is a large proportion of functions being drawn that are negative at $\theta = 0$. The information that

$$f(0) > 0 \tag{34}$$

is added to prevent the function from becoming negative at $\theta = 0$. The second graph of Figure 8 is the plot of the pointwise median, 2.5th and 97.5th percentiles from the facilitator's posterior distribution for the function when the facilitator uses this extra information. In this case, 93,000 functions must be drawn from the facilitator's posterior distribution to obtain 10,000 functions that are positive for all θ in the area of interest.

The complex nature of incorporating sign judgements in the model slows the computational speed of the method. However, by including just one extra sign judgement, the number of negative valued functions being drawn can be significantly reduced or a reported mode for θ can be forced to be a mode in the facilitator's posterior distribution for $f(.)$.

3 Application: river discharge at Lobith

To demonstrate the technique on a real set of elicited judgements, we use part of a study done on Dutch dike rings and expected river discharge that is described in Frijters et al. (1999) and Cooke and Slijkhuis (2003). The study was interested in the uncertainty in

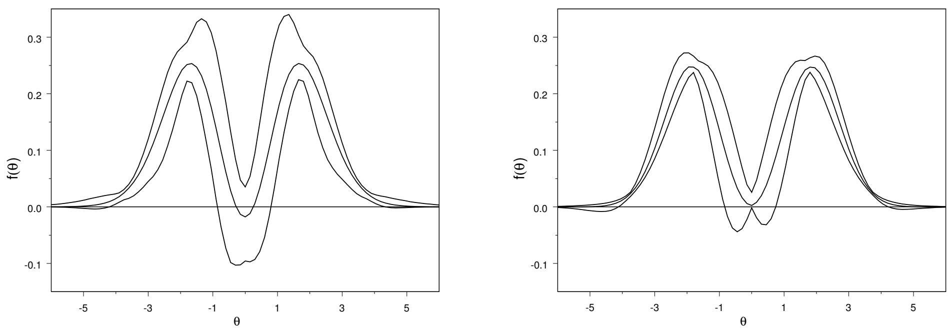

Figure 8: The median and pointwise 95% credible intervals for the expert's density function (solid lines) before and after $f(0) > 0$ is added.

the influence of climate change and human intervention (dams in the river, deforestation and urbanisation) on the extreme water discharges of the Rhine at Lobith, Netherlands over a period of 100 years amongst other flood risk variables. One uncertain quantity they elicited expert opinion on was the discharge measured 100 years from now given that a discharge of $16000m^3/s$ is measured at Lobith today and the same measurement technique will be used in 100 years; we will call this uncertain quantity θ. They asked the expert to state the 5th, 25th, 50th, 75th and 95th percentiles of their probability distribution for θ.

The expert gave the following percentiles from their distribution for θ:

$$P_5 = 15000m^3/s,\ P_{25} = 16500m^3/s,\ P_{50} = 17000m^3/s,$$
$$P_{75} = 17500m^3/s \quad \text{and} \quad P_{95} = 19000m^3/s.$$

Figures 9 and 10 show the results of using a Gaussian process prior with an underlying normal distribution and an underlying t distribution respectively. It can be seen that with an underlying normal distribution density functions with have local modes around $15000m^3/s$ and $17000m^3/s$. This is not consistent with the expert's beliefs about θ. The method with an underlying t distribution is more flexible, and the range of possible densities shown in figure 10 is more representative of our beliefs about the expert's density for θ.

4 Discussion and further work

After the facilitator elicits judgements from the expert about some parameter, his uncertainty about the true form of the expert's density for the parameter can be quantified using the method. By employing a t distribution as the underlying density, the expert's distribution is allowed to be heavy-tailed and the facilitator's uncertainty about the tails of the expert's distribution is no longer understated.

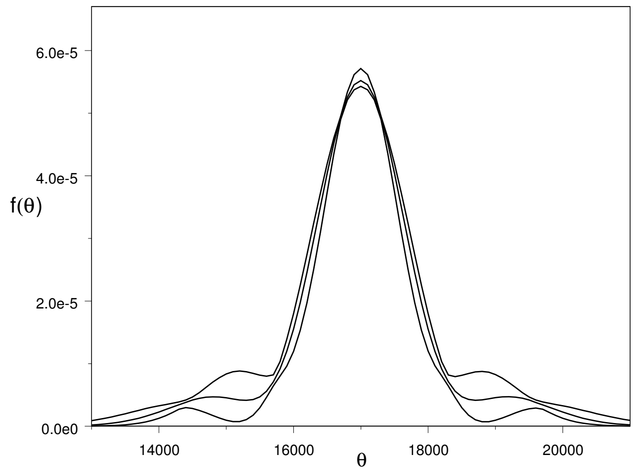

Figure 9: The median and pointwise 95% credible intervals for the expert's density function using an underlying normal distribution.

