Biometrika (2007), **94**, 2, *pp*. 427–441
doi:10.1093/biomet/asm031
Advance Access publication 13 May 2007
Printed in Great Britain

Uncertainty in prior elicitations: a nonparametric approach

BY JEREMY E. OAKLEY AND ANTHONY O'HAGAN

Department of Probability and Statistics, University of Sheffield, Sheffield, S3 7RH, U.K.
j.oakley@sheffield.ac.uk a.ohagan@sheffield.ac.uk

SUMMARY

A key task in the elicitation of expert knowledge is to construct a distribution from the finite, and usually small, number of statements that have been elicited from the expert. These statements typically specify some quantiles or moments of the distribution. Such statements are not enough to identify the expert's probability distribution uniquely, and the usual approach is to fit some member of a convenient parametric family. There are two clear deficiencies in this solution. First, the expert's beliefs are forced to fit the parametric family. Secondly, no account is then taken of the many other possible distributions that might have fitted the elicited statements equally well. We present a nonparametric approach which tackles both of these deficiencies. We also consider the issue of the imprecision in the elicited probability judgements.

Some key words: Expert elicitation; Gaussian process; Nonparametric density estimation.

1. INTRODUCTION

1·1. *Expert elicitation*

We consider the elicitation of an expert's beliefs about some unknown continuous variable θ. The objective of the elicitation is to identify the underlying density function $f(\theta)$ that represents the expert's beliefs. For convenience of exposition, we consider an analyst who receives the expert's elicited statements, and suppose that the analyst wishes to make inferences about the expert's density function $f(\theta)$. To avoid convoluted language, we let the expert be female and the analyst be male.

The use of elicited prior distributions will always be contentious in practice, irrespective of any theoretical justification, and even disregarding any considerations of bias. One important concern is the extent to which the density function resulting from the elicitation represents what the expert actually believes. Winkler (1967) argues that the expert does not have a 'true' prior density function waiting to be elicited, only a 'satisficing' prior distribution that she is 'content to live with at a particular moment of time'. If we accept this position, it follows that there can be no unique 'satisficing' prior distribution, and almost certainly different analysts, using different elicitation methods, would elicit different distributions from the same expert.

This issue is generic to all elicitation; with the exception of some trivial cases we doubt whether an analyst could ever claim that a single density function $f(\theta)$ was the sole correct representation of an expert's beliefs. Any subsequent analysis that only considers a single $f(\theta)$ must therefore be ignoring the uncertainty that the analyst still has about the expert's genuine beliefs. We believe that, to improve the credibility and robustness of expert elicitation, this uncertainty needs to be assessed formally. Rather than producing a single $f(\theta)$, the analyst needs to derive a set of all the distinct density functions that he believes to be consistent with the judgements provided by the expert. This in itself is not a novel suggestion; Berger & Berliner (1986) advocate exploring robustness of posterior inferences to the choice of prior. A crucial distinction here is that, by considering prior beliefs about the nature of the expert's density function, such as smoothness, the analyst can probabilistically assess his uncertainty about the expert's beliefs, and so conduct a probabilistic robustness analysis if required.

1·2. *Uncertainty about the expert's distribution*

It is assumed that the expert is only able to provide certain summaries of her distribution such as the mean or various percentiles. A finite set of elicited summaries does not of course identify $f(\theta)$ uniquely, and we identify this as the primary source of uncertainty about the expert's beliefs. A common strategy is to choose a density function that fits those summaries as closely as possible from some convenient parametric family of distributions. Some examples of this are Kadane et al. (1980), O'Hagan (1998), Oakley (2000) and Garthwaite & Al-Awahdi (2002). Two deficiencies in this approach are that it forces the expert's beliefs to fit the parametric family, and that it fails to acknowledge the fact that many other densities might have fitted those same summaries equally well. Practical responses to these criticisms include using a process of feedback to verify that the fitted density is a satisfactory approximation to the expert's beliefs, and checking the sensitivity of subsequent inferences or decisions to variations in the fitted $f(\theta)$.

We present here an approach that addresses both deficiencies together in a single coherent framework. Our fitted estimate of the expert's $f(\theta)$ is nonparametric and we explicitly quantify the uncertainty around this estimate that results from the expert having stated only a finite set of summaries. We treat the problem of fitting a density function to given summaries as an exercise in Bayesian inference; we have an unknown quantity of interest, $f(\theta)$, the analyst formulates his prior beliefs about this quantity and then receives data in the form of summaries of the expert's density function. The analyst then derives his posterior distribution for $f(\theta)$. The analyst'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 second source of uncertainty in an elicitation is the expert's inability to make probability judgements with absolute precision. A complete assessment of the uncertainty about the expert's beliefs should also allow for this. In §6 we suggest a simple extension of our method for acknowledging this imprecision, though we do not aim to give a comprehensive treatment of this issue in this paper.

1·3. *The analyst and the expert*

The field of elicitation is diverse, and addresses a number of related but distinct problems. Lindley et al. (1979), French (1980), Lindley (1982) and Lindley & Singpurwalla (1986) consider the problem of updating the analyst's own beliefs about θ given judgements, possibly incoherent, from one or more experts. It is important to emphasize that our objective is quite different from that of these other authors. In this paper, the analyst is making inference about the expert's density function $f(\theta)$, not about θ. Another way to view the distinction is that in the approach of French, Lindley and Singpurwalla it is the analyst who is the decision-maker, and therefore whose distribution for θ is ultimately required, whereas in our method the decision-maker is implicitly the expert, because it is the expert's distribution for θ that is required. We emphasize that the separation of the roles of the analyst and the expert is primarily for convenience of exposition. In principle, they could be the same person.

There is much more to the process of eliciting expert knowledge than is addressed in this article. Questions should be carefully formulated in terms that the expert understands, and so as to avoid known ways in which experts tend to misjudge probabilities. There are many other important practical questions regarding how the analyst should actually conduct the elicitation, and for advice we refer the reader to Chaloner (1996), Clemen (1996), Garthwaite et al. (2005) and O'Hagan et al. (2006).

