Elicitation of Multivariate Prior Distributions: A nonparametric Bayesian approach

Fernando A. Moala[1] and Anthony O'Hagan[2]

1 Department of Mathematics and Statistics, FCT-UNESP, Brazil

2 Department of Probability and Statistics, University of Sheffield,UK

Abstract

In the context of Bayesian statistical analysis, elicitation is the process of formulating a prior density $f(\cdot)$ about one or more uncertain quantities to represent a person's knowledge and beliefs. Several different methods of eliciting prior distributions for one unknown parameter have been proposed. However, there are relatively few methods for specifying a multivariate prior distribution and most are just applicable to specific classes of problems and/or based on restrictive conditions, such as independence of variables. Besides, many of these procedures require the elicitation of variances and correlations, and sometimes elicitation of hyperparameters which are difficult for experts to specify in practice. Garthwaite, Kadane and O'Hagan (2005) discuss the different methods proposed in the literature and the difficulties of eliciting multivariate prior distributions. We describe a flexible method of eliciting multivariate prior distributions applicable to a wide class of practical problems. Our approach does not assume a parametric form for the unknown prior density $f(\cdot)$, instead we use nonparametric Bayesian inference, modelling $f(\cdot)$ by a Gaussian process prior distribution. The expert is then asked to specify certain summaries of his/her distribution, such as the mean, mode, marginal quantiles and a small number of joint probabilities. The analyst receives that information, treating it as a data set D with which to update his/her prior beliefs to obtain the posterior distribution for $f(\cdot)$. Theoretical properties of joint and marginal priors are derived and numerical illustrations to demonstrate our approach are given.

Key words: Elicitation, expert, analyst, Gaussian process, prior distribution.

1 Introduction

In any practical statistical analysis there will always be some available form of knowledge about the issue apart from the experimental data, and Bayesian inference allows one to use this knowledge in the form of a prior distribution to achieve a more informative posterior distribution. Prior information is often sought from subject matter experts, and the process of obtaining that information and expressing it in the form of a probability distribution is known as elicitation. Despite the fundamental role that prior information plays in the Bayesian approach, much published Bayesian analysis employs 'noninformative' prior distributions without explicit consideration of any available prior information. This can be justified when the data are extensive, because sufficiently strong data will generally overwhelm any contribution from the prior distribution. However, another reason why genuine prior information appears to be so rarely used in practice may be the fact that elicitation is not a simple task, and techniques for accurate and reliable elicitation of expert knowledge are not readily available or widely known.

Note that the elicitation of the expert's beliefs in the form of probability distributions does not have to be thought of as 'prior' to some experimental data. O'Hagan (1998) presents two examples that illustrate this situation. In his second example, there is no practical possibility to obtain any experimental data after the elicitation is carried out. The expert's opinions are neither 'prior' nor 'posterior' but just represent the available knowledge of the experts at the time. Probabilistic risk analysis is an area that has made important applications of elicitation in this sense. Some examples of elicitation in risk analysis are given by Cooke (1991).

Garthwaite, Kadane and O'Hagan (2005), O'Hagan et al (2006) and Wolfson (1995) provide good reviews of elicitation concepts and the approaches proposed in the literature. A variety of methods for specifying opinions have been developed for unidimensional prior distributions. However, practical statistical models invariably contain many unknown parameters, and it is usually important to obtain the expert's information about several parameters. Yet there are few methods for specifying a multivariate prior distribution. Furthermore, most of these methods are just applicable to specific classes of problems, or rely on restrictive conditions such as independence of variables, or else require the elicitation of variances and covariances. It is known that experts are not generally able to specify second order moments reliably (Kadane and Wolfson, 1998). This has motivated us to propose a more flexible method of eliciting multivariate prior distributions, that is applicable to a wide class of practical problems.

We consider the elicitation of a single expert's beliefs about some unknown continuous random vector $\boldsymbol{\theta} = (\theta_1, \theta_2, . . ., \theta_k)$. An expert is a person who has background in the study area and enough knowledge to answer questions related to $\boldsymbol{\theta}$. The objective of the elicitation is to identify the underlying continuous density function $f(\boldsymbol{\theta})$ that represents the expert's beliefs.

It is assumed that the expert can state certain summaries of his/her distribution, such as the mean or percentiles, and can make probability judgements about $\boldsymbol{\theta}$ directly. It will often be necessary in practice to give training in these tasks (Alpert and Raiffa (1982) and Chaloner et al (1993)).

In most of the elicitation methods proposed in the literature, some convenient parametric form of the expert's prior is assumed to represent the unknown parameter $\boldsymbol{\theta}$; a conjugate prior is usually chosen. Based on the expert's assessments about $\boldsymbol{\theta}$ the hyperparameters of this prior are evaluated. Some examples of this are Al-Awadhi and Garthwaite (1998; 2001), Garthwaite and O'Hagan (2000) and Chaloner et al (1993).

The approach proposed here is a multivariate extension of Oakley and O'Hagan (2007). We do not assume a parametric form for $f(\boldsymbol{\theta})$; instead it is considered a random function $f(\cdot)$. In Bayesian nonparametric inference, a prior distribution for $f(\cdot)$ is then needed. To clarify this, we distinguish between the expert, whose probability density function for $\boldsymbol{\theta}$ is $f(\cdot)$, and the analyst, who is the person who elicits the expert's knowledge with a view to identifying $f(\cdot)$. The prior distribution for $f(\cdot)$ will express the analyst's personal prior beliefs about $f(\cdot)$. The role of the expert and analyst, and the nature of the analyst's prior information are discussed in Oakley and O'Hagan (2007). For convenience, we follow the convention that the expert is female and the analyst is male.

The expert provides various summaries of her distribution, such as the mean, mode, marginal quartiles and some joint probabilities. The analyst then uses the expert's stated summaries as data to update his prior and obtain his posterior distribution for $f(\cdot)$. The analyst's posterior mean can then be offered as a 'best estimate' for $f(\cdot)$, while his posterior variance quantifies the remaining uncertainty around this estimate.

We emphasize that our focus, as in Oakley and O'Hagan (2007), is on the analyst's Bayesian inference about $f(\cdot)$. The analyst's data are in the form of elicited summaries from the expert, representing her beliefs about $\boldsymbol{\theta}$. It may be that $f(\cdot)$ will be used as a prior distribution in some other Bayesian analysis to make inference about $\boldsymbol{\theta}$, but equally it may be elicited for some other purpose such as a probabilistic risk analysis. We will refer to $f(\cdot)$ simply as the expert's distribution, and any references to prior and posterior distributions concern the analyst's beliefs about $f(\cdot)$.

A very useful and flexible distribution to represent the analyst's prior beliefs about $f(\cdot)$ is the Gaussian process. Gaussian process models have been used in a wide variety of contexts to make inference about unknown functions and several properties and examples of its use are given in O'Hagan (1978), O'Hagan (1991), Neal (1999), Kennedy and O'Hagan (1996), Oakley and O'Hagan (2002).

The Gaussian process avoids forcing the expert's beliefs to fit a specified parametric family. Instead, it gives prior support across the space of all density functions, and so allows the true $f(\cdot)$ to have any form at all. Although we formulate the prior distribution with an expectation that the expert's distribution will approximate to a member of a parametric family, given enough elicited summaries from the expert the analyst's posterior distribution will converge to the true $f(\cdot)$, no matter how far that may be from his prior expectation. Furthermore, the analyst's posterior distribution explicitly quantifies the uncertainty around the estimate of $f(\cdot)$, given the (typically small) set of summaries elicited from the expert. Posterior distributions of marginal densities, cumulative distribution functions and moments derived from $f(\cdot)$ can be easily obtained.

This paper is structured as follows. In the next section we describe the analyst's prior beliefs about $f(\cdot)$. Section 3 describes the expert's data and Section 4 derives posterior inference about the expert's density. In Section 5 we present numerical illustrations of our procedure, while some conclusions are offered in Section 6. Proofs of posterior distributions for marginal densities, cumulative distribution functions and moments are given in the Appendix.

2 Gaussian process as prior representation for $f(\cdot)$

We assume that the expert's prior density is a relatively smooth function $f(\cdot)$: $\Re^k \rightarrow \Re$ of an unknown continuous random vector $\boldsymbol{\theta}$ =(θ_1, θ_2, . . ., θ_k). We will model a nonparametric prior for $f(\cdot)$ hierarchically in terms of a vector $\boldsymbol{\alpha}$ of hyperparameters. An appropriate choice to represent the analyst's prior beliefs about $f(\cdot)$ is a Gaussian process, which we denote by:

$$f(\cdot)\,|\,\boldsymbol{\alpha} \sim GP\Big(g(\cdot), \sigma^2 C(\cdot,\cdot)\Big) \tag{1}$$

where the mean and covariance functions

$$E\{f(\boldsymbol{\theta})\,|\,\boldsymbol{\alpha}\} = g(\boldsymbol{\theta}) \qquad Cov[f(\boldsymbol{\theta})\,,\, f(\boldsymbol{\phi})\,|\,\boldsymbol{\alpha}] = \sigma^2 C(\boldsymbol{\theta}\,,\boldsymbol{\phi}) \tag{2}$$

depend on the hyperparameters $\boldsymbol{\alpha}$.

The mean function expresses the analysts's prior expectation that the expert's distribution will be smooth, unimodal and roughly symmetric. Specifically, the analyst's prior mean is a multivariate normal distribution:

$$g(\boldsymbol{\theta}) = \frac{1}{(2\pi)^{\frac{k}{2}}}|\Sigma|^{-\frac{1}{2}} \exp\Big\{-\frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{m})^{\mathrm{t}}\Sigma^{-1}(\boldsymbol{\theta}-\boldsymbol{m})\Big\} \tag{3}$$

where the hyperparameters $\boldsymbol{m}$ and Σ are components $\boldsymbol{\alpha}$. However, the model is nonparametric and allows the true $f(\cdot)$ to have any form at all. In practice, the analyst may feel that the expert's distribution would approximate better to a member of some other parametric family. For instance, if $\boldsymbol{\theta}$ is a vector of positive random variables a good choice for $g(\cdot)$ might be a multivariate lognormal density, if $f(\cdot)$ may be expected to have a heavy tail then $g(\cdot)$ could be a multivariate t-Student, a Dirichlet distribution could be used if $\boldsymbol{\theta}$ is a vector of proportions, and so on. The basic theory developed here would apply in all of these cases, but we assume the multivariate normal distribution here because it will often be appropriate and because it allows certain integrals to be evaluated analytically.

