Nonparametric Prior Elicitation

Jeremy Oakley and Anthony O'Hagan

Deparment of Probability and Statistics, University of Sheffield, The Hicks Building,
Hounsfield Road, Sheffield, S3 7RH, UK;
E-mail: j.oakley@sheffield.ac.uk

1. INTRODUCTION

Global sensitivity analysis is recognized as an essential tool for investigating the effects of input parameter uncertainty in a complex model. To obtain meaningful results from a sensitivity analysis, it is important that the probability distributions for all the uncertain input parameters in the model accurately represent the beliefs of the model user or decision-maker. When little or no data related to these parameters are available, parameter distributions must be specified largely on the basis of expert knowledge. This is rarely a simple task.

A particular difficulty in this scenario is that to perform the global sensitivity analysis the full joint probability distribution is required for all the uncertain input parameters in the model. However, a full probability distribution implies an infinite number of probability judgments by the expert about the parameters, clearly something the expert is unable to provide. In practice it is only going to be possible to elicit a finite and typically small number of probability statements from the expert. These statements will typically take the form quantiles of the distribution, perhaps the mode and sometimes the mean or other moments. 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. The parametric family may imply additional beliefs about the

Further author information:

J.O.: Telephone: +44 (0)114 222 3853

parameters that the expert does not agree with. Second, no account is then taken of the many other possible distributions that might have fitted the elicited statements equally well. This clearly has consequences for a global sensitivity analysis; other distributions might produce very different results when the uncertainty they are describe is propagated through the computer model under investigation.

We present an approach which tackles both of these deficiencies. Our model is non-parametric, allowing the expert's distribution to take any continuous form. It also quantifies the uncertainty in the resulting elicited distribution. Formally, the expert's density function is treated as an unknown function, about which we make inference. The result is a posterior distribution for the expert's density function. The posterior mean serves as a 'best fit' elicited distribution, while the variance around this fit expresses the uncertainty in the elicitation.

Specifically, this is achieved by using a Gaussian process to describe our own beliefs about the expert's distribution. Our prior specification contains proper prior beliefs about the smoothness of the expert's distribution, but is ultimately vague in that we do not include any of our own beliefs about likely values of the uncertain input parameter. Data then comes in the form of the expert's summaries, such as their mean and various quantiles. Properties of Gaussian processes can then be exploited to update our beliefs about the expert's distribution analytically, conditional on various hyperparameters in our Gaussian process model. Finally, Markov Chain Monte Carlo methods are used to remove the conditioning on these hyperparameters to give a full, probabilistic description of our uncertainty about the expert's distribution.

Illustrations of our method are given using some simple real elicitation exercises.

2. THE ELICITATION METHOD.

Here we give a brief overview of the elicitation method. Full details can be found in [1]. The idea is to think of eliciting a prior distribution as a standard problem in Bayesian inference. We, the analyst, wish to make inferences about an unknown function $f(\theta)$, the expert's prior density function for θ. We first formulate our own prior beliefs about

$f(\theta)$. We then ask the expert for probability judgments about θ which we think of as data about $f(\theta)$. We then update our beliefs about $f(\theta)$ in light of this data.

2.1. A prior distribution for $f(\theta)$

We 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 [2] and [3], classification [3] and numerical analysis [4].

The Gaussian process is specified by giving its mean function and variance-covariance function. We will 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. Thus

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

Now 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

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

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 $c(.,.)$ must ensure that the prior variance-covariance matrix of any set of observations of $f(.)$ (or functionals of $f(.)$) is positive semi definite. Here we choose the function

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

This will be seen to be a mathematically convenient choice, and implies that $f(.)$ is infinitely differentiable with probability 1.

This formulation was given in [5], who were interested in quadrature for computationally expensive density functions.

Our model represents a belief that the expert's density function $f(\theta)$ will, to some extent, approximate to a member of the parametric family $g(\theta \,|\, u)$. However, 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 its prior mean, and so governs how well it approximates to the 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.