2. A prior distribution for the unknown density function

2·1. *The scaled Gaussian process*

A very useful and flexible prior model for an unknown function is the Gaussian process, and we will assume that the analyst's prior beliefs about $f(\theta)$ can be represented by a Gaussian process. In particular, the analyst's prior distribution for any finite set of points on this function is multivariate normal. Gaussian process priors for functions have been proposed in various different settings, including regression (O'Hagan, 1978; Neal, 1999), classification (Neal, 1999) and numerical analysis (O'Hagan, 1992). A recent treatment of both Gaussian process regression and classification is given in Rasmussen & Williams (2006).

Since $f(\theta)$ is a density function, we need to apply the constraint $f(\theta) \geqslant 0$ for all θ. This is accomplished by first updating the analyst's prior given the elicited data D, ignoring the constraint, and then conditioning on $f(\theta) \geqslant 0$ for all θ after we have derived $p\{f(\theta)|D\}$. It is also known that $\int_{-\infty}^{\infty} f(\theta)d\theta = 1$ and we apply this constraint as part of the data in §3.

The Gaussian process is specified by giving its mean function and variance-covariance function. We model these hierarchically in terms of a vector α of hyperparameters. First let the analyst's prior expectation of $f(\theta)$ be some member $g(\theta \mid u)$ of a suitable parametric family with parameters u, contained within α. Thus

$$E\{f(\theta) \mid \alpha\} = g(\theta \mid u).$$

It would not be realistic to suppose that the variance of $f(\theta)$ would be the same for all θ. In general, where the analyst expects $f(\theta)$ to be smaller his prior variance should be smaller in absolute terms. We reflect this in our model by supposing that the variance-covariance function has the scaled stationary form

$$\text{cov}\{f(\theta), f(\phi) \mid \alpha\} = g(\theta \mid u)\, g(\phi \mid u)\, \sigma^2 c(\theta, \phi),$$

where $c(\theta, \phi)$ is a correlation function that takes the value 1 at $\theta = \phi$ and is a decreasing function of $|\theta - \phi|$. In general, the function $c(\cdot, \cdot)$ must ensure that the prior variance-covariance matrix of any set of observations of $f(\cdot)$, or functionals of $f(\cdot)$, is positive semidefinite. Here we choose the function

$$c(\theta, \phi) = \exp\{-\frac{1}{2b}(\theta - \phi)^2\}.$$

This will be seen to be a mathematically convenient choice, and implies that $f(\cdot)$ is infinitely differentiable with probability 1. This formulation was given in Kennedy & O'Hagan (1996), who were interested in quadrature for computationally expensive density functions.

Although the analyst's prior expectation is that $f(\theta)$ will be a member of the parametric family $g(\theta \mid u)$, the model is nonparametric and allows the true $f(\theta)$ to have any form at all. The hyperparameter σ^2 specifies how close the true density function will be to the analyst's parametric family. The hyperparameter b controls the smoothness of the true density. If b is large, then two points $f(\theta)$ and $f(\phi)$ will be highly correlated even if θ and ϕ are far apart.

2·2. *Hierarchical prior distribution*

The hyperparameters of this model are $\alpha = (u, \sigma^2, b)$. In general, we would wish to express noninformative prior distributions for the components of u. Although the analyst may in practice have some idea himself about what values to expect for θ, and so may have prior knowledge about u, our objective is to elicit the expert's beliefs about θ, and we would not wish this to be coloured by the analyst's beliefs.

We would also wish to express noninformative prior beliefs about σ^2, and so allow the analyst to learn just how well the expert's density function approximates the underlying parametric family. However, there are good reasons for wishing to specify some informative beliefs about the smoothness parameter b. The first reason is technical: an improper prior for b can lead to an improper posterior regardless of the data observed. The data will not be able to distinguish between two sufficiently small but distinct values of b, as the differences in smoothness would only be detectable on a very fine scale. Furthermore, the analyst will not expect the true density to be highly multimodal. If u contains a variance parameter v, then it is reasonable to suppose that b is of the same order of magnitude as v, since this then expresses a belief in moderate smoothness of the true density over the range for which $f(\theta)$ is non-negligible. This implies prior beliefs about the ratio b/v.

In the following sections we will consider the particular case where $u = (m, v)$ and $g(\cdot \mid u)$ is the normal density with mean m and variance v. We reparameterize the covariance function as follows:

$$c(\theta, \phi) = \exp\{-\frac{1}{2vb^*}(\theta - \phi)^2\},$$

where $b^* = b/v$. We now let the prior distribution of $\alpha = (m, v, \sigma^2, b^*)$ have the form

$$p(m, v, \sigma^2, b^*) \propto v^{-1}\sigma^{-2}p(b^*),$$

where $p(b^*)$ will be an informative prior for b^* expressing the above beliefs. We represent our uncertainty about $\log b^*$ with a standard normal distribution. Informal justification for this prior is given in Fig. **??**, where we show realizations from the prior distribution of the ratio $f(\theta)/g(\theta)$. Since we believe $f(\cdot)$ to be a smooth function, we would not expect this ratio to fluctuate excessively. In addition, we believe that $f(\cdot)$ may deviate substantially from a normal density function, and so we would not assign much prior weight to values of b^* that imply an almost constant ratio. The plots in Fig. **??** suggest that $(-2, 2)$ is an appropriate range to consider for $\log b^*$ a priori. We have chosen a normal distribution to give more weight to the centre of the range, as $\log b^* = \pm 2$ appears to give realizations that are too rough/smooth. We note that it will be important to investigate robustness to this choice of prior.