We now need to choose a form for the covariance function. Equation (1) expresses this in terms of a common variance hyperparameter σ^2, controlling how closely the true density function will follow the analyst's prior mean $g(\cdot)$, and a scaled covariance function $C(\cdot,\cdot)$. However, it would not be realistic to set $C(\boldsymbol{\theta},\boldsymbol{\theta}) = 1$ for all $\boldsymbol{\theta}$, because in practice the variance of $f(\cdot)$ would not be the same for all $\boldsymbol{\theta}$. In general, where the analyst expects $f(\cdot)$ to be smaller his prior variance should be smaller in absolute terms. Based on the above considerations, and following Oakley and O'Hagan (2007), we let $C(\boldsymbol{\theta},\boldsymbol{\phi}) = g(\boldsymbol{\theta})g(\boldsymbol{\phi})c(\boldsymbol{\theta},\boldsymbol{\phi})$, and hence

$$Cov\big[f(\boldsymbol{\theta})\,,\,f(\boldsymbol{\phi})\,|\,\boldsymbol{\alpha}\big] = \sigma^2 C(\boldsymbol{\theta}\,,\boldsymbol{\phi}) = \sigma^2 g(\boldsymbol{\theta})g(\boldsymbol{\phi})c(\boldsymbol{\theta},\boldsymbol{\phi}). \tag{4}$$

This corresponds to modelling the ratio $r(\boldsymbol{\theta}) = \frac{f(\boldsymbol{\theta})}{g(\boldsymbol{\theta})}$ as a stationary Gaussian process with a constant mean 1 and covariance function $Cov[r(\boldsymbol{\theta})\,,\,r(\boldsymbol{\phi})\,|\,\boldsymbol{\alpha}] = \sigma^2 c(\boldsymbol{\theta}\,,\boldsymbol{\phi})$, where $c(\cdot,\cdot)$ is a correlation function satisfying $c(\boldsymbol{\theta},\boldsymbol{\theta}) = 1$.

The function $c(\boldsymbol{\theta}\,,\boldsymbol{\phi})$ should express the analyst's belief that the expert's density function $f(\cdot)$ is smooth, so that $r(\boldsymbol{\theta})$ and $r(\boldsymbol{\phi})$ will be more highly correlated the closer $\boldsymbol{\theta}$ is to $\boldsymbol{\phi}$, according to some appropriate metric. A useful and widely used correlation function $c(\cdot,\cdot)$ in Gaussian process modelling is $c(\boldsymbol{\theta},\boldsymbol{\phi}) = \exp\{-\frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\phi})^{\mathrm{t}}B^{-1}(\boldsymbol{\theta}-\boldsymbol{\phi})\}$. This form ensures that the covariance matrix of any set of observations of $f(\cdot)$ or functionals of $f(\cdot)$) is positive semi definite, and that $f(\cdot)$ is infinitely differentiable with probability 1. It also has convenient mathematical properties, which together with (3) allows some necessary integrals to be evaluated analytically. The matrix B describes how rough the true prior density $f(\cdot)$ is in each dimension of the input $\boldsymbol{\theta}$. We again follow Oakley and O'Hagan (2007) in arguing that the expert's belief that the shape of $f(\cdot)$ will approximate to that of the parametric family $g(\cdot)$ may be represented by assuming that $B = b\Sigma$. Thus, we formulate the covariance function as

$$c(\boldsymbol{\theta},\boldsymbol{\phi}) = \exp\Big\{-\frac{1}{2b}(\boldsymbol{\theta}-\boldsymbol{\phi})^{\mathrm{t}}\Sigma^{-1}(\boldsymbol{\theta}-\boldsymbol{\phi})\Big\}. \tag{5}$$

The hyperparameters of our model will be represented by $\boldsymbol{\alpha} = (\boldsymbol{m},\Sigma,b,\sigma^2)$ and the hierarchical model is completed by formulating the analyst's prior distribution for $\pi(\boldsymbol{\alpha})$ to reflect his prior beliefs about the unknown hyperparameters $\boldsymbol{\alpha}$.

As discussed by Oakley and O'Hagan (2007), the analyst's prior information should be restricted to generic information about the kind of density functions that the expert is likely to have, rather than any context specific information about $\boldsymbol{\theta}$ which would interfere with the objectives of eliciting the expert's knowledge. In particular, we express non-informative prior distributions for $\boldsymbol{m}, \Sigma$ and σ^2 via

$$\pi(\boldsymbol{\alpha}) \propto \frac{1}{\sigma^2}|\Sigma|^{-\frac{1}{2}(k+1)}\pi_b(b), \tag{6}$$

where $\pi_b(b)$ is the analyst's prior distribution for b. Oakley and O'Hagan (2007) point out that using a non-informative (improper) prior for b could lead to an improper posterior regardless of the data observed. We follow their proposal of a lognormal prior

$$log\, b \sim N(0\,; 1) \tag{7}$$

and therefore $\pi_b(b)$ represents this informative prior.

3 Data supplied by expert

Clearly, it is unreasonable to expect the expert to be able to state values of her density function at various values of $\boldsymbol{\theta}$. Thus it would be more reasonable to ask for probabilities such as $P\{\boldsymbol{\theta} \in A\}$, where A is a region of parameter space Θ. Since $f(\cdot)$ has a Gaussian process, the distribution of $P_A = P\{\boldsymbol{\theta} \in A\} = \int_A f(\boldsymbol{\theta})d\boldsymbol{\theta}$ is Normal with mean and variance given, respectively, by

$$E(P_A \,|\, \boldsymbol{\alpha}) = \int_A N_k(\boldsymbol{\theta} \,|\, m, \Sigma)d\boldsymbol{\theta} \tag{8}$$

and

$$cov\big[P_A, P_{A^*} \,|\, \boldsymbol{\alpha}\big] = \sigma^2\Big(\frac{b}{b+2}\Big)^{\frac{k}{2}} \int_A \int_{A^*} N_{2k}((\boldsymbol{\theta}^t, \boldsymbol{\phi}^t)^t \,|\, M, S)d\boldsymbol{\theta}d\boldsymbol{\phi} \qquad (9)$$

where the integrals represent probabilities of k- and $2k$-dimensional multivariate normal distributions with means $\boldsymbol{m}$ and $M = (\boldsymbol{m}^t, \boldsymbol{m}^t)^t$ and covariance matrices Σ and

$$S = \begin{bmatrix} \frac{1+b}{2+b}\Sigma & \frac{1}{2+b}\Sigma \\ \frac{1}{2+b}\Sigma & \frac{1+b}{2+b}\Sigma \end{bmatrix} = \begin{bmatrix} \frac{1+b}{2+b} & \frac{1}{2+b} \\ \frac{1}{2+b} & \frac{1+b}{2+b} \end{bmatrix} \otimes \Sigma, \qquad (10)$$

respectively, with $\otimes$ denoting the Kronecker product. Equations (8) and (9) are proved in the Appendix A1.

In general, P_A is a joint probability, which may also not be easy for the expert to specify. However, if A is a cylinder set, P_A is a marginal probability, which will usually be easier to specify. Marginal probabilities may also be specified implicitly as marginal quantiles; for instance, by specifying that u is the upper quartile of her distribution for θ_1, the expert implies that $P_A = 0.75$, where $A = (-\infty, u) \times \Re^{k-1}$. We will illustrate in Section 5 how in practice we can obtain good results by eliciting marginal probabilities supplemented by a small number of joint probabilities. It is also possible to incorporate elicited values for means or other moments, and also modes, within our framework (Oakley and O'Hagan, 2007), but for simplicity we assume that the elicited summaries from the expert comprise a data set $D = (P_1, \ldots, P_n)^t$ where P_i is the expert's assessment of the chance that $\boldsymbol{\theta} \in A_i$.

Since we know that $f(\cdot)$ is a proper density function, we include this in D by setting one of the A_is to be $\Re^k$ and giving the corresponding P_i the 'elicited' value of 1.

Conditional on $\boldsymbol{\alpha}$, D is normally distributed with mean H, variance $\sigma^2 W$ where the elements of H and W are given by (8) and (9), respectively. Consequently the likelihood $L(\boldsymbol{\alpha} \,|\, D)$ is given by

$$L(\boldsymbol{\alpha} \,|\, D) = \frac{|W|^{-\frac{1}{2}}}{\sigma^n} exp\Big\{-\frac{1}{2\sigma^2}(D-H)^t W^{-1}(D-H)\Big\}. \qquad (11)$$

4 Posterior distributions

We now derive the analyst's posterior distribution for the expert's distribution $f(\cdot)$ and other quantities derived from it, such as marginal densities. Using the hierarchical structure of the model, we will first obtain the posterior distribution of the hyperparameters, and discuss how to compute inferences using Markov chain Monte Carlo (MCMC) methods. We then derive posterior distributions for $f(\cdot)$ and other relevant inferences conditional on the hyperparameters $\boldsymbol{\alpha}$.

4.1 Hyperparameters $\boldsymbol{\alpha}$

Thus, from (11) and (6) the posterior density for $\boldsymbol{\alpha}$ =($\boldsymbol{m}$, Σ, b, σ^2) is given by

$$p(\boldsymbol{m}, \Sigma, b, \sigma^2 \mid D) \propto \pi_b(b)|\Sigma|^{-\frac{1}{2}(k+1)} \frac{|W|^{-\frac{1}{2}}}{\sigma^{n+2}} exp\Big\{-\frac{1}{2\sigma^2}(D-H)^t W^{-1}(D-H)\Big\}. \tag{12}$$

The conditional posterior distribution of σ^2 given the other hyperparameters can be seen to have the inverse-chi-square form

$$\sigma^2 \mid \boldsymbol{m}, \Sigma, b \sim \chi^{-2}(ns^2, n) \tag{13}$$

or equivalently

$$\frac{ns^2}{\sigma^2} \sim \chi^{-2}(n) \tag{14}$$

where $s^2 = \frac{1}{n}(D-H)^t W^{-1}(D-H)$.

The posterior distribution (12) can now be integrated with respect to σ^2 , to yield

$$p(\boldsymbol{m}, \Sigma, b \mid D) \propto |\Sigma|^{-\frac{1}{2}(k+1)} \left[(D-H)^t W^{-1}(D-H)\right]^{-\frac{n}{2}} |W|^{-\frac{1}{2}} \pi_b(b). \quad (15)$$

The joint posterior (15) is a very complex function of $\boldsymbol{m}$, Σ and b, and we cannot obtain marginal posteriors analytically. Although the analyst will not generally be interested in the hyperparameters as such, all subsequent inferences will be derived conditionally on these hyperparameters, and so it is important to be able to compute expectations with respect to (15).

Because of the intractability of the joint density (15) we will employ Markov chain Monte Carlo (MCMC) techniques to draw a large sample from the posterior distribution of these hyperparameters. At each iteration of the Metropolis-Hastings algorithm, each parameter $\boldsymbol{m}$, Σ and b is updated in turn by sampling a new value from its proposal function. This way, after a proposal value $\boldsymbol{m}_{(i+1)}$ is generated from the normal distribution $N_k(\boldsymbol{m}_{(i)}\,; V)$, which it is accepted or rejected according to the usual Metropolis-Hastings rule, then a new proposal $\Sigma_{(i+1)}$ is generated from the inverse Wishart distribution $IW(d\ ,\ \frac{1}{d-k-1}\Sigma_{(i)})$, which it is accepted or rejected and finally a proposal $b_{(i+1)}$ is generated from the lognormal distribution $LN(ln\, b_{(i)}\,; \nu)$, which is accepted or rejected.