Once the facilitator's posterior distribution for $f(.)$ has been derived, any number of functions can be simulated from that distribution. This is done in order to see how the different functions affect the analyses stemming from the facilitator's beliefs about $f(.)$. In cases where the facilitator has elicited many judgements from the expert, the facilitator's posterior uncertainty about $f(.)$ will be small and the functions drawn from the facilitator's posterior distribution will be close to the posterior mean for $f(.)$. If the facilitator's posterior uncertainty is large, the facilitator could use the posterior distribution for $f(.)$ to choose intervals on which more judgements could be elicited from the expert. This sequential updating of the facilitator's beliefs could lead to better understanding of the expert's distribution.

The underlying distribution has an impact on the results of the method. Any distribution can be employed in this role provided it matches the facilitator's prior beliefs about the expert's density. For instance, when the parameter is strictly positive, a gamma or a log-normal distribution could work well as the underlying distribution. Ultimately, the judgements of the expert should dictate the facilitator's beliefs about $f(.)$. The influence of the underlying distribution is reduced by eliciting more judgements from the expert. However, the underlying distribution must be flexible enough to allow for the judgements from the expert otherwise computational difficulties will arise. The proper prior distribution for b^* and σ^2 developed in this paper was created for an underlying t distribution: further work is required to investigate the changes needed for different underlying distributions.

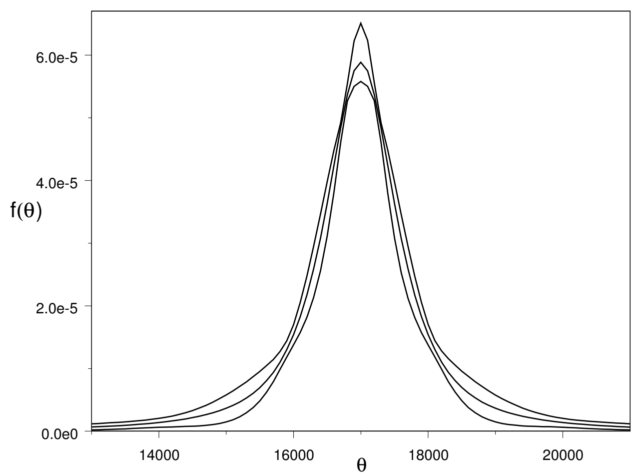

Figure 10: The median and pointwise 95% credible intervals for the expert's density function using an underlying t-distribution.

When a decision depends on expert opinion alone, it is unrealistic to depend on a distribution that is the mathematical idealization that Bernardo and Smith (1994) discuss in regards to elicitation. This paper shows how that the facilitator's uncertainty about $f(.)$ can be quantified and the great dependence on selecting an appropriate parametric family to fit an expert's judgements can be removed.

The judgements from the expert's distribution in the examples in this paper have been taken as being correct without any error. In practice, if the expert states that the modal value for some parameter is M, it may be difficult for them to justify why they use M instead of $M \pm \delta$ for sufficiently small δ. An important extension to this model is the accommodation of the uncertainty about the elicited judgements. The incorporation of uncertainty about the expert's judgements in this model is discussed in Oakley and O'Hagan (2007) and Daneshkhah et al. (2006).

This paper has only considered the elicitation of judgements about an univariate parameter in this paper, yet there are often multivariate prior distributions to consider. The original Gaussian process model for nonparametric regression has been used in multidimensional settings, for example, in Rasmussen (1996) and Wernisch (2004). The theory could be extended to cover multivariate prior elicitation.

GUI software has been developed that implements the techniques of this paper. *ROBEO* (Representing Our Beliefs about Expert Opinions) is currently available on the web as a test beta. For more details of this software and tutorials, see `http://j-p-gosling.staff.shef.ac.uk`.

Appendix: Prior beliefs about σ^2 and b^*

Here we develop a proper prior distribution for the hyperparameters σ^2 and b^*.