2·3. *Alternative models*

We have described a model for the unknown density function in which $f(\cdot)$ is allowed to deviate nonparametrically from a parametric class of density functions. This is a common

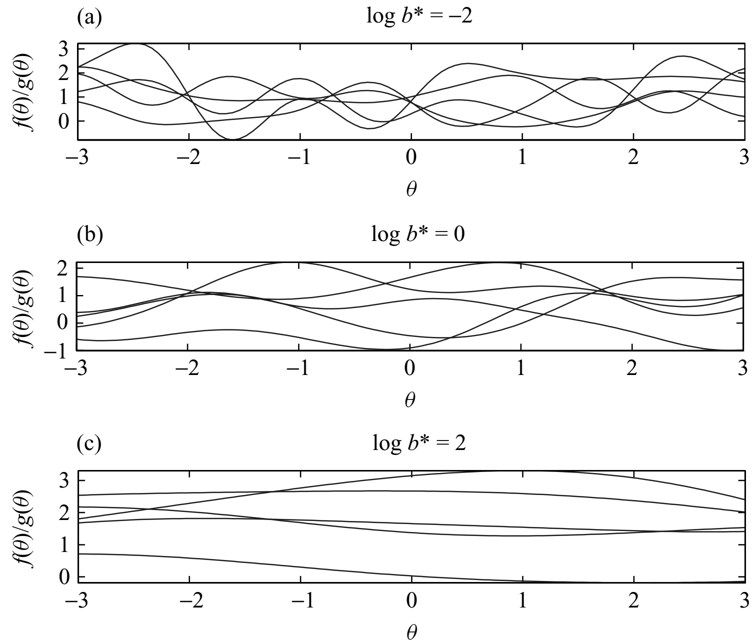

Fig. 1. Realizations from the prior distribution of $f(\theta)/g(\theta)$ for (a) $\log b^* = -2$, (b) $\log b^* = 0$, (c) $\log b^* = 2$.

theme in Bayesian nonparametric inference; see for example Leonard (1978), who modelled the log of the unknown density function as a Gaussian process, Hjort (1996) and Walker et al. (1999). In general these alternative modelling approaches guarantee a positive density function, but they were developed for the problem of density estimation, where inference is required for an unknown distribution based on a sample of observations from that distribution. The elicitation context is quite different, and our data comprise elicited summaries of the distributions, such as quantiles. These alternative models would become completely intractable for such data. For instance, if $\log f(\theta)$ has a Gaussian process distribution then we cannot even write down the likelihood for data which comprise integrals of $f(\theta)$.

West (1988) presents a prior distribution for an expert's quantile function $Q(U)$ for $U \in [0, 1]$. His formulation leads to a Dirichlet model for the expert's distribution. As in Lindley et al. (1979), this is in the context where the analyst is the decision-maker and so it is the analyst's beliefs about θ rather than $f(\theta)$ that are of interest. Note that a possibility here would have been to model the quantile function as a Gaussian process; this would solve one problem in that $Q(U)$ can be negative, but would create a new one in that $Q(U)$ must be nondecreasing.

3. Prior to posterior updating

Our model for the analyst's prior distribution is designed to handle the kinds of information commonly elicited from an expert. This includes quantiles of the distribution and simple moments. Since $f(\theta)$ has a Gaussian process distribution conditional on the hyperparameters α, the distribution of any linear functional of $f(\theta)$, which includes moments and other expectations such as the probabilities that define given quantiles, is normal.

Formally, let the data comprise a vector D of elicited summaries of this form. Remember that D will also include the information that $\int_{-\infty}^{\infty} f(\theta)d\theta = 1$. Then the joint distribution of D and any finite set of points on the function $f(\theta)$ is multivariate normal. Denote the mean vector and variance-covariance matrix of D by H and $\sigma^2 A$ respectively, and denote the covariance between D and $f(\theta)$ by $\sigma^2 t(\theta)$. Note that H will be a function of m and v, while A and $t(\theta)$ will be functions of m, v and b^*. The forms of these quantities for some common summaries are given explicitly in the Appendix. It follows immediately from properties of the multivariate normal distribution that $f(\theta)|D, m, v, b, \sigma^2$ also has a normal distribution with

$$E\{f(\theta) \mid \alpha\} = g(\theta) + t(\theta)^{\mathrm{T}} A^{-1}(D - H),$$
$$\mathrm{cov}\{f(\theta), f(\phi) \mid \alpha\} = \sigma^2\{g(\theta \mid u)g(\phi \mid u)c(\theta, \phi) - t(\theta)^{\mathrm{T}} A^{-1} t(\phi)\},$$

In fact, conditional on α the analyst's posterior distribution of $f(\theta)$ is again that of a Gaussian process, with these as its mean and variance-covariance functions.

The posterior distribution of $\alpha = (m, v, \sigma^2, b^*)$ is easily found from the multivariate normal likelihood for D given α:

$$p(m, v, \sigma^2, b^* \mid D) \propto v^{-1}|A|^{-\frac{1}{2}}\sigma^{-(n+2)}\frac{1}{b^*}\exp\left\{-\frac{1}{2}(\log b^*)^2\right\} \times \exp\left\{-\frac{1}{2\sigma^2}(D - H)^T A^{-1}(D - H)\right\}, \quad (1)$$

where n denotes the number of elements in D. The posterior marginal distribution of $f(\theta)$ is then the result of integrating its Gaussian process conditional distribution with respect to this marginal posterior of the hyperparameters. We can integrate σ^2 out of (1) analytically to obtain

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

where

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

We can then use Markov chain Monte Carlo to obtain a sample of values of $\{m, v, b^*\}$ from their joint posterior. Finally, given m, v and b^*, we can sample a density function $f(\cdot)$ at a finite number of values of θ from the posterior $p\{f(\cdot) \mid D, m, v, b^*\}$, as in Oakley & O'Hagan (2002) the joint posterior distribution of $\{f(\theta_1), \ldots, f(\theta_n)\}$ conditional on D, m, v and b^* is multivariate t, for any set $\{\theta_1, \ldots, \theta_n\}$. At this point we apply the constraint $f(\theta) \geqslant 0$ for all θ by discarding any density functions that are negative over a suitable finite range. Repeating this will give us a sample of functions from the analyst's posterior distribution of $f(\cdot)$, and estimates and pointwise bounds for $f(\theta)$ can then be reported.

4. EXAMPLE

4·1. *A bimodal test case*

The aim of this synthetic example is to demonstrate how the analyst is able to identify a density function that is of a markedly different shape from his prior expectation of a normal density, given sufficient judgements from the expert. Suppose that the expert has the following density function for θ:

$$f(\theta) = \frac{0{\cdot}4}{\surd(2\pi)} \exp\{-\frac{1}{2}(\theta+2)^2\} + \frac{0{\cdot}6}{\surd(4\pi)} \exp\{-\frac{1}{4}(\theta-1)^2\}.$$

The analyst asks the expert to state $\text{pr}(\theta < x)$ for $x \in \{-3, -2, -1, 0, 1, 2, 3\}$.