In the examples of Section 5, we ran the chain for 20000 iterations. The proposal distribution parameters V, d and ν were chosen to obtain good mixing of the chains.

The MCMC algorithm produces sampled values of $\boldsymbol{\alpha}^* = (\boldsymbol{m}, \Sigma, b)$. If we wish to have samples of $\boldsymbol{\alpha}$, we can supplement each sampled $\boldsymbol{\alpha}^*_{(i)}$ by a value $\sigma^2_{(i)}$ sampled from its full conditional distribution (13) given $\boldsymbol{\alpha}^* = \boldsymbol{\alpha}^*_{(i)}$.

4.2 The expert's density function $f(\cdot)$

Conditional on $\boldsymbol{\alpha}$, we can determine the posterior distribution $f(\cdot) \mid D, \boldsymbol{\alpha}$ analytically using well-known properties of multivariate normal distributions. This is because $f(\boldsymbol{\theta})$ together with all the items in D have a multivariate normal distribution. First, the prior covariance of $f(\boldsymbol{\theta})$ with D has elements

$$Cov\big[f(\boldsymbol{\theta}), P_A|\,\boldsymbol{\alpha}\big] = Cov\Big[f(\boldsymbol{\theta}), \int_A f(\boldsymbol{\phi})d\boldsymbol{\phi}\,|\,\boldsymbol{\alpha}\Big] = \int_A Cov\big[f(\boldsymbol{\theta}), f(\boldsymbol{\phi})\,|\,\boldsymbol{\alpha}\big]d\boldsymbol{\theta} \tag{16}$$

$$= \sigma^2 g(\boldsymbol{\theta}) \int_A g(\boldsymbol{\phi}) exp\Big\{-\frac{1}{2b}(\boldsymbol{\theta}-\boldsymbol{\phi})^t\Sigma^{-1}(\boldsymbol{\theta}-\boldsymbol{\phi})\Big\}d\boldsymbol{\phi} \tag{17}$$

using (4) and (5). After some algebra (17) can be rewritten as

$$cov\big[f(\boldsymbol{\theta})\,, P_A\,|\,\boldsymbol{\alpha}\big] = \sigma^2\Big(\frac{b}{b+2}\Big)^{\frac{k}{2}} \int_A N_{2k}\Big((\boldsymbol{\theta}^t, \boldsymbol{\phi}^t)^t\,|\,M, S\Big)d\boldsymbol{\phi} \tag{18}$$

where $M = (\boldsymbol{m}^t, \boldsymbol{m}^t)^t$ and $S = \begin{bmatrix} \frac{b+1}{b+2}\Sigma & \frac{1}{b+2}\Sigma \\ \frac{1}{b+2}\Sigma & \frac{b+1}{b+2}\Sigma \end{bmatrix}$. Therefore,

$$cov\big[f(\boldsymbol{\theta})\,, P_A\,|\,\boldsymbol{\alpha}\big] = \sigma^2\Big(\frac{b}{b+2}\Big)^{\frac{k}{2}} N_k\Big(\boldsymbol{\theta}\,|\,\boldsymbol{m}, \frac{b+1}{b+2}\Sigma\Big) \int_A N_k\Big(\boldsymbol{\phi}\,|\,\boldsymbol{\theta}, \,\boldsymbol{m}^*, \Sigma^*\Big)d\boldsymbol{\phi} \tag{19}$$

where $\boldsymbol{m}^* = \boldsymbol{m} + \frac{1}{b+1}(\boldsymbol{\theta} - m)$ and $\Sigma^* = \frac{b}{b+1}\Sigma$. The prior covariance of $f(\boldsymbol{\theta})$ with D is then

$$cov[f(\boldsymbol{\theta}), D|\,\boldsymbol{\alpha}\big] = \begin{bmatrix} cov\big[f(\boldsymbol{\theta}),\, P_{A_1}\,|\,\boldsymbol{\alpha}\big] \\ cov\big[f(\boldsymbol{\theta}),\, P_{A_2}\,|\,\boldsymbol{\alpha}\big] \\ \vdots \\ cov\big[f(\boldsymbol{\theta}),\, P_{A_n}\,|\,\boldsymbol{\alpha}\big] \end{bmatrix} = \sigma^2\boldsymbol{t}(\boldsymbol{\theta}). \tag{20}$$

having elements (18).

Theorem 1: After observing D distributed as $D\,|\,\boldsymbol{\alpha} \sim N_n(H,\, \sigma^2 W)$ with $cov\big[D, f(\boldsymbol{\theta})\,\mid\,\boldsymbol{\alpha}\big] = \sigma^2\boldsymbol{t}(\boldsymbol{\theta})$ and $f(\cdot)$ having the Gaussian prior distribution given by (1) the posterior distribution $f(\cdot)\,|\,D, \boldsymbol{\alpha}$ is a Gaussian process with parameters

$$E\big[f(\boldsymbol{\theta})\,|\,D, \boldsymbol{\alpha}\big] = g^*(\boldsymbol{\theta}) = g(\boldsymbol{\theta}) + \boldsymbol{t}(\boldsymbol{\theta})^t W^{-1}(D - H) \tag{21}$$

and

$$Cov\big[f(\boldsymbol{\theta})\,, f(\boldsymbol{\phi})\,|\,D, \boldsymbol{\alpha}\big] = \sigma^2 C^*(\boldsymbol{\theta}, \boldsymbol{\phi}) = \sigma^2\big[C(\boldsymbol{\theta}, \boldsymbol{\phi}) - \boldsymbol{t}(\boldsymbol{\theta})^t W^{-1}\boldsymbol{t}(\boldsymbol{\phi})\big]. \quad (22)$$

Proof: The joint distribution of D, $f(\boldsymbol{\theta})$ and $f(\boldsymbol{\phi})$ is given by

$$\begin{pmatrix} D \\ f(\boldsymbol{\theta}) \\ f(\boldsymbol{\phi}) \end{pmatrix} \sim N\left[\begin{pmatrix} H \\ g(\boldsymbol{\theta}) \\ g(\boldsymbol{\phi}) \end{pmatrix}, \begin{pmatrix} \sigma^2 W & \sigma^2\boldsymbol{t}(\boldsymbol{\theta}) & \sigma^2\boldsymbol{t}(\boldsymbol{\phi}) \\ \sigma^2\boldsymbol{t}(\boldsymbol{\theta})^t & C(\boldsymbol{\theta}, \boldsymbol{\theta}) & C(\boldsymbol{\theta}, \boldsymbol{\phi}) \\ \sigma^2\boldsymbol{t}(\boldsymbol{\phi})^t & C(\boldsymbol{\theta}, \boldsymbol{\phi}) & C(\boldsymbol{\phi}, \boldsymbol{\phi}) \end{pmatrix}\right]. \quad (23)$$

The posterior distribution $f(\cdot)\,|\,D, \boldsymbol{\alpha}$ is simply obtained by conditioning on the data set D. From (23), the conditional distribution of two arbitrary points $f(\boldsymbol{\theta})$ and $f(\boldsymbol{\phi})$ on the function f is normal, and the argument generalizes trivially to show that the posterior distribution of any p points on the function is multivariate normal. Hence the posterior distribution of f is $GP(g^*(\cdot), C^*(\cdot,\cdot))$, with mean and covariance functions given by (21) and (22). ■

By integrating out σ^2, the conditional posterior distribution of $f(\cdot)$ given $\boldsymbol{\alpha}^* = (\boldsymbol{m}, \Sigma, b)$ becomes a t process. The conditional posterior distribution of $f(\cdot)$ is

$$f(\boldsymbol{\theta})|\, D, \boldsymbol{\alpha}^* \sim t_1\Big(g^*(\boldsymbol{\theta})\,, s^2 C^*(\boldsymbol{\theta},\, \boldsymbol{\theta})\,, n\Big), \quad (24)$$

or equivalently,

$$\frac{f(\boldsymbol{\theta}) - g^*(\boldsymbol{\theta})}{s\sqrt{C^*(\boldsymbol{\theta},\, \boldsymbol{\theta})}} \sim t_1(n). \quad (25)$$

We cannot now integrate out $\alpha^* = (\boldsymbol{m}, \Sigma, b)$ analytically to obtain the marginal posterior distribution, but can do so numerically using the MCMC sampling of hyperparameters described in Section 4.1. Given a sample $\boldsymbol{\alpha}^*_{(i)} =$

$(\boldsymbol{m}_{(i)}, \Sigma_{(i)}, b_{(i)})$, $i = 1, 2, . \ . \ ., N$ of N sets of these hyperparameter values, we can make inferences about $f(\cdot)$ using two alternative approaches.

First, we can use the law of iterated expectations to compute appropriate posterior moments and probabilities. For instance, one inference that will always be of interest is the posterior mean of $f(\cdot)$, since this is the analyst's 'best' point estimate of the expert's density. It is given by

$$E\big[f(\boldsymbol{\theta}) \,|\, D\big] = E\big[E\big[f(\boldsymbol{\theta}) \,|\, D, \boldsymbol{\alpha}^*\big]\big] = \int E\big[f(\boldsymbol{\theta}) \,|\, D, \boldsymbol{\alpha}^*\big] p(\boldsymbol{\alpha}^* \,|\, D) d\boldsymbol{\alpha}^*, \quad (26)$$

and may be computed from the MCMC sample for sufficiently large N as

$$E\big[f(\boldsymbol{\theta}) \,|\, D\big] = \frac{1}{N} \sum_{i=1}^{N} E\big[f(\boldsymbol{\theta}) \,|\, D, \boldsymbol{\alpha}^*_{(i)}\big] \quad (27)$$

where $E\big[f(\boldsymbol{\theta}) \,|\, D, \boldsymbol{\alpha}_{(\boldsymbol{i})}\big]$ is evaluated from (21). Note that since the 'observation' $\int_\Omega f(\boldsymbol{\theta}) d\boldsymbol{\theta} = 1$ is included in D this will be true a posteriori with probability 1, and in particular the mean will integrate to 1.

This approach can be used for any inference about $f(\cdot)$ that can be expressed as a posterior expectation. For instance, we can compute the posterior variance of $f(\boldsymbol{\theta})$ for any $\boldsymbol{\theta}$, or the probability that $f(\boldsymbol{\theta})$ exceeds any particular value.