The hyperparameters of this model are $\alpha = (u, \sigma^2, b)$. Non-informative priors are given for u and σ^2, and a (proper) lognormal prior is assumed for the ratio b/v (reflecting a belief that the expert's density will be smooth over the range implied by v)

2.2. Prior to posterior updating

Data will come in the form of quantiles of the distribution and simple moments. Conditional on α, the posterior distribution of $f(\theta)$ can be derived analytically. Markov Chain Monte Carlo sampling is then used to remove the dependence on α (details in [1]

3. EXAMPLE

We will illustrated the method with a simple synthetic example. Suppose that the expert has the following density function for θ:

$$f(\theta) = \frac{0.4}{\sqrt{2\pi}} \exp\{-\frac{1}{2}(\theta+2)^2\} + \frac{0.6}{\sqrt{4\pi}} \exp\{-\frac{1}{4}(\theta-1)^2\}. \qquad (4)$$

It is further assumed that the expert can state $P(\theta < x)$ for any x. The expert is asked to give probabilities for the following x: {-3,-2,-1,0,1,2,3}. These probability judgments constitute our data. We do not ask for the mean in this example, though we assume that the expert has given us $P(\theta < \infty) = 1$.

We now use MCMC to sample from the posterior distribution of $\alpha(u, \sigma^2, b)$. For each sampled α, we generate one random density function from the conditional distribution of

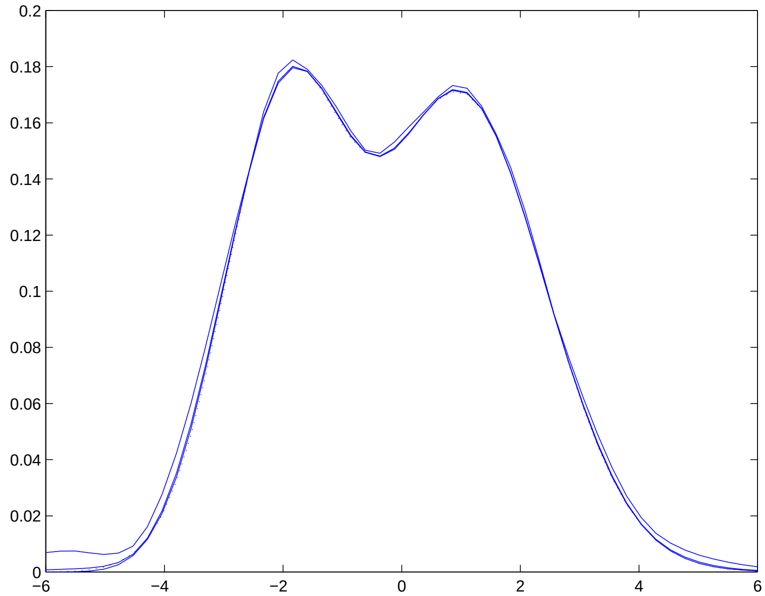

Figure 1. The mean and pointwise 95% intervals for the expert's density function (solid lines), and the true density function (dotted line).

$f(\theta)$ given α and the data. Given that we must have $f(\theta) > 0$, we discard any generated density functions that are negative over the range of interest. We then plot the pointwise mean, 2.5th and 97.5th percentiles from the distribution of the density function.

In figure 1 we can see that, without specifying a bimodal density function $f(\theta)$ in our prior for $f(\theta)$, we have correctly recovered the bimodal shape. We are also able to report our remaining uncertainty about $f(\theta)$ after eliciting the seven percentiles, which in this case is small.

References

[1] J. E. Oakley and A. O'Hagan. Uncertainty in prior elicitations: a nonparametric approach. Technical report, Department of Probability and Statistics, University of Sheffield, 2003.

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

[3] R. Neal. Regression and classification using gaussian process priors. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, editors, *Bayesian Statistics 6*, pages 69–95. Oxford: University Press, 1999.

[4] A. O'Hagan. Some Bayesian numerical analysis. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, editors, *Bayesian Statistics 4*, pages 345–363. Oxford: University Press, 1992.

[5] M. C. Kennedy and A. O'Hagan. Iterative rescaling for Bayesian quadrature. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, editors, *Bayesian Statistics 5*, pages 639–645. Oxford: University Press, 1996.