We now use Markov chain Monte Carlo to sample from the posterior distribution of $\{m, v, b^*\}$. The chain is run for 20 000 iterations, and the first 1000 runs are discarded. For each sampled set of $\{m, v, b^*\}$, we generate one random density function. Functions that are negative anywhere for $\theta \in (-6, 6)$ are rejected. We plot the pointwise median, 2·5th and 97·5th percentiles from the distribution of the density function in Fig. **??** (a). Note that the median is almost indistinguishable from the true density function.

4·2. *Feedback*

The analyst can show the expert his posterior distribution of her density function, although, with the exception of features such as number of modes and skewness, she may not to be able to assess whether or not the analyst has a good representation of her beliefs. However, if the analyst presents his posterior distribution of her distribution function, she can then see if values of $\text{pr}(\theta < \theta_0)$ for new values of θ_0 suggested by his posterior match her own beliefs. We plot a pointwise 95% interval for the distribution function in Fig. **??**(b).

4·3. *Robustness*

We now repeat the analysis with weaker prior information about the parameter b^*: we double the prior standard deviation so that $\log b^* \sim N(0, 4)$. In Fig. **??**(a) we show pointwise posterior intervals for $f(\theta)$ resulting from both the $N(0, 1)$ and $N(0, 4)$ prior distributions. We can see that posterior uncertainty about $f(\theta)$ is sensitive to the choice of prior; posterior uncertainty about the cumulative distribution function is less sensitive, but we do not show the plot here.

Increasing prior uncertainty about b^* has led to more realizations from the posterior distribution of $f(\theta)$ with three or more modes; these account for approximately 20% of the realizations compared with 5% using the $N(0, 1)$ prior. In practice, information about

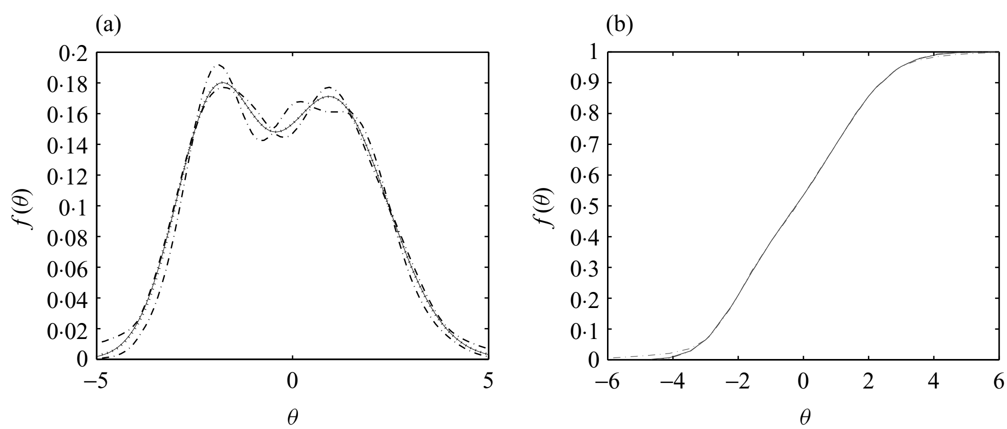

Fig. 2. Synthetic example. (a) The median (dotted), pointwise 95% intervals for the expert's density function (dot-dashed), and the true density function (solid). (b) Pointwise 95% intervals for the expert's distribution function (dot-dashed), and the true distribution function (solid).

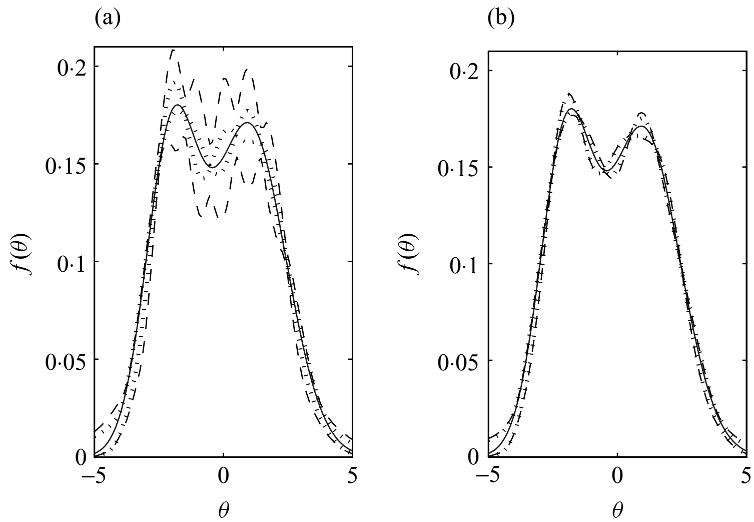

Fig. 3. Synthetic example. (a) Pointwise 95% intervals for the expert's density function using $\log b^* \sim N(0, 1)$ (dotted) and $\log b^* \sim N(0, 4)$ (dashed), and the true density function (solid). (b) as in panel (a), but with all density functions with three or more modes excluded.

the number of modes in the expert's distribution could be obtained through discussion between the analyst and the expert. If we now choose to reject any realization that has three or more modes, we find that the two prior distributions for b^* give similar results; see Fig. ??(b). Consequently, given additional information about the number of modes in the expert's distribution, it could be argued that in this example the only important difference between the two priors is that the $N(0, 1)$ prior will lead to more efficient computation, as fewer realizations will need to be rejected.

5. Uncertainty about posterior beliefs

Whether or not the uncertainty about the expert's distribution matters in practice will depend upon the use to which it will be put. One obvious use for an elicited distribution is to serve as a prior distribution $f(\theta)$ for a Bayesian analysis of some data s. It is then of interest to derive the uncertainty that is implied for the resulting posterior density $f(\theta|s)$. We consider inference about a posterior probability $\text{pr}(\theta \in A|s)$. We can simulate random probabilities from the analyst's distribution of $\text{pr}(\theta \in A|s)$, by first simulating density functions $f(\cdot)$ from the analyst's posterior distribution of $f(\cdot)|D$, and then by calculating $\text{pr}(\theta \in A|s)$ using numerical integration.