The posterior covariance and variance of $f(\boldsymbol{\theta})$ are estimated respectively by

$$cov\big[f(\boldsymbol{\theta}), f(\boldsymbol{\phi}) \,|\, D\big] = E\Big[cov\big[f(\boldsymbol{\theta}), f(\boldsymbol{\phi}) \,|\, D, \alpha^*\big] \,|\, D, \alpha^*\Big] +$$
$$+ cov\Big[E\big[f(\boldsymbol{\theta}) \,|\, D, \alpha^*\big], E\big[f(\boldsymbol{\phi}) \,|\, D, \alpha^*\big] \,|\, D, \alpha^*\Big] \quad (28)$$

and

$$var\big[f(\theta) \,|\, D\big] = E\Big[var\big[f(\theta) \,|\, D, \alpha^*\big] \,|\, D, \alpha^*\Big] + var\Big[E\big[f(\theta) \,|\, D, \alpha^*\big] \,|\, D, \alpha^*\Big] \quad (29)$$

The second approach is more useful to compute quantiles of the posterior distribution or other inferences that are not readily expressed in terms of expectations. We simulate a realization of $f(\cdot)$ from its Gaussian process conditional posterior distribution given $\boldsymbol{\alpha} = \boldsymbol{\alpha}_{(i)}$, for $i = 1,2,\ldots,N$, using the approach of Oakley and O'Hagan (2002). This represents a sample from the unconditional posterior distribution of $f(\cdot)$, from which any desired inference can be computed.

In the following subsections we present the analyst's posterior distributions for various other aspects of the expert's distribution such as the expert's marginal densities, joint and marginal distribution functions and moments. This would allow the analyst to identify further features of the expert's density still not evaluated in order to have a better representation of her beliefs. These inferences, in particular marginal distribution functions and moments, will also enable the analyst to give the expert feedback. In order to verify that the expert has provided a true reflection of her beliefs, graphical presentation of marginal densities and probabilities derived from marginal and joint function distributions may be provided by the analyst. The expert would be then asked to account for any discrepancies from these estimated data and would be given an opportunity to reflect upon the results of the elicitation process.

4.3 Posterior distribution for marginal prior densities

The following result shows that the analyst's posterior distribution for the expert's marginal density

$$f_i(\theta_i) = \int_{\Re^{k-1}} f(\boldsymbol{\theta}) d\boldsymbol{\theta}_{-i}. \tag{30}$$

is a Gaussian process conditional on $\boldsymbol{\alpha}$.

Theorem 2: Under the conditions given by Theorem 1, the marginal distribution $f_i(\theta_i)= \int_{\Re^{k-1}} f(\boldsymbol{\theta}) d\boldsymbol{\theta}_{-i}$, $i = 1, \ldots, k$ has a Gaussian process posterior distribution conditional on hyperparameters $\boldsymbol{\alpha}$ with mean and covariance functions:

$$E\big[f_i(\theta_i)\,|\,D,\boldsymbol{\alpha}\big] = g_i^*(\theta_i) = g_i(\theta_i) + \boldsymbol{t}_i(\theta_i)^t W^{-1}(D-H) \tag{31}$$

and

$$Cov\big[f_i(\theta_i)\,,f_i(\phi_i)\,|\,D,\boldsymbol{\alpha}\big] = \sigma^2 C_i^*(\theta_i,\phi_i) = \sigma^2\Big[C_i(\theta_i,\phi_i) - \boldsymbol{t}_i(\theta_i)^t W^{-1}\boldsymbol{t}_i(\phi_i)\Big] \tag{32}$$

respectively, where

$$g_i(\theta_i) = \frac{1}{\sqrt{2\pi}\sigma_i} exp\big\{-\frac{1}{2\sigma_i^2}(\theta_i - m_i)^2\big\}, \tag{33}$$

$$C_i(\theta_i,\phi_i) = \Big(\frac{b}{b+2}\Big)^{\frac{k-1}{2}} g_i(\theta_i) g_i(\phi_i) c_i(\theta_i\,,\phi_i), \tag{34}$$

with

$$c_i(\theta_i\,,\phi_i) = exp\big\{-\frac{1}{2b\sigma_i^2}(\theta_i - \phi_i)^2\big\}, \tag{35}$$

and the elements of vector $\boldsymbol{t}_i(\theta_i)$ are given by

$$\Big(\frac{b}{b+2}\Big)^{\frac{k}{2}} N_1\big(\theta_i|\, m_i\,, \frac{b+1}{b+2}\sigma_i^2\big) \int_A N_k\Big(\boldsymbol{\phi}\,|\,\boldsymbol{m}(\theta_i), S_i^*\Big) d\boldsymbol{\phi} \tag{36}$$

for each region A considered in the data set D, with $\boldsymbol{m}(\theta_i)= \boldsymbol{m} + \frac{1}{1+b}\frac{1}{\sigma_i^2}(\theta_i - m_i)\Sigma_i$, $S_i^* = \frac{1+b}{2+b}\Big(\Sigma - \frac{1}{(1+b)^2}\frac{1}{\sigma_i^2}\Sigma_i\Sigma_i^t\Big)$ and $\Sigma_i^t = (\sigma_{1i}\,, \sigma_{2i},\, \ldots, \sigma_i^2, ..., \sigma_{ki})$.

It's important to point out that in practice we still need to average over $\boldsymbol{\alpha}$ (e.g., using MCMC).

Now, by integrating out σ^2 we obtain the posterior distribution of $f_i(\cdot)$ given $\boldsymbol{\alpha}^* =(\boldsymbol{m}, \Sigma, b)$ by

$$f_i(\theta_i) | \boldsymbol{\alpha}^*, D \sim t_1\Big(g_i^*(\theta_i)\,,\, s^2 C_i^*(\theta_i, \phi_i)\,,\, n\Big)$$

or equivalently

$$\frac{f_i(\theta_i) - g_i^*(\theta_i)}{s\sqrt{C_i^*(\theta_i, \phi_i)}} \sim t(n). \tag{37}$$

where $s^2 = \frac{1}{n}(D-H)^t W^{-1}(D-H)$. ■

Proof: See Appendix A3.

Theorem 2 can be extended to give the analyst's posterior distribution for the expert's marginal distribution of any subset of the elements of $\boldsymbol{\theta}$ as conditionally a Gaussian process; see Moala (2006).

4.4 Posterior distribution for expert's distribution functions

Theorem 3: Let $f(\cdot)$ be a prior density distributed as Gaussian process given by (1) and D the data set composed by the expert's summaries. The analyst's posterior distribution for $F(\boldsymbol{x}) = \int_{-\infty}^{\boldsymbol{x}} f(\boldsymbol{\theta}) d\boldsymbol{\theta}$ is a Gaussian process with

$$E\big[F(\boldsymbol{x})\,|\,D, \boldsymbol{\alpha}\big] = \Phi_k(\boldsymbol{x}\,|\,\boldsymbol{m}, \Sigma) + \boldsymbol{t}^*(\boldsymbol{x})^t W^{-1}(D-H) \tag{38}$$

and

$$Cov\big[F(\boldsymbol{x})\,,\, F(\boldsymbol{y})\,|\,D, \boldsymbol{\alpha}\big] = \sigma^2\big[C_{\boldsymbol{xy}}(\boldsymbol{x}, \boldsymbol{y}) - \boldsymbol{t}^*(\boldsymbol{x})^t W^{-1} \boldsymbol{t}^*(\boldsymbol{y})\big] \tag{39}$$

where

$$C_{\boldsymbol{xy}}(\boldsymbol{x},\boldsymbol{y}) = \Big(\frac{b}{b+2}\Big)^{\frac{k}{2}} \Phi_{2k}\big((\boldsymbol{x},\boldsymbol{y})\,|\,M,S\big)\,, \tag{40}$$

and $\boldsymbol{t}^{*}(\boldsymbol{x})$ is a vector with elements

$$\int_{-\infty}^{\boldsymbol{x}}\int_{A} N_{2k}\Big((\boldsymbol{\theta},\boldsymbol{\phi})\,|\,M,S\Big)d\boldsymbol{\theta}d\boldsymbol{\phi} \tag{41}$$

for each region A considered in the data set D, and where Φ_k and Φ_{2k} denote the cumulative functions of multivariate normal distributions with the given mean vectors and covariance matrices defined in Section 3.

Proof: See Appendix A4. ■

Again, by integrating out σ^2, the conditional posterior distribution of $F(\cdot)$ given $\alpha^{*}=(\boldsymbol{m},\Sigma,b)$ becomes

$$\frac{F(\boldsymbol{x}) - E\big[F(\boldsymbol{x})\,|\,D,\,\boldsymbol{\alpha}^{*}\big]}{S\sqrt{C_{\boldsymbol{xy}}(\boldsymbol{x},\,\boldsymbol{y}) - \boldsymbol{t}^{*}(\boldsymbol{x})^{t}W^{-1}\boldsymbol{t}^{*}(\boldsymbol{y})}} \sim t(n). \tag{42}$$

Note that the approach proposed by Oakley and O'Hagan (2007) generates a distribution function interpolated by the probabilities elicited from expert given in D. Indeed, from equations (39) to (41) we can show that the posterior variance in those points is zero.

Suppose now we are interested in the probability that a single element θ_i of $\boldsymbol{\theta}$ takes on a value less than or equal to a number s. Consider therefore the marginal distribution function $F_i(x_i)$ in one of the dimensions independently from the other dimensions. This marginal distribution can be computed as

$$F_i(s) = F(+\infty,+\infty,\ldots,\,s,+\infty,\,\ldots,+\infty)\,. \tag{43}$$

Thus, the analyst's posterior distribution of $F_i(s)$ can be derived directly from the posterior distribution of the joint function distribution $F(\boldsymbol{x})$, resulting in a Gaussian process with

$$E\big[F_i(s)\,|\,D,\boldsymbol{\alpha}\big] = \Phi_1(s\,|m_i,\sigma_i^2) + \boldsymbol{t}_i^*(s)^t W^{-1}(D-H) \tag{44}$$

and

$$Cov\big[F_i(s)\,,F_i(p)\,|\,D,\boldsymbol{\alpha}\big] = \sigma^2\big[C_{sp}(s,p) - \boldsymbol{t}_i^*(s)^t W^{-1}\boldsymbol{t}_i^*(p)\big] \tag{45}$$

where

$$C_{sp}(s,p) = \Big(\frac{b}{b+2}\Big)^{\frac{k}{2}}\Phi_2\big((s,p)|\,\boldsymbol{m}_i,\sigma_i^2 V\big) \tag{46}$$

and $\boldsymbol{t}_i^*(s)$ is a vector with elements given by

$$\Big(\frac{b}{b+2}\Big)^{\frac{k}{2}}\int_{-\infty}^{s}\int_{A_j} N_{k+1}\Big((\boldsymbol{\theta},\boldsymbol{\phi})\,|\,M_i,S_i\Big)d\boldsymbol{\theta}d\boldsymbol{\phi} \tag{47}$$

for $\boldsymbol{m}_i = (m_i,m_i)$, $V = \begin{bmatrix}\frac{b+1}{b+2} & \frac{1}{b+2}\\ \frac{1}{b+2} & \frac{b+1}{b+2}\end{bmatrix}$, $M_i = (m_i,\boldsymbol{m})$, $S_i = \begin{bmatrix}\frac{b+1}{b+2}\sigma_i^2 & \frac{1}{b+2}\Sigma_i^t\\ \frac{1}{b+2}\Sigma_i & \frac{b+1}{b+2}\Sigma\end{bmatrix}$ and $\Sigma_i^t = (\sigma_{1i}\,,\sigma_{2i},\,\ldots,\sigma_i^2,...,\sigma_{ki})$.