A sufficiently small σ^2 will bring the simulated functions close to the underlying function. Consider the ratio of a simulated function and the underlying function, $h(\theta) = f(\theta)/g(\theta)$, then $h(\theta)$ is being modelled as a stationary Gaussian process with a constant mean of one. It follows immediately that

$$h(\theta) \sim N(1, \sigma^2), \tag{35}$$

for all possible values of θ. Therefore, for small values of σ^2, the ratio of the two densities will be close to one. In contrast, the ratio could be far from one for large values of σ^2. This indicates a greater difference between $f(.)$ and the underlying density. Figure 11 shows five random functions drawn from the facilitator's prior distribution for $f(.)$ conditional on different values of σ^2 and b^* that demonstrates this behaviour. The underlying function $g(.)$ in these simulations is a standard normal density.

A relatively small smoothness parameter b^* implies that the value of the correlation function will be close to zero. This means that the dependence between neighbouring points of the function will diminish and functions drawn from the distribution will appear *rough*. Alternatively, a relatively large b^* will lead to a greater affinity between $f(.)$ and $g(.)$. The effect of different values of b^* can be seen in Figure 11.

To further investigate the impact of different values for σ^2 and b^*, the hyperparameters of the underlying distribution are given arbitrary values, then points on the (b^*, σ^2)-plane are selected. The hyperparameters of the underlying distribution are arbitrary because in the ratio of $f(.)/g(.)$ the effect of these hyperparameters is cancelled out. For each pair selected from the (b^*, σ^2)-plane, five hundred functions are simulated from the facilitator's distribution for $f(.)$. These functions are used to discover the values of σ^2 and b^* that lead to simulated functions consistent with the facilitator's prior beliefs about $f(.)$.

The first condition is that $f(.)$ is not negative. For one-hundred equally-spaced values of θ between -3 and 3, the number of times that $f(\theta)$ is negative is recorded, then, for each function, the proportion of points that are negative is calculated. By using this information, an area of the (b^*, σ^2)-plane can be identified as containing the values of b^* and σ^2 that reduce the chance of a function being negative. In practice, we discard drawn functions that are negative at some point; however, if we are discarding a high percentage of functions, computational time can be greatly increased.

As the facilitator believes that $f(.)$ is similar to a t density, the facilitator should believe that $f(.)$ does not have many modes. Some of the lower values of b^* can be ruled out by investigating how many modes the simulated functions have. Modes are searched for in the interval between -3 and 3 because this is where the bulk of the expert's density lies.

The difference between the function drawn from the facilitator's prior distribution for $f(.)$ and the underlying density can be calculated using the simulated functions.