Berger & O'Hagan (1988) give an example of two engineers stating prior probabilities for the mean lifetime θ of a proposed new industrial engine; for conciseness we only consider one engineer here. The engineer states $\text{pr}(\theta \in I_i)$ for a set of intervals $I_1, \ldots, I_6$, and these probabilities are listed in Table 1. In addition to these prior beliefs there are data s consisting of two observations with likelihood function $f(s|\theta) = \theta^{-2} \exp(-4500/\theta)$. In Berger & O'Hagan (1988), the objective was to compute maximum and minimum values for posterior probabilities for various classes of prior distributions, rather than to quantify probabilistically the uncertainty in the prior/posterior distribution. Consequently, we can only compare the two approaches informally.

Table 1. *Engineers example. Prior probabilities* (p_i)*, minimum and maximum posterior probabilities under two classes of prior distribution* Π_0 *and* Π_2'*, and an approximate* 99·9% *interval for the posterior probability using the Gaussian process model* (GP).

I_i	p_i	Π_0	Π_2'	GP
[0, 1000)	0·15	(0, 0·111)	(0·020, 0·023)	(0·017, 0·031)
[1000, 2000)	0·15	(0·088, 0·255)	(0·162, 0·184)	(0·149, 0·178)
[2000, 3000)	0·20	(0·235, 0·391)	(0·284, 0·327)	(0·286, 0·301)
[3000, 4000)	0·20	(0·197, 0·349)	(0·248, 0·288)	(0·247, 0·260)
[4000, 5000)	0·15	(0·125, 0·233)	(0·149, 0·175)	(0·149, 0·159)
[5000, ∞)	0·15	(0, 0·146)	(0, 0·121)	(0·077, 0·123)

We now simulate values from the distribution of $\text{pr}(\theta \in I_i | s)$ for the engineer. Approximate 99·9% intervals are reported in Table 1 for each interval. For comparison, ranges of posterior probabilities are given for two prior families of distributions considered in Berger & O'Hagan (1988): the family Π_0 contains any prior distribution that has the correct elicited prior probabilities given by the expert, and Π_2' contains any prior distribution that has the correct elicited prior probabilities, is unimodal and has a particular upper bound for the maximum value of the density.

Note the discrepancy between the Gaussian process model and the Π_2' class regarding interval I_2. This occurs because the probabilities specified by the engineer are incompatible with a distribution that is both smooth and unimodal; a smooth distribution would have to have a turning point in the interval (0, 2000).

6. Uncertainty in the elicited probabilities

A difficulty in any elicitation is that the expert will not be able to specify any probability with absolute precision. This issue was considered in Walley (1991), who proposed bounding a probability P with upper and lower probabilities, P_{U} and P_{L} respectively. This still leaves unresolved the issue of how to specify P_{U} and P_{L} with absolute precision. In addition, the expert may also feel that values in the centre of the interval $[P_{\text{L}}, P_{\text{U}}]$ represent their uncertainty more appropriately than values towards the ends of the interval. This is an important yet under-researched theme in elicitation. In this paper, we suggest a simple pragmatic approach for acknowledging this imprecision.

We interpret the reported probability $\text{pr}^*(\theta \in A)$ as the expert's 'true' probability plus a small additive 'error' which represents the imprecision in the stated probability:

$$\text{pr}^*(\theta \in A) = \text{pr}(\theta \in A) + \varepsilon.$$

Lindley et al. (1979) and Dickey (1980) both consider models for errors in probability judgements, but in the context of incoherence rather than imprecision. We then assume that $\varepsilon \sim N(0, \tau^2)$ for some appropriate value of τ^2. The variance parameter τ^2 describes the imprecision in the probability assessment, and may vary for different probability judgements. The variances could be chosen in consultation with the expert.

With noise in the data, it is now necessary to use a proper prior distribution for σ^2. By considering the posterior distribution of σ^2 in noise-free examples, we have found that $\sigma^{-2} \sim \text{Ga}(1, 1)$ is a suitable candidate for $f(\sigma^2)$. As with the proper prior for b^*, robustness

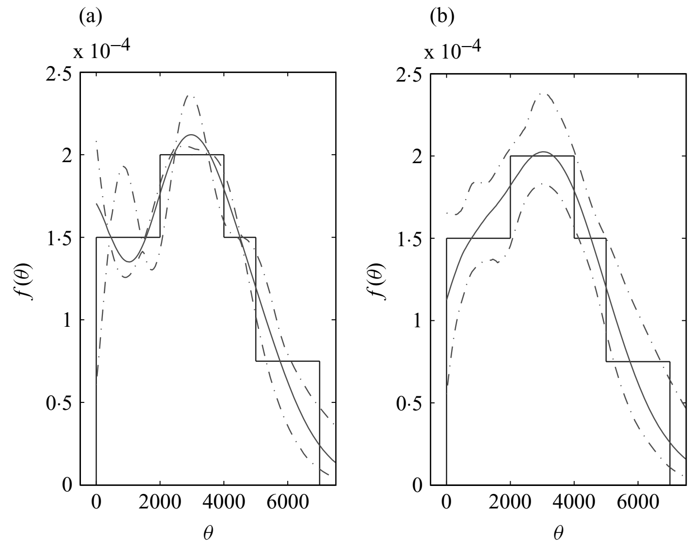

Fig. 4. Engineers example. Curves show 2·5th, 50th and 97·5th percentiles of $f(\theta)$ with (a) no imprecision and (b) imprecision in the elicited probabilities. A histogram, truncated at 7000, based on the elicited probabilities is also given for comparison.

of posterior inferences to this choice of prior should be checked, though we have found that alternatives such as $\sigma^{-2} \sim \text{Ga}(2, 1)$ and $\sigma^{-2} \sim \text{Ga}(1, 0{\cdot}5)$ give similar results.