4.5 Posterior distribution for moments of θ

By deriving the moments and central moments of the random variable $\boldsymbol{\theta}$, such as mean, variance and covariance, we will have information about the behavior of $\boldsymbol{\theta}$ and therefore they can be used as measures of various characteristics of the expert's density $f(\cdot)$.

Theorem 4: Under the Gaussian process proposed by the analyst in Theorem 1 the posterior distribution of the marginal moment of order r,

$$\mu_i^{(r)} = E(\theta_i^r) = \int_{-\infty}^{+\infty}\theta_i^r f_i(\theta_i)d\theta_i\ \ ,\ i=1,\,\ldots,\,k, \tag{48}$$

conditional on $\boldsymbol{\alpha}$, is normal with mean and variance, respectively, given by:

$$E\big[\mu_i^{(r)}\,|\,D,\boldsymbol{\alpha}\big] = \int_{-\infty}^{+\infty} \theta_i^r g_i(\theta_i) d\theta_i + \big[\boldsymbol{t}_i^{(r)}\big]^t W^{-1}(D-H)\,, \tag{49}$$

and

$$var\big[\mu_i^{(r)}|\,D,\boldsymbol{\alpha}\big] = \sigma^2\Big[C_i^{(r)} - \big[\boldsymbol{t}_i^{(r)}\big]^t W^{-1}\boldsymbol{t}_i^{(r)}\Big], \tag{50}$$

where

$$C_i^{(r)} = \Big(\frac{b}{b+2}\Big)^{\frac{k}{2}} \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \theta_i^r \phi_i^r N_2((\theta_i,\phi_i)\,|\,\boldsymbol{m}_i,\sigma_i^2 V) d\theta_i d\phi_i, \tag{51}$$

and $\boldsymbol{t}_i^{(r)}$ is a vector σ^2 with elements given by

$$\Big(\frac{b}{b+2}\Big)^{\frac{k}{2}} \int_A \psi(\boldsymbol{\phi}) N_k\big(\boldsymbol{\phi}|\,\boldsymbol{m}, \frac{b+1}{b+2}\Sigma\big) d\boldsymbol{\phi}, \tag{52}$$

$\psi(\boldsymbol{\phi}) = \int_{-\infty}^{+\infty} \theta_i^r N_1\big(\theta_i\,|\,m_i^*, \sigma_i^{*2}\big) d\theta_i$ is the moment of order r of normal distribution and $m_i^* = m_i + \frac{b+2}{b+1}\Sigma_i^t\Sigma^{-1}(\boldsymbol{\phi}-\boldsymbol{m})$ and $\sigma_i^{*2} = \frac{b+1}{b+2}\sigma_i^2 - \frac{b+2}{b+1}\Sigma_i^t\Sigma^{-1}\Sigma_i$ for $i = 1, \ldots, k$.

Proof: See Appendix A5. ■

The most immediate use for Theorem 4 is to give the analyst's posterior distribution for the expert's mean value of each θ_i.

Corollary 5: The expert's mean $\mu_i = E(\theta_i)$ has a normal posterior distribution conditional on $\boldsymbol{\alpha}$ with parameters

$$E\big[\mu_i\,|\,D,\,\boldsymbol{\alpha}\big] = m_i + \big[\boldsymbol{t}_i^{(1)}\big]^t W^{-1}(D-H)\,, \tag{53}$$

$$var\big[\mu_i\,|\,D,\boldsymbol{\alpha}\big] = \sigma^2\Big[\Big(\frac{b}{b+2}\Big)^{\frac{k}{2}}\big(m_i^2 + \frac{1}{b+2}\sigma_i^2\big) - \big[\boldsymbol{t}_i^{(1)}\big]^t W^{-1}\boldsymbol{t}_i^{(1)}\Big] \tag{54}$$

where $\boldsymbol{t}_i^{(1)}$ is a vector with elements

$$\Big(\frac{b}{b+2}\Big)^{\frac{k}{2}}\Big[m_i\int_A N_k\big(\theta\,|\,\boldsymbol{m},\frac{b+1}{b+2}\Sigma\big)d\theta + \frac{b+2}{b+1}\int_A \Sigma_i^t\Sigma^{-1}(\boldsymbol{\phi}-\boldsymbol{m})N_k\big(\boldsymbol{\phi}\,|\,\boldsymbol{m},\frac{b+1}{b+2}\Sigma\big)d\boldsymbol{\phi}\Big] \tag{55}$$

Proof: This is simply the case $r = 1$ of Theorem 4. ■

Moala (2006) generalizes Theorem 4 to give the posterior distribution of the joint raw moments of θ_1, . . ., θ_k of the form

$$E(\theta_1^{r_1}\ldots\theta_k^{r_k}) = \int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\theta_1^{r_1}\ldots\theta_k^{r_k}f(\theta_1,\ldots,\theta_k)d\theta_1\ldots d\theta_k\,, \tag{56}$$

where the r_i's are 0 or any positive integer. Although we cannot derive the distributions for central moments, the following are obtained by Moala (2006):

$$E\big[var(\theta_i)\,|\,D,\alpha\big] = m_i^2 + \sigma_i^2\; + \big[\boldsymbol{t}_i^{(2)}\big]^t W^{-1}(D-H)\; -$$

$$\sigma^2\Big[\Big(\frac{b}{b+2}\Big)^{\frac{k}{2}}\big(m_i^2 + \frac{1}{b+2}\sigma_i^2\big) - \big[\boldsymbol{t}_i^{(1)}\big]^t W^{-1}\boldsymbol{t}_i^{(1)}\Big] - \Big[m_i + \big[\boldsymbol{t}_i^{(1)}\big]^t W^{-1}(D-H)\Big]^2 \tag{57}$$

for $i = 1,\ldots,k$.

$$E\big[cov(\theta_i,\theta_j)\,|D,\alpha\big] = m_i m_j + \sigma_{ij} + \boldsymbol{t}_{ij}^{*t}W^{-1}(D-H) - \Big[m_i + \big[\boldsymbol{t}_i^{(1)}\big]^t W^{-1}(D-H)\Big]$$

$$\Big[m_j + \big[\boldsymbol{t}_j^{(1)}\big]^t W^{-1}(D-H)\Big] - \sigma^2\Big[\Big(\frac{b}{b+2}\Big)^{\frac{k}{2}}(m_i m_j + \frac{\sigma_{ji}}{2+b}) - \big[\boldsymbol{t}_i^{(1)}\big]^t W^{-1}\boldsymbol{t}_j^{(1)}\Big] \tag{58}$$

where

$$\boldsymbol{t}_{ij}^{*} = \Big(\frac{b}{b+2}\Big)^{\frac{k}{2}} \begin{bmatrix} \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\theta_i\theta_j\, N_2\big((\theta_i,\theta_j),\boldsymbol{m}_{ij}\,,\Sigma_{ij}^{*}\big)\int_{A_1} N_k\big(\boldsymbol{\phi}|\,\boldsymbol{m}^{(2)},S^{(2)}\big)d\boldsymbol{\phi}d\theta_i\theta_j \\ \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\theta_i\theta_j\, N_2\big((\theta_i,\theta_j),\boldsymbol{m}_{ij}\,,\Sigma_{ij}^{*}\big)\int_{A_2} N_k\big(\boldsymbol{\phi}|\,\boldsymbol{m}^{(2)},S^{(2)}\big)d\boldsymbol{\phi}d\theta_i\theta_j \\ \vdots \\ \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\theta_i\theta_j\, N_2\big((\theta_i,\theta_j),\boldsymbol{m}_{ij}\,,\Sigma_{ij}^{*}\big)\int_{A_n} N_k\big(\boldsymbol{\phi}|\,\boldsymbol{m}^{(2)},S^{(2)}\big)d\boldsymbol{\phi}d\theta_i\theta_j \end{bmatrix} \tag{59}$$

for $i = 1, \ldots, k$ and $j = 1, \ldots, k$.

5 Numerical Illustration

Although the procedure proposed in this paper has been developed for any number of parameters, the illustrative examples considered in this section deal just with eliciting bivariate distributions. Besides the fact that the bivariate case is already challenging enough and computationally intensive, we need to learn how the procedure works in two dimensions better before attempting higher dimensions. Moreover, the elicitation of bivariate distributions is by itself quite important in many practical situations.

In the two examples in this section we chose which joint probabilities to elicit based on already knowing the true distribution. This ensured that we obtained probabilities able to capture the main features of the true distribution such as modes, saddlepoints, heavy tails, etc. Our examples show that our approach is capable of identifying realistic joint distributions using a realistic number of elicited probabilities, at least if those probabilities are well chosen. The complexities of a serious practical elicitation are beyond the scope of this article, but will be addressed in a forthcoming paper which begins to demonstrate that effective elicitation can indeed be achieved in practice using our methods.

5.1 Example 1: Bimodal density

We suppose that the expert has the following density function for $\boldsymbol{\theta} = (\theta_1,\, \theta_2)$ proposed in O'Hagan (1994) and given by

$$f(\theta_1, \theta_2) = 0.4 \times N\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}0.5 & 0\\0 & 0.5\end{bmatrix}\right) + 0.6 \times N\left(\begin{bmatrix}1.5\\2\end{bmatrix}, \begin{bmatrix}0.5 & 0.3\\0.3 & 0.5\end{bmatrix}\right) \quad (60)$$

Four regions containing important features of this density such as saddlepoint and modes were identified and we suppose that the expert assesses their probabilities accurately. The probabilities are $P\{0 \leq \theta_1 \leq 1{,}0.3 \leq \theta_2 \leq 1.5\} = 0.12$, $P\{0 \leq \theta_1 \leq 1.2{,}1 \leq \theta_2 \leq +\infty\} = 0.11$, $P\{1.2 \leq \theta_1 \leq 2{,}1.5 \leq \theta_2 \leq +\infty\} = 0.21$ and $P\{2 \leq \theta_1 \leq +\infty{,}1.5 \leq \theta_2 \leq +\infty\} = 0.14$. Besides, the expert provides the marginal quartiles for θ_1 given by 0.15, 1.0, and 1.7; and for θ_2 given by 0.2, 1.4 and 2.2. With this information we can compose the data set with ten probabilities as $D = \{0.25{,}0.25{,}0.25{,}0.25{,}0.25{,}0.25{,}0.12{,}0.11{,}0.21{,}0.14\}$.

After generating the values of hyperparameters $\boldsymbol{m}$, Σ, b by MCMC a computer algorithm is performed and the posterior mean function of Gaussian process is constructed as a best elicited distribution for the expert density $f(\cdot)$. The plots of the expert density $f(\cdot)$ and the analyst's expected density are shown in Figure 1.

Figure 1 shows us that although the analyst's prior beliefs indicate a unimodal shape for $f(\cdot)$ the data elicited from the expert are sufficiently informative to yield a bimodal analyst's expected density very close to the expert's density.