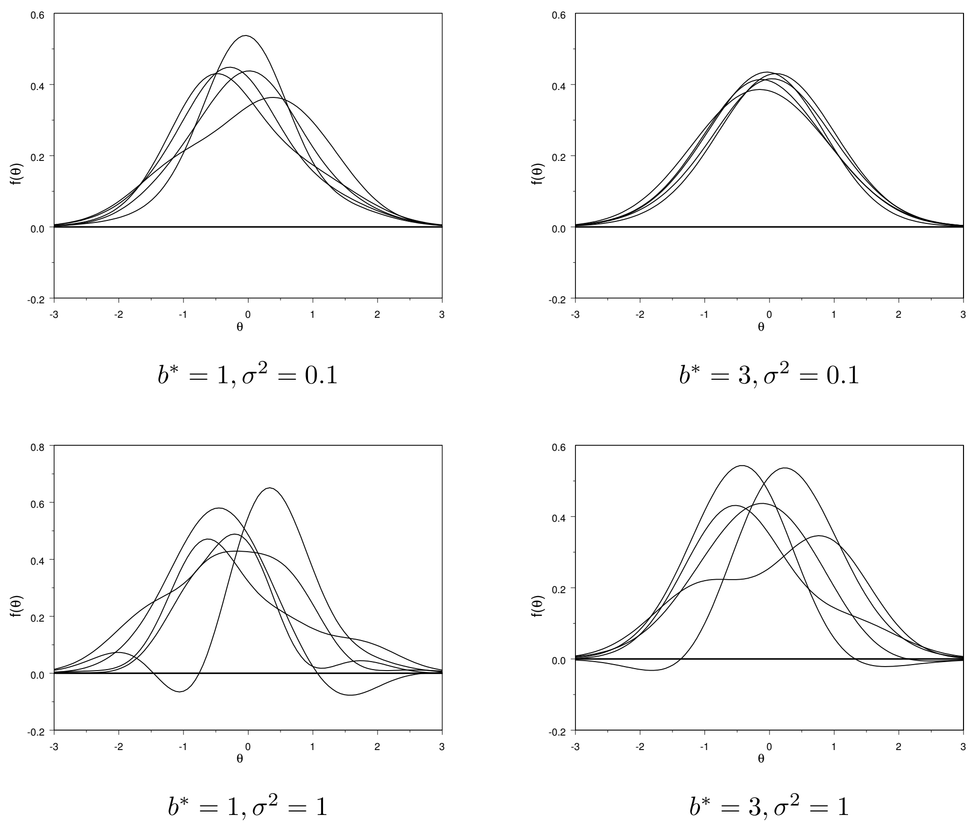

Figure 11: Functions drawn from the facilitator's prior distribution for $f(.)$.

The maximum absolute difference proportional to the underlying density for each of the one-hundred θ values gives a measure of how far $f(.)$ is from $g(.)$. Relatively large values of b^* and small σ^2 lead the simulated functions to follow the underlying density closely. This behaviour should be avoided as it restricts the facilitator's uncertainty about $f(.)$.

The (b^*, σ^2)-plane can be partitioned to give an area that will be likely to yield sensible functions from the facilitator's prior distribution for $f(.)$ based on these investigations. The area labelled by X in Figure 12 is the area that provides values for b^* and σ^2 that is most likely give sensible density functions. The vertical line on the left is positioned at the value of b^* where it is expected that about 20% of functions drawn from the facilitator's prior distribution for $f(.)$ will have three or more modes; as the facilitator believes that $f(.)$ is roughly unimodal, any (b^*, σ^2) pair that falls to left of this line should have low probability in the facilitator's prior distribution. The upper line $u(b^*)$ lies where the proportion of points that are negative is about 0.1, and the proportion is higher for values of σ^2 above the line. The lower line $l(b^*)$ joins points at which the maximum absolute difference proportional to the underlying density between a random function and the underlying distribution is 0.25.

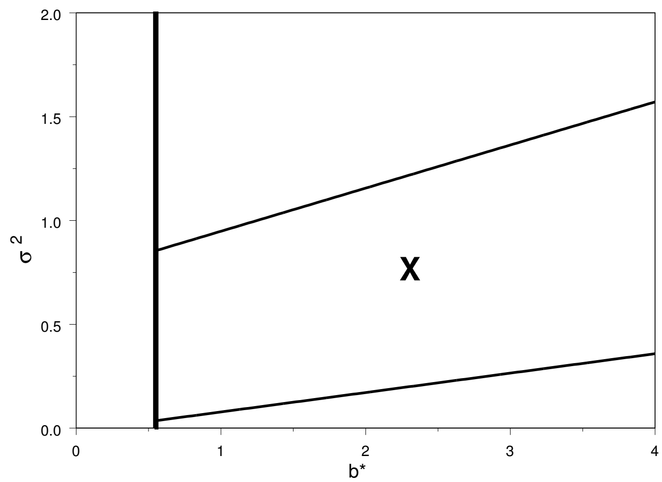

Figure 12: Division of (b^*, σ^2) plane obtained through investigation.

The boundary lines in Figure 12 are given by:

$$\begin{aligned} l(b^*) &= 0.094\, b^* - 0.015, \\ u(b^*) &= 0.208\, b^* + 0.740. \end{aligned} \tag{36}$$

The coefficients are set by a simple linear least-squares fit, and $l(b^*) = 0.0004$ if $b^* < 0.16$. There is still great prior uncertainty as to where inside the boundaries σ^2 and b^* should be selected. A prior distribution for b^* is used to reflect the facilitator's uncertainty:

$$\begin{aligned} \log(b^*) &\sim N\left(\frac{\log(5)+\log(0.55)}{2}, \left(\frac{\log(5)}{2\Phi^{-1}(0.01)}\right)^2\right) \\ &\sim N\left(0.56, 0.27^2\right). \end{aligned} \tag{37}$$

The parameters in this distribution come from setting $P(b^* < 0.55) = 0.01$ and $P(b^* > 5) = 0.01$. The vertical line in Figure 12 is at $b^* = 0.55$, and the value of $b^* = 5$ gives good coverage of the desirable area. For values of $b^* > 5$ numerical problems can arise in the inversion of the correlation matrix.