We return to the example of the elicited probabilities from the engineer. Recall that these were incompatible with a smooth, unimodal distribution, and we now consider the effect of allowing for imprecision in the elicited probabilities. We suppose that the six prior probabilities in Table 1 were imprecise probabilities $p_1^*, \ldots, p_6^*$, with

$$p_i^* = p_i + \varepsilon_i.$$

We need to choose a variance for each error ε_i. Since the absolute imprecision must be smaller for probabilities close to 0 or 1, we set $\varepsilon_i \sim N[0, \{0{\cdot}1 \times (1 - p_i^*)p_i^*\}^2]$. The effect of including this 'noise' in the data is shown in Fig. **??**. In Fig. **??**(a), we show the analyst's pointwise 2·5th, 50th and 97·5th percentiles of $f(\theta)$ with no imprecision in the elicited probabilities. In Fig. **??**(b), the imprecision has been added, and the analyst now expects the engineer to have a smooth, unimodal distribution.

7. Example: the bisection method

One significant concern in elicitation is the ability of experts to assess appropriately the tails of their distributions. Alpert & Raiffa (1982) report an experiment in which subjects were asked to provide 98% intervals for various uncertain quantities known only to the experimenter. It was found that these intervals contained the true values only 53% of the time. In an attempt to reduce overconfidence, various elicitation schemes have been suggested in which the expert is not asked for extreme percentiles. One such scheme is the bisection method, described in Raiffa (1968), in which the expert is asked for her median and quartiles. Murphy & Winkler (1974) report good results with this method, though Garthwaite & O'Hagan (2000) suggest that overconfidence is reduced more effectively if the expert is asked to provide three equally likely intervals.

Here we show that, given only the median and quartiles, posterior uncertainty about $f(\theta)$ can be considerable, particularly if the expert's distribution is skewed. We consider a parameter θ that is known to be positive, but is likely to be close to zero. It is straightforward to handle this scenario by transforming θ to the log scale, and then model the density function of $\log\theta$ as a Gaussian process, with the expectation of $f(\log\theta)$ given by a normal density function. The same prior as described in §2·2 is used for b^*. We can then transform back to the original scale when reporting the posterior distribution of $f(\theta)$.

The initial elicited probabilities are $\text{pr}(\theta < 1) = 0{\cdot}25$, $\text{pr}(\theta < 1{\cdot}7) = 0{\cdot}5$ and $\text{pr}(\theta < 2{\cdot}7) = 0{\cdot}75$. We then compare the analyst's posterior distribution for $f(\theta)$ with three parametric distributions, namely $\theta \sim \text{Ga}(2{\cdot}127, 1{\cdot}058)$, $\log\theta \sim N(0{\cdot}512, 0{\cdot}737^2)$ and $\theta^{-1} \sim \text{Ga}(2{\cdot}066, 2{\cdot}844)$. All three parametric distributions give good approximations to the elicited probabilities, although, as they do not match the elicited probabilities exactly, we again allow for imprecision in the Gaussian process model, with $\varepsilon_i \sim N[0, \{0{\cdot}1 \times (1 - p_i^*)p_i^*\}^2]$ as before.

In Fig. **??**(a), we plot the analyst's pointwise 99% intervals for $f(\theta)$ together with these three parametric estimates. For this comparison we have chosen to reject any bimodal realization from the analyst's posterior distribution for $f(\theta)$. Note that there is considerably more uncertainty about $f(\theta)$ given the three percentiles only. Of particular interest here is the right-hand tail. In Fig. **??**(b) we can see that the analyst's posterior uncertainty encompasses the different tail behaviours exhibited by the three parametric approximations.

We can also consider the analyst's uncertainty about the expert's mean of θ. For a skewed distribution, it is very unlikely that the expert would be able to specify her mean value directly, and so it could only be deduced by considering her complete density function. The three parametric fits give mean values of 2·010, 2·189 and 2·668 respectively. The

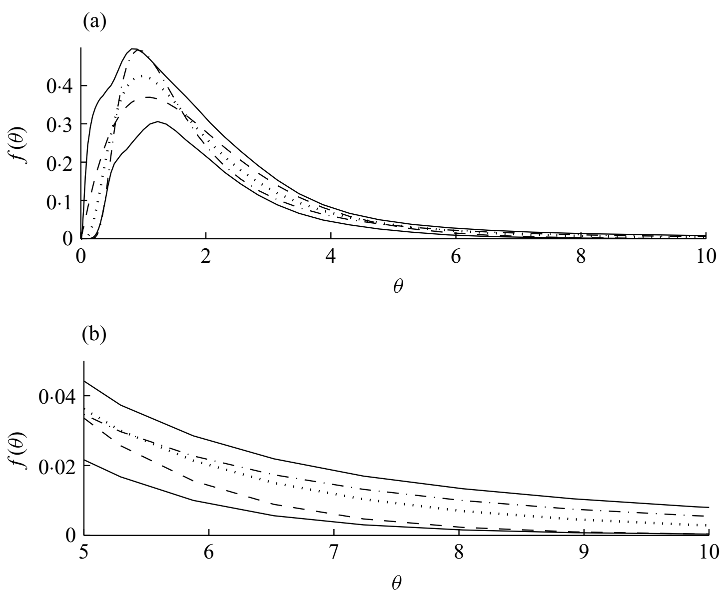

Fig. 5. Bisection method example. Analyst's 99% pointwise interval for $f(\theta)$ (solid), fitted gamma distribution (dashed), fitted log-normal distribution (dotted) and fitted inverse gamma distribution (dot-dashed), (a) for the whole range, (b) for the upper tail.

analyst's 99% interval for the expert's mean is (1·91, 2·83); changing the prior distribution for b^* to $\log b^* \sim N(0, 4)$ has negligible effect, although again the proportion of rejected realizations is higher. The uncertainty in the mean value is not trivial and could, for example, be important if the expert needed to identify $E(\theta)$ for use in a decision problem. To reduce this uncertainty, the expert would have to make a judgement about the tail of her distribution, such as its 95th percentile. Alternatively, if the expert were unwilling to make a tail judgement, she could use the analyst's posterior expectation of $E(\theta)$ as her mean value, if she were willing to accept the analyst's prior assumptions about $f(\theta)$.