Figure 2 displays both expert's and analyst's expected density contours of $f(\cdot)$ in order to give us a visual comparison of how accurate the fit is.

The contours of the analyst's expected density agree very well around the main mode while they are just reasonable around the second mode. The analyst's density seems to provide an adequate representation of the expert's beliefs.

We also note in the contour plot a little influence of the normal prior mean function proposed in the analyst's prior beliefs. This is caused because we have not had a sufficient number of judgements of $\boldsymbol{\theta}$ around the saddlepoint and because of the small number of elicited joint probabilities.

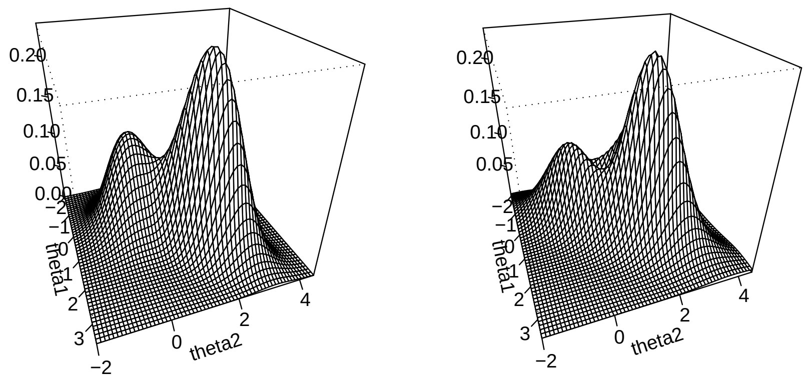

Figure 1: Plots of expert's density (a) given by (60) and analyst's expected density (b).

Thus, our procedure is able to capture with a good accuracy the most important features of the expert's distribution by just making use of a small amount of data, allowing us therefore to validate our procedure.

One way to assess the accuracy of fitting can be made through the posterior variance function given in (29) which is plotted in Figure 3 along with contour plot.

Another indicator is provided by the contours of fifth and ninety-fifth percentiles of the posterior distribution of $f(\cdot)$ are plotted in Figure 4 where dotted lines indicate the level curves of 5th percentiles and solid lines indicate level curves of 95th percentiles. The percentiles allow the analyst to describe the uncertainty surrounding the density $f(\cdot)$. Very small differences between the two bounds correspond to low uncertainty about the density function value. On the other hand, large differences will imply that the remaining uncertainty is still large.

Figure 4 shows that the elicited data have not identified the expert's density without some appreciable posterior uncertainty. After the expert's

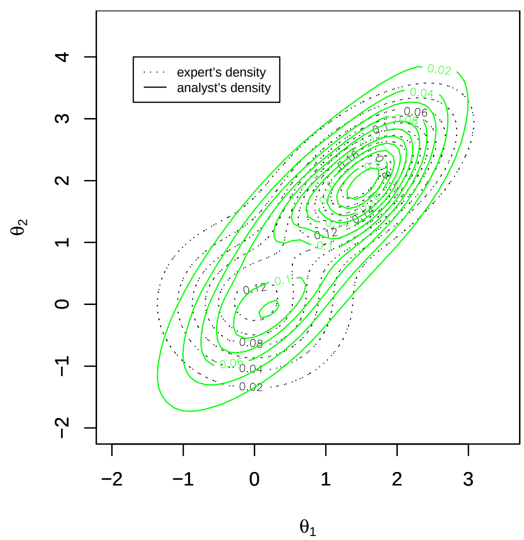

Figure 2: Contours of the analyst's expected and expert's densities.

joint density is derived, the analyst might be interested in having more details about her joint density such as moments and marginal densities. We have seen that the possibility to obtain the marginal densities is an interesting result provided by our approach. From expert's joint density (60) the marginal densities are given by

$$f_1(\theta_1) = 0.4 \times N(0\,, 0.5) + 0.6 \times N(1.5, 0.5) \tag{61}$$

and

$$f_2(\theta_2) = 0.4 \times N(0\,, 0.5) + 0.6 \times N(2, 0.5) \tag{62}$$

To visually assess how good the estimation is, we plot the expert's and analyst's posterior marginal densities in Figure 5. Also plotted are the 5th and 95th percentiles from the distribution of the density function indicating pointwise 90% credible intervals.

Figure 5 shows that the marginal distributions have been captured well by the elicitation procedure. The example is a challenging one not just

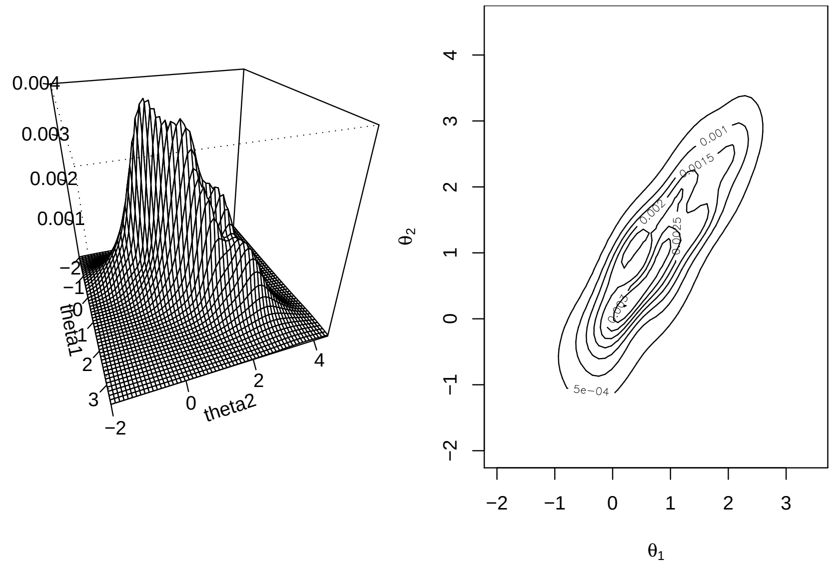

Figure 3: Variance of the posterior Gaussian process given in (29).

because the joint distribution is bimodal (which will be unusual in practice) but because this bimodality gives a bimodal margin for θ_2 but not for θ_1. Nevertheless the θ_1 margin has a shoulder which is nearly a second mode. The analyst's posterior nicely shows these features, where it is quite clear about the bimodality of θ_2 but more equivocal about θ_1.

Marginal and joint cumulative functions can also be derived through our approach. Based on the estimated cumulative functions we can derive the estimated probabilities of 'elicited' regions in order to compare them with the probabilities supplied by the expert. To visually determine whether the expert's knowledge has been elicited adequately, we can examine the plots of marginal cumulative distribution functions as shown in Figure 6.

From marginal densities (61) and (62) the marginal distribution functions are given by

$$F_1(x) = 0.4 \times \Phi_1(x \,|\, 0\,, 0.5) + 0.6 \times \Phi_1(x \,|\, 1.5, 0.5) \tag{63}$$

and

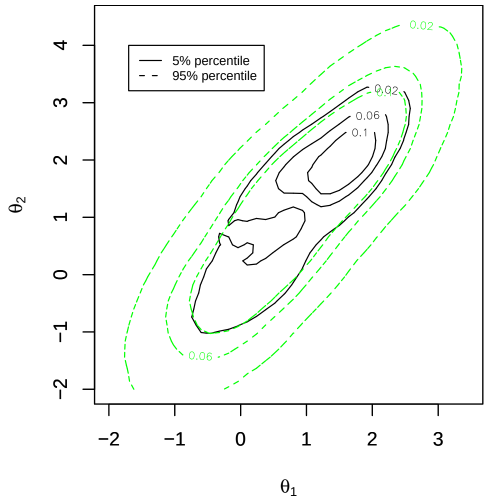

Figure 4: Contours of 5th and 95th percentiles for the density $f(\cdot)$.

$$F_2(y) = 0.4 \times \Phi_1(y \,|\, 0\,, 0.5) + 0.6 \times \Phi_1(y \,|\, 2, 0.5) \tag{64}$$

These plots clearly show a better accuracy of the estimated marginal distribution functions than the densities. This was expected since in the Oakley and O'Hagan (2007) approach the distribution function $F(x)$ exactly interpolates the expert's judgements. Thus, the uncertainty goes down close to the judgements.

From the posterior means of the moments given in Section 4.5, the summaries of expert's density can also be evaluated. The estimates of mean and variance of the distribution of $\boldsymbol{\theta}$ are compared with the true values and shown in Table 1.

Table 1: Marginal and joint moments of $(\theta_1\,,\,\theta_2)$

measures	Exact Values	Analyst' expected values
$E(\theta_1)$	0.9	0.905 (0.031*)
$E(\theta_2)$	1.2	1.199 (0.049*)
$var(\theta_1)$	1.04	1.150
$var(\theta_2)$	1.06	1.935
$cov(\theta_1, \theta_2)$	0.9	1.199

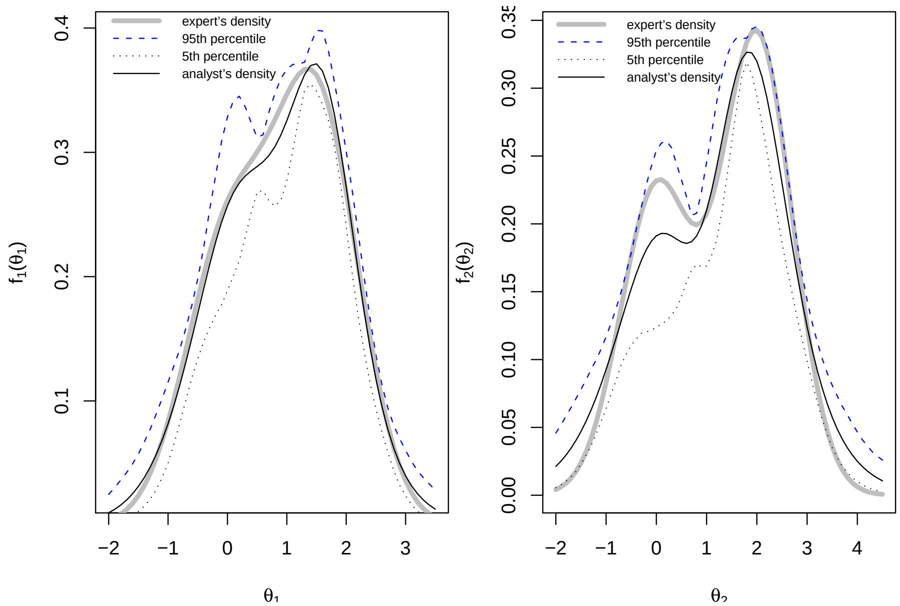

Figure 5: Plot of analyst's and expert's marginal densities.

∗ standard deviation

If we are interested in the expectations $\mu_1 = E(\theta_1)$ or $\mu_2 = E(\theta_2)$ our approach allows us to obtain this information with good accuracy. Variances and covariances can also be estimated by this procedure with good accuracy. By Figure 8 and Table 1 we can observe that four joint probabilities have provided enough information about the structure of dependence between the variables and, hence, the covariance was obtained with good precision. This is a difficult task to carry out and very important information in the modeling of multivariate distributions.