A conditional prior distribution $p(\sigma^2|b^*)$ can be constructed to utilise the boundary lines given in equations 36 and 36. The following distribution satisfies this condition:

$$\log(\sigma^2)|b^* \sim N(M, S^2), \tag{38}$$

where M and S are to be determined using $l(b^*)$ and $u(b^*)$. A log-normal distribution for $\sigma^2|b^*$ has been chosen here to help prevent σ^2 from getting too close to zero. The facilitator believes that there should be only a small probability of σ^2 falling outside the two curves, $l(b^*)$ and $u(b^*)$, for any b^*; this probability is set at 0.01. The values of M and S are given by

$$\begin{aligned} M &= \frac{\log(u(b^*)) + \log(l(b^*))}{2}, \\ S &= \frac{\log(l(b^*)) - \log(u(b^*))}{2\Phi^{-1}(0.005)}. \end{aligned} \quad (39)$$

The prior distribution, shown in Figure 13, follows the boundary lines suggested in Figure 12.

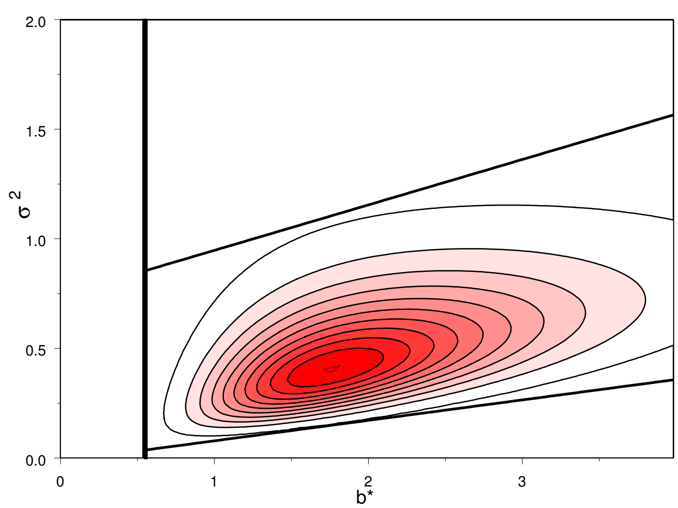

Figure 13: A contour plot of the facilitator's prior density for b^* and σ^2 and the boundary lines of Figure 12.

Given this prior distribution for b^* and σ^2, the prior probability of $f(.)$ being bimodal is approximately 0.09 and the prior probability of $f(.)$ having three of more modes is approximately 0.0005. This matches the facilitator's beliefs that $f(.)$ is likely to be unimodal. Also, the expected proportion of points that $f(.)$ will be negative at is approximately 0.03. Hence, the chance of $f(.)$ being negative is quite small. In addition to this, the expected maximum absolute difference proportional to the underlying density is 0.35, which allows for functions that can be quite different from the underlying density.

References

Ang, M. and Buttery, N. (1997). "An approach to the application of subjective probabilities in level 2 PSAs." *Reliability Engineering and System Safety*, 58: 145–156. 694

Belyaev, L. (1959). "Analytic random processes." *Theor. Probability Appl.*, 4: 402–409. 696

Bernardo, J. and Smith, A. (1994). *Bayesian Theory*. Chichester: Wiley. 693, 711

Blight, B. and Ott, L. (1975). "A Bayesian approach to model inadequacy for polynomial regression." *Biometrika*, 62: 79–88. 695

Cairns, J. and Shackley, P. (1999). "What price information? Modelling threshold probabilities of fetal loss." *Social Science and Medicine*, 49: 823–830. 694

Chaloner, K. and Rhame, F. (2001). "Quantifying and documenting prior beliefs in clinical trials." *Statistics in Medicine*, 20: 581–600. 694

Cooke, R. and Slijkhuis, K. (2003). "Expert Judgment in the Uncertainty Analysis of Dike Ring Failure Frequency." In *Case Studies in Reliability and Maintenance* (eds. W.R. Blischke and D.N. Prabhakar Murthy), 331–352. Chichester: Wiley. 708

Coolen, F., Mertens, P., and Newby, M. (1992). "A Bayes-competing risk model for the use of expert judgement in reliability estimation." *Reliability Engineering and System Safety*, 35: 23–30. 694

Daneshkhah, A., Oakley, J., and O'Hagan, A. (2006). "Nonparametric prior elicitation with imprecisely assessed probabilities." Technical report, *BEEP working report*, Department of Probability and Statistics, University of Sheffield. (Available from `http://www.shef.ac.uk/beep/publications.html`). 711