8. DISCUSSION

We have presented a means of eliciting an expert's prior density function $f(\cdot)$ that both avoids having to assume a parametric form for $f(\cdot)$ and allows us to measure our uncertainty about $f(\cdot)$ given a limited number of judgements from the individual. A number of potential practical advantages of this approach offer directions for its further development. In practice, elicitation is a dialogue between the expert and the analyst. For instance, after eliciting a set of summaries the analyst can consider his posterior uncertainty about the expert's density function before offering a final inference. For example, he may choose to elicit more summaries if his posterior uncertainty is too large. Our model may then help to identify which additional summaries would be most informative. The dialogue can also be used to check the validity of the analyst's prior model for $f(\theta)$.

One obvious area for further development is the multivariate case. In some cases the existing method can be applied simply by considering alternative parameterizations where independence may be assumed, such as a baseline and relative risk instead of two absolute risks. In principle, the method extends readily to multivariate elicitation, but some practical issues arise of computation and of deciding what summaries to elicit from the expert. This is an area of active research.

ACKNOWLEDGEMENT

We would like to thank an associate editor and two referees for their very helpful comments regarding the content and clarity of this paper.

APPENDIX

Likelihood for the expert's summaries

In §3 it is stated that the data D comprising the expert's summaries are normally distributed. We give here details of the means, variances and covariances for summaries of particular forms when $g(\theta \mid u)$ is the $N(m, v)$ density.

Expectation of θ. Let $\mu = E(\theta)$. Then μ is normally distributed with

$$E_f(\mu|m, v, b^*, \sigma^2) = m,$$

$$\text{cov}_f\{f(\theta), \mu|m, v, b^*, \sigma^2\} = \sigma^2 g(\theta) \int_{-\infty}^{\infty} \phi c(\theta, \phi) g(\phi) d\phi$$

$$= \sigma^2 g(\theta) \exp\left\{-\frac{(\theta - m)^2}{2v(b^* + 1)}\right\} \frac{(\theta + mb^*)\surd b^*}{(1 + b^*)^{3/2}}$$

$$\mathrm{var}_f(\mu|m, v, b^*, \sigma^2) = \sigma^2 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \theta\phi c(\theta, \phi) g(\theta) g(\phi) d\theta d\phi$$

$$= \sigma^2 \frac{\surd b^*(2m^2 + b^* m^2 + v)}{(2 + b^*)^{3/2}}.$$

Percentiles of θ. A value x is chosen, and the expert gives her probability that $\theta \leqslant x$. We write $P_x = \int_{-\infty}^{x} f(\theta) d\theta$. This also has a normal distribution, with

$$E(P_x|m, v, b^*) = \Phi\left(\frac{x - m}{v}\right),$$

$$\mathrm{cov}\{P_x, f(\theta)|m, v, b^*, \sigma^2\} = \sigma^2 g(\theta) \int_{-\infty}^{x} c(\theta, \phi) g(\phi) d\phi$$

$$= \sigma^2 g(\theta) \left(\frac{b^*}{1 + b^*}\right)^{1/2} \exp\left\{-\frac{(\theta - m)^2}{2v(b^* + 1)}\right\} \Phi\left\{\left(x - \frac{\theta + mb^*}{b^* + 1}\right) \surd\left(\frac{1 + b^*}{vb^*}\right)\right\},$$

where $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal distribution. The covariance between any two percentiles is given by

$$\mathrm{cov}(P_x, P_y|m, v, b^*, \sigma^2) = \sigma^2 \int_{-\infty}^{x}\int_{-\infty}^{y} g(\theta) g(\phi) c(\theta, \phi) d\theta d\phi$$

$$= \frac{\sigma^2}{2\pi v} \int_{-\infty}^{x}\int_{-\infty}^{y} \exp\left\{-\frac{1}{2vb^*}(\theta - \phi)^2 - \frac{1}{2v}(\theta - m)^2 - \frac{1}{2v}(\phi - m)^2\right\} d\theta d\phi$$

$$= \frac{|S|^{1/2}\sigma^2}{2\pi v} \int_{-\infty}^{x}\int_{-\infty}^{y} |S|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\theta - m, \phi - m) S^{-1} (\theta - m, \phi - m)^{\mathrm{T}}\right\} d\theta d\phi,$$

where

$$S^{-1} = \begin{pmatrix} \frac{1 + b^*}{vb^*} & -\frac{1}{vb^*} \\ -\frac{1}{vb^*} & \frac{1 + b^*}{vb^*} \end{pmatrix}.$$

Making the transformation

$$z_1 = \surd\left(\frac{b^* + 2}{v(b^* + 1)}\right)(\theta - m),$$

$$z_2 = \surd\left(\frac{b^* + 1}{b^* v}\right)(\phi - m) - \surd\left(\frac{1}{b^* v(b^* + 1)}\right)(\theta - m),$$

we have

$$\mathrm{cov}(P_x, P_y|m, v, b^*, \sigma^2) = \frac{\sigma^2}{2\pi}\surd\left(\frac{b}{b + 2v}\right) \int_{-\infty}^{l_1}\int_{-\infty}^{l_2} \exp\left\{-\frac{1}{2}(z_1^2 + z_2^2)\right\} dz_1 dz_2$$

$$= \frac{\sigma^2}{\surd(2\pi)}\surd\left(\frac{b}{b + 2v}\right) \int_{-\infty}^{l_1} \exp\left(-\frac{1}{2}z_1^2\right) \Phi(l_2) dz_1,$$

where

$$l_1 = \surd\left(\frac{b^* + 2}{v(b^* + 1)}\right)(x - m),$$

$$l_2 = \surd\left(\frac{b^* + 1}{b^* v}\right)(y - m) - \frac{1}{\surd\{b^*(b^* + 2)\}} z_1.$$

We also need the covariance between percentiles and the mean. This is given by

$$\operatorname{cov}(P_y, \mu | m, v, b^*, \sigma^2)$$
$$= \frac{\sigma^2}{2\pi}\sqrt{\left(\frac{b^*}{b^*+2}\right)}\int_{-\infty}^{\infty}\int_{-\infty}^{l_2}\left\{\sqrt{\left(\frac{v(b^*+1)}{b^*+2}\right)}z_1 + m\right\}\exp\left\{-\frac{1}{2}(z_1^2+z_2^2)\right\}dz_1dz_2.$$

REFERENCES

ALPERT, M. & RAIFFA, H. (1982). A progress report on the training of probability assessors. In *Judgement and Uncertainty: Heuristics and Biases*, Ed. D. Kahneman, P. Slovic and A. Tversky, pp. 294–305. Cambridge: Cambridge University Press.