5.2 Example 2: Strictly positive variables with skew density

In this example we will illustrate a situation in which θ_1 and θ_2 are positive variables and the expert's density has strong skewness. We assume

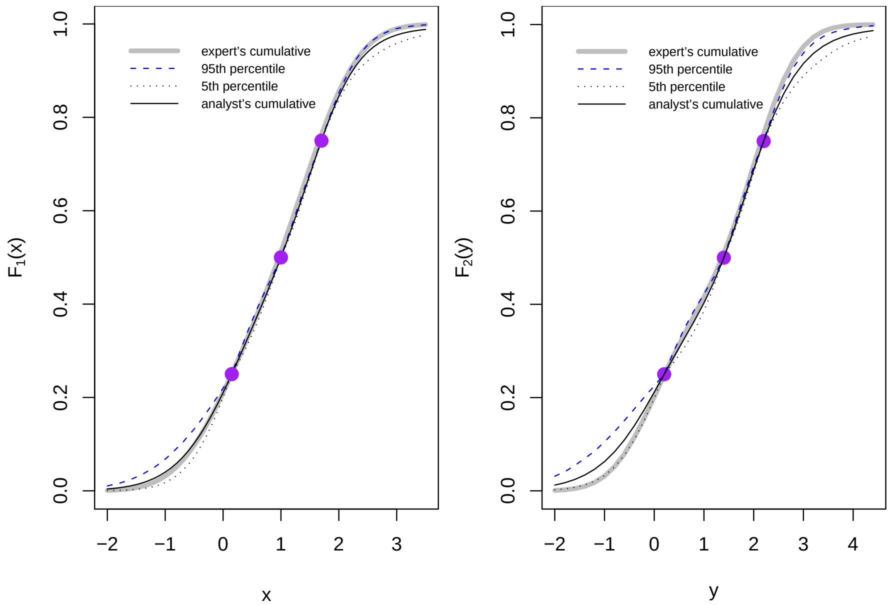

Figure 6: Plot of analyst's and expert's marginal function distributions.

a bivariate Lognormal distribution for $\boldsymbol{\theta} = (\theta_1,\, \theta_2)$ as the expert's density given by

$$(log\,\theta_1, log\,\theta_2) \sim N\left(\begin{bmatrix}2.8\\3\end{bmatrix}, \begin{bmatrix}0.1 & 0.05\\0.05 & 0.1\end{bmatrix}\right). \tag{65}$$

Suppose now that the expert provides a probability assessment for marginal θ_1, giving quartiles: 13.3, 16.5 and 20.4; and for θ_2: 16.3, 20.1 and 24.9. She also provides three joint probabilities given by: $P\{0 \leq \theta_1 \leq 20, 0 \leq \theta_2 \leq 14.7\} = 0.15$, $P\{0 \leq \theta_1 \leq 13.4, 14.7 \leq \theta_2 \leq 25\} = 0.15$ and $P\{19 \leq \theta_1 \leq 35, 18 \leq \theta_2 \leq 37\} = 0.25$.

Figure 7 shows the plots of the expert density $f(\cdot)$ and the analyst's expected density.

Although a normal density for the prior mean function is not an appropriate choice in the case where both variables θ_1 and θ_2 are known to be positive, our procedure has provided a good fit for the expert's density. We can observe this good fit in Figure 8, where both analyst's expected and expert's density contours are plotted.

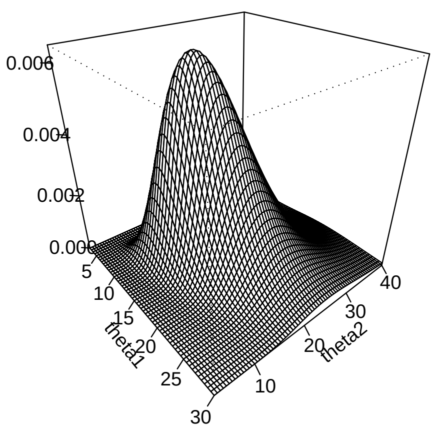

Figure 7: Plots of expert's density (a) given by (65) and analyst's expected density (b).

We can also visually assess the remaining uncertainty associated with the density $f(\cdot)$ by plotting the contours of 90% credible interval bounds as shown in Figure 9.

From the expert's joint density (65) the marginal distributions are given by

$$log\,\theta_1 \sim N(2.8, 0.1) \quad and \quad log\,\theta_2 \sim N(3.0, 0.1) \tag{66}$$

An inspection of the posterior mean of marginal densities in Figure 10 shows a good estimation of the expert's densities. Nevertheless, we can observe that the uncertainty surrounding the tails increases slightly.

In order to examine the accuracy of the point estimates of moments (means, variances and covariance) we compare them with real values given in Table 2. We can observe a reasonable accuracy of these estimators.

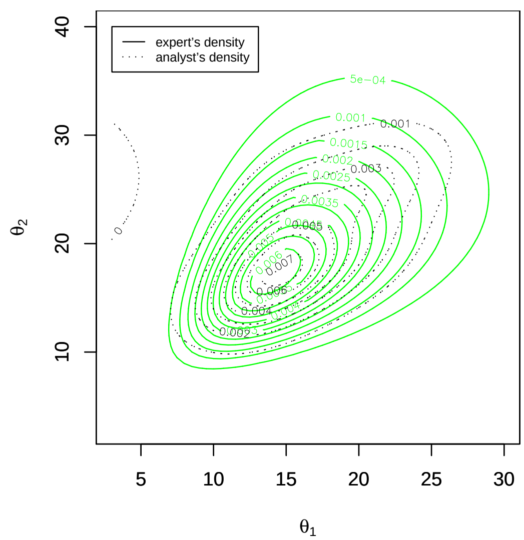

Figure 8: Contours of the analyst's expected density and expert's density.

Table 2: Marginal and joint moments of (θ_1 , θ_2)

measures	Exact Values	Analyst' expected values
$E(\theta_1)$	17.288	16.136 (1.254*)
$E(\theta_2)$	21.115	19.611 (1.515*)
$var(\theta_1)$	31.432	31.429
$var(\theta_2)$	46.891	46.385
$cov(\theta_1, \theta_2)$	18.716	26.693

∗ standard deviation

6 Conclusion

Eliciting joint probability distributions from experts is a difficult problem but an important one. Previous work in the literature has been reviewed by O'Hagan et al (2006), and suffers from a number of restrictions. There are various methods based on assuming that the expert's distribution lies within

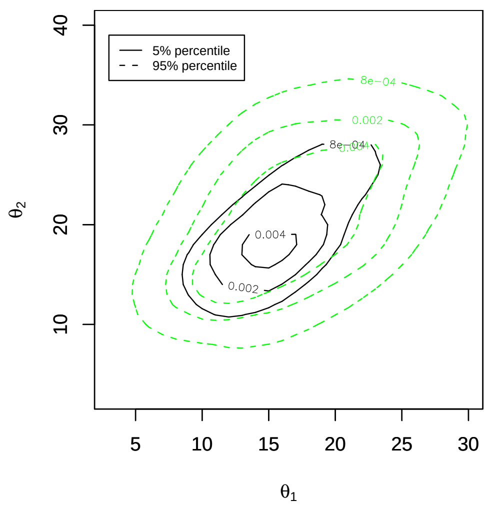

Figure 9: Contours of 5th and 95th percentiles for the density $f(\cdot)$.

a particular family of multivariate distributions (e.g. the multivariate normal family), and then eliciting enough summaries from the expert to identify the parameters of this distribution. Our approach does not make any such assumptions, and indeed is fully nonparametric. We do use a multivariate normal distribution as a prior expectation, but simply on the basis that it is reasonable to suppose that the expert's distribution is not too dissimilar to such a standard parametric form (and other forms can be used). Nevertheless, as we have seen particularly in our first example in Section 5, the method can adapt to expert opinions that do not fit this prior expectation.

Some authors have allowed the marginal distributions to be arbitrary, but have assumed a specific copula form to convert these to a joint distribution. Again, this is a restrictive assumption that is not made in our approach. Also, such methods typically require the direct elicitation of correlation measures from the expert, and the psychological literature suggests that this is not a task that experts can perform reliably. Other approaches have required the elicitation of moments, which have the same psychological difficulties. Our method can make use of elicited moments if these are felt to be reliable, but is based primarily on eliciting just probabilities.

Finally, all other methods result in just a single fitted distribution to represent the expert's elicited summaries, with no indication of uncer-

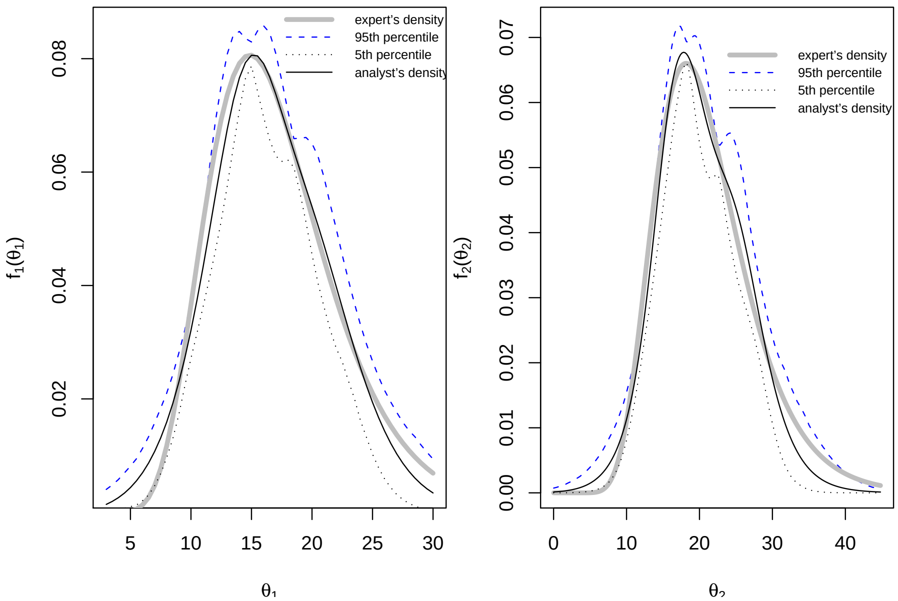

Figure 10: Plot of analyst's and expert's marginal densities.

tainty, yet it is clear that no matter how many elicited values have been obtained from the expert they cannot be enough to identify a joint distribution uniquely. Our approach makes use a realistically small number of elicited probabilities and characterises uncertainty around the 'fitted' distribution (the analyst's posterior mean) by the posterior variance (and indeed a full posterior distribution).

We believe therefore that our approach is uniquely suitable for the practical elicitation of expert knowledge about two or more uncertain quantities, particularly when there is no strong reason to believe that that knowledge can be represented by a particular parametric family. There are, nevertheless, a number of outstanding issues to be addressed.