Dominitz, J. (1998). "Earning expectations, revisions, and realisations." *The Review of Economics and Statistics*, 80: 374–388. 694

Frijters, M., Cooke, R., Slijkhuis, K., and Noortwijk, J. V. (1999). "Expertmening Onzekerheidsanalyse." Technical report, Ministerie van verkeer en Waterstaat, Directoraat-Generaal Rijkswaterstaat, Bouwdienst Rijkswaterstaat ONIN-1-99006. 708

Garthwaite, P., Kadane, J., and O'Hagan, A. (2005). "Statistical methods for eliciting probability distributions." *J. Amer. Statist. Ass.*, 100: 680–701. 694

Grisley, W. and Kellogg, E. (1983). "Farmers' subjective probabilities in Northern Thailand: an elicitation analysis." *American Journal of Agricultural Economics*, 77: 74–82. 694

Gustafson, D., Sainfort, F., Eichler, M., Adams, L., Bisognamo, M., and Steudel, H. (2003). "Developing and testing a model to predict outcomes of organizational change." *Health Services Research*, 38: 751–776. 694

Hogarth, R. (1975). "Cognitive processes and the assessment of subjective probability distributions." *J. Amer. Statist. Ass.*, 70: 271–294. 694

Horrace, W. (2005). "Some results on the multivariate truncated normal distribution." *Journal of Multivariate Analysis*, 94: 209–221. 706

Kadane, J. and Wolfson, L. (1998). "Experiences in elicitation." *The Statistician*, 47: 3–19. 694, 695

Kahneman, D. and Tversky, A. (1973). "On the psychology of prediction." *Psychological review*, 80: 237–251. 694

Kimeldorf, G. and Wahba, G. (1970). "A correspondence between Bayesian estimation on stochastic processes and smoothing by splines." *Ann. Math. Statist.*, 41: 495–502. 695

Must, A., Phillips, S., Stunkard, A., and Naumova, E. (2002). "Expert opinion on body mass index percentiles for figure drawings at menarche." *International Journal of Obesity*, 26: 876–879. 694

Oakley, J. and O'Hagan, A. (2002). "Bayesian inference for the uncertainty distribution of computer model outputs." *Biometrika*, 89: 769–784. 698

— (2007). "Uncertainty in prior elicitations." Technical report, 521/02, Department of Probability and Statistics, University of Sheffield. To appear in *Biometrika*. 695, 711

O'Hagan, A. (1978). "Curve fitting and optimal design for prediction (with discussion)." *J. R. Statist. Soc. Ser. B*, 40: 1–42. 695

— (1991). "Bayes-Hermite quadrature." *J. Statist. Planning and Inference*, 91: 245–260. 698

— (1992). "Some Bayesian numerical analysis." In *Bayesian Statistics 4* (eds. Bernado, J.M. *et al.*), 345–363. Oxford: Oxford University Press. 699

O'Hagan, A., Buck, C., Daneshkhah, A., Eiser, J., Garthwaite, P., Jenkinson, D., Oakley, J., and Rakow, T. (2006). *Uncertain judgements: eliciting expert probabilities*. Chichester: Wiley. 694

O'Hagan, A. and Leonard, T. (1976). "Bayes estimation subject to uncertainty about parameter constraints." *Biometrika*, 63: 201–203. 706

Ramachandran, G., Banerjee, S., and Vincent, J. (2003). "Expert judgement and occupational hygiene: application to aerosol speciation in the nickel primary production industry." *Annals of Occupational Hygiene*, 47: 461–476. 694

Rasmussen, C. (1996). "Evaluation of Gaussian processes and other methods for non-linear regression." Ph.D. thesis, Graduate Department of Computer Science, University of Toronto. 711

Sexsmith, R. (1999). "Probability-based safety analysis — value and drawbacks." *Structural Safety*, 21: 303–310. 694

Smith, J. and Mandac, A. (1995). "Subjective versus objective yield distributions as measures of production risk." *American Journal of Agricultural Economics*, 77: 152–161. 694

Wernisch, L. (2004). "Nonlinear regression with Gaussian processes." *Seminar Slides*, Birbeck College, London. (Available from `http://people.cryst.bbk.ac.uk/~wernisch/`). 711

Acknowledgments

The project was funded by the Engineering and Physical Sciences Research Council. We thank Roger Cooke for providing us with the results from the elicitation exercise that are analysed in this paper. We would also like to thank the anonymous referees for their helpful comments on an earlier draft of this paper.