BERGER, J. O. & BERLINER, L. M. (1986). Robust Bayes and empirical Bayes analysis with ϵ-contaminated priors. *Ann. Statist.* **14**, 461–86.

BERGER, J. O. & O'HAGAN, A. (1988). Ranges of posterior probabilities for unimodal priors with specified quantiles (with Discussion). In *Bayesian Statistics 3*, Ed. J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, pp. 45–65. Oxford: Oxford University Press.

CHALONER, K. (1996). The elicitation of prior distributions. In *Case Studies in Bayesian Biostatistics*, Ed. D. Berry and D. Stangl, pp. 141–56. New York: Marcel Dekker.

CLEMEN, R. T. (1996). *Making Hard Decisions*. Belmont, CA: Duxberry.

DICKEY, J. M. (1980). Beliefs about beliefs, a theory for stochastic assessements of subjective probabilities (with Discussion). In *Bayesian Statistics*, Ed. J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, pp. 471–87. Valencia: Valencia University Press.

FRENCH, S. (1980). Updating of belief in the light of someone else's opinion. *J. R. Statist. Soc.* A **143**, 43–8.

GARTHWAITE, P. H. & AL-AWAHDI, S. A. (2002). Non-conjugate prior distribution assessment for multivariate normal sampling. *J. R. Statist. Soc.* B **50**, 462–74.

GARTHWAITE, P. H., KADANE, J. B. & O'HAGAN, A. (2005). Elicitation. *J. Am. Statist. Assoc.* **100**, 680–701.

GARTHWAITE, P. H. & O'HAGAN, A. (2000). Quantifying expert opinion in the UK water industry: an experimental study. *Statistician* **49**, 455–77.

HJORT, N. L. (1996). Bayesian approaches to non- and semiparametric density estimation. In *Bayesian Statistics 5*, Ed. J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, pp. 223–54. Oxford: Oxford University Press.

KADANE, J. B., DICKEY, J. M., WINKLER, R. L., SMITH, W. S. & PETERS, S. C. (1980). Interactive elicitation of opinion for a normal linear model. *J. Am. Statist. Assoc.* **75**, 845–54.

KENNEDY, M. C. & O'HAGAN, A. (1996). Iterative rescaling for Bayesian quadrature. In *Bayesian Statistics 5*, Ed. J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, pp. 639–45. Oxford: Oxford University Press.

LEONARD, T. (1978). Density estimation, stochastic processes and prior information (with Discussion). *J. R. Statist. Soc.* B **40**, 113–46.

LINDLEY, D. V. (1982). The improvement of probability judgments. *J. R. Statist. Soc.* A **148**, 117–26.

LINDLEY, D. V. & SINGPURWALLA, N. D. (1986). Reliability (and fault tree) analysis using expert opinions. *J. Am. Statist. Assoc.* **81**, 87–90.

LINDLEY, D. V., TVERSKY, A. & BROWN, R. V. (1979). On the reconciliation of probability assessments (with Discussion). *J. R. Statist. Soc.* A **142**, 146–80.

MURPHY, A. H. & WINKLER, R. L. (1974). Credible interval temperature forecasting: some experimental results. *Mon. Weath. Rev.* **102**, 784–94.

NEAL, R. (1999). Regression and classification using Gaussian process priors. In *Bayesian Statistics 5*, Ed. J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, pp. 69–95. Oxford: Oxford University Press.

OAKLEY, J. E. (2002). Eliciting Gaussian process priors for complex computer codes. *Statistician* **51**, 81–97.

OAKLEY, J. E. & O'HAGAN, A. (2002). Bayesian inference for the uncertainty distribution of computer model outputs. *Biometrika* **89**, 769–84.

O'HAGAN, A. (1978). Curve fitting and optimal design for prediction (with Discussion). *J. R. Statist. Soc.* B **40**, 1–42.

O'HAGAN, A. (1992). Some Bayesian numerical analysis. In *Bayesian Statistics 4*, Ed. J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, pp. 345–63. Oxford: Oxford University Press.

O'HAGAN, A. (1998). Eliciting expert beliefs in substantial practical applications (with Discussion). *Statistician* **47**, 21–35.

O'HAGAN, A., BUCK, C. E., DANESHKHAH, A., EISER, J. R., GARTHWAITE, P. H., JENKINSON, D., OAKLEY, J. E. & RAKOW, T. (2006). *Uncertain Judgements: Eliciting Experts' Probabilities*. Chichester: Wiley.

RAIFFA, H. (1968). *Decision Analysis: Introductory Lectures on Choice Under Uncertainty*. Reading, MA: Addison-Wesley.

RASMUSSEN, C. E. & WILLIAMS, C. K. I. (2006). *Gaussian Processes for Machine Learning*. Cambridge, MA: MIT Press.

WALKER, S. G., DAMIEN, P., LAUD, P. W. & SMITH, A. F. M. (1999). Bayesian nonparametric inference for random distributions and related functions (with Discussion). *J. R. Statist. Soc.* B **61**, 485–527.

WALLEY, P. (1991). *Statistical Reasoning with Imprecise Probabilities*. London: Chapman and Hall.

WEST, M. (1988). Modelling expert opinion (with Discussion). In *Bayesian Statistics 3*, Ed. J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, pp. 493–508. Oxford: Oxford University Press.

WINKLER, R. L. (1967). The assessment of prior distributions in Bayesian analysis. *J. Am. Statist. Assoc.* **62**, 776–800.

[*Received July* 2005. *Revised October* 2006]