Foremost among these is the practicality of the approach in serious elicitation. We have demonstrated it here only in some hypothetical examples, which have been enough to demonstrate its power to represent expert opinion adequately given only a modest number of elicited probabilities (and, importantly, only a small number of joint probabilities). A forthcoming paper will offer a strategy for practical elicitation and a serious practical example. However, much more practical experience would be required to assess its viability in a range of elicitation contexts.

Another important limitation of the present work (and of the forthcoming paper) is that we have only tried the method for bivariate elicitation. Eliciting a joint distribution for three or more uncertain quantities will be more demanding, of course. Nevertheless, even if higher dimensions prove to be intractable, we believe that a method to elicit bivariate distributions which addresses all the deficiencies listed above of existing approaches is a major step forward in the field.

Our method is a generalization of that of Oakley and O'Hagan (2007), which has one deficiency in that the analyst's posterior distribution does not enforce the natural requirement that $f(\boldsymbol{\theta}) \geq 0$ for all $\boldsymbol{\theta}$. In practice, this is rarely a problem because the elicited probabilities ensure that the posterior probability of the expert's density being negative at any point is small. For instance, in the first example of Section 5, the variance plotted in Figure 3 gives a posterior standard deviation which is comfortably less than half of the posterior mean plotted in Figure 2. If negative density is an issue, then it can be addressed in practice by using the approach of simulating realizations of $f(\cdot)$ as described in Section 4.2 and rejecting realizations that become negative.

We should also mention that our method assumes that the expert can accurately assess the probabilities that the analyst requests. In practice, of course, the expert's judgement of such values will be fallible, and so there is another element of imprecision in the elicitation which our analyst's posterior distribution does not account for. This is another question that is briefly addressed by Oakley and O'Hagan (2007) in the univariate case, but which would benefit from much closer study.

In conclusion, our approach to multivariate elicitation uniquely allows the expert's true underlying density to have arbitrary form, does not make unrealistic demands of the expert and incorporates an assessment of the uncertainty in the fitted distribution. We have shown that it has the potential to achieve an adequate representation of the expert's distribution in realistic bivariate examples. There are a number of outstanding issues to be addressed, and in particular ongoing work aims to provide a practical protocol for using the method in bivariate elicitation.

7 References

Al-Awadhi, S.A. and Garthwaite, P.H. (1998) An elicitation method for multivariate normal distributions. Communications in Statistics: Theory and Methods,} 27, 1123-42.

Al-Awadhi, S.A. and Garthwaite, P.H. (2001) Prior distribution assessment for a multivariate normal distribution: An experimental study. Journal of Applied Statistics}, 28, 5-23.

Chaloner, K.M., Church, T., Louis, T.A., and Matts, J.P. (1993). Graphical elicitation of a prior distribution for a clinical trial. The Statistician, 42, 341-353.

Cooke, R. M. (1991) Experts in uncertainty: opinion and subjective probability in Science, Oxford University Press.

Garthwaite, P. H., Kadane, J. B. and O'Hagan, A. (2005). Statistical methods for eliciting probability distributions. Journal of the American Statistical Association, 100, 680-701.

Garthwaite, P. and O'Hagan, A. (2000). Quantifying expert opinion in the water industry: an experimental study. The Statistician 49, 455-477.

Kadane, J. B. and Wolfson, L. J. (1998) Experiences in elicitation. Statistician, 47,3-19.

Kennedy, M. C. and O'Hagan, A. (1996). Iterative rescaling for Bayesian quadrature, in Bayesian Statistics 5, edited by Bernardo, J. M., Berger, J. O., Dawid, A. P. and Smith, A. F. M., pp. 639-645, Oxford: University Press.

Moala, F.A. (2006). Elicitation of Multivariate Prior Distributions. PhD Thesis. Department of Statistics. University of Sheffield.

Neal, R. M. (1999). Regression and classification using Gaussian process priors (with discussion). In Bernardo, J. M., Berger, J. O., Dawid, A. P., and Smith, A. F. M., editors, Bayesian Statistics 6, pages 475-501. Oxford University Press.

Oakley, J. E. and O'Hagan, A. (2007). Uncertainty in prior elicitations. Biometrika 94, 427–441.

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

O'Hagan, A. (1978). Curve Fitting and Optimal Design for Prediction (with discussion). Journal of the Royal Statistical Society B 40(1), 1–42.

O'Hagan, A. (1991). Bayes-Hermite quadrature. Journal of Statistical Planning and Inference 29, 245-260.

O'Hagan, A. (1998). Eliciting Expert Beliefs in Substantial Practical Applications, Statistician 47: (1), pp. 21-35 (with discussion, pp 55-68).

O'Hagan, A. (1994). Kendall's Advanced Theory of Statistics: Bayesian Inference. v.B. Wiley.

O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. E., Garthwaite, P. H., Jenkinson, D. J., Oakley, J. E. and Rakow, T. (2006). Uncertain Judgements: Eliciting Expert Probabilities. Wiley.

Wolfson, L. J. (1995). Elicitation of priors and utilities for Bayesian Analysis. PhD Thesis. Department of Statistics, Carnegie Mellon University, Pittsburgh.

Appendix: proof of results

In this Appendix we provide some formal proofs of the results used in this paper.

A1. Distribution of data supplied by expert

In Section 3 it is established that the expert's summaries composing the data D are normally distributed. We formalize this with a detailed proof.

Since $f(\cdot)$ has a Gaussian process distribution and $P_A = \int_A f(\boldsymbol{\theta})d\theta$ is a linear functional of $f(\cdot)$, then distribution of P_A is normally distributed with mean

$$E(P_A \mid \boldsymbol{\alpha}) = \int_A E\big[f(\boldsymbol{\theta}) \mid \alpha\big]d\theta = \int_A g(\boldsymbol{\theta})d\boldsymbol{\theta} = \int_A N_k(\boldsymbol{\theta} \mid m, \Sigma)d\boldsymbol{\theta} \qquad \text{(A1)}$$

and variance obtained from:

$$cov(P_A\,, P_{A^*}|\ \boldsymbol{\alpha}) = \int_A\int_{A^*} cov\big[f(\boldsymbol{\theta}), f(\boldsymbol{\phi}) \mid \boldsymbol{\alpha}\big]d\boldsymbol{\theta}d\boldsymbol{\phi} = \sigma^2 \int_A\int_{A^*} g(\boldsymbol{\theta})g(\boldsymbol{\phi})c(\boldsymbol{\theta},\boldsymbol{\phi})d\boldsymbol{\theta}d\boldsymbol{\phi} =$$
$$= \sigma^2 \frac{1}{(2\pi)^k}|\Sigma|^{-1} \int_A\int_{A^*} exp\Big\{-\frac{1}{2}[(\boldsymbol{\theta}-\boldsymbol{m})^t\Sigma^{-1}(\boldsymbol{\theta}-\boldsymbol{m}) + (\boldsymbol{\phi}-\boldsymbol{m})^t\Sigma^{-1}(\boldsymbol{\phi}-\boldsymbol{m}) +$$
$$+ \frac{1}{b}(\boldsymbol{\theta}-\boldsymbol{\phi})^t\,\Sigma^{-1}(\boldsymbol{\theta}-\boldsymbol{\phi})\big]\Big\}\, d\boldsymbol{\theta}d\boldsymbol{\phi}. \qquad \text{(A2)}$$

After some algebra, (A2) can be rewrite as

$$cov(P_A, P_{A^*}|\boldsymbol{\alpha}) = \frac{\sigma^2}{(2\pi)^k}|\Sigma|^{-1}|S|^{\frac{1}{2}} \int_A\int_{A^*} |S|^{-\frac{1}{2}} exp\Big\{-\frac{1}{2}[(\boldsymbol{\theta},\boldsymbol{\phi}) - (\boldsymbol{m},\boldsymbol{m})]S^{-1}[(\boldsymbol{\theta},\boldsymbol{\phi}) -$$
$$- (\boldsymbol{m},\boldsymbol{m})]^T\Big\}\, d\boldsymbol{\theta}d\boldsymbol{\phi} \qquad \text{(A3)}$$

that is,

$$cov(P_A\,, P_{A^*}|\ \boldsymbol{\alpha}) = \sigma^2|\Sigma|^{-1}|S|^{\frac{1}{2}} \int_A\int_{A^*} N_{2k}\Big((\boldsymbol{\theta},\boldsymbol{\phi}) \mid M, S\Big)d\boldsymbol{\theta}d\boldsymbol{\phi}, \qquad \text{(A4)}$$

where $M = (\boldsymbol{m}, \boldsymbol{m})$ and $S^{-1} = \begin{bmatrix} \frac{1+b}{b}\Sigma^{-1} & -\frac{1}{b}\Sigma^{-1} \\ -\frac{1}{b}\Sigma^{-1} & \frac{1+b}{b}\Sigma^{-1} \end{bmatrix} = V^{-1} \otimes \Sigma^{-1}$ with $V^{-1} = \begin{bmatrix} \frac{1+b}{b} & -\frac{1}{b} \\ -\frac{1}{b} & \frac{1+b}{b} \end{bmatrix}$ and denoting the Kronecker product. Now, $S^{-1} = V^{-1} \otimes \Sigma^{-1}$ implies

$$S = (V^{-1} \otimes \Sigma^{-1})^{-1} = V \otimes \Sigma = \begin{bmatrix} \frac{1+b}{2+b} & \frac{1}{2+b} \\ \frac{1}{2+b} & \frac{1+b}{2+b} \end{bmatrix} \otimes \Sigma = \begin{bmatrix} \frac{1+b}{2+b}\Sigma & \frac{1}{2+b}\Sigma \\ \frac{1}{2+b}\Sigma & \frac{1+b}{2+b}\Sigma \end{bmatrix}. \quad \text{(A5)}$$

From properties of the Kronecker product $|S| = |V \otimes \Sigma| = |V|^k |\Sigma|^2$. But $|\Sigma|^{-1}|S|^{\frac{1}{2}} = |\Sigma|^{-1}\left[|V|^k|\Sigma|^2\right]^{\frac{1}{2}} = |\Sigma|^{-1}|V|^{\frac{k}{2}}|\Sigma| = |V|^{\frac{k}{2}} = \left(\frac{b}{b+2}\right)^{\frac{k}{2}}$
then

$$cov(P_A, P_{A^*}|\, \boldsymbol{\alpha}) = \sigma^2\Big(\frac{b}{b+2}\Big)^{\frac{k}{2}} \int_A \int_{A^*} N_{2k}((\boldsymbol{\theta}, \boldsymbol{\phi})\,|\, M, S) d\boldsymbol{\theta} d\boldsymbol{\phi} \quad \text{(A6)}$$

A2. Posterior distribution for the hyperparameter σ^2

The posterior distribution for the hyperparameter σ^2 is given by $\sigma^2 \,|\, \boldsymbol{m}, \Sigma, b, D \sim$Inv-gamma$(v = \frac{n}{2}, \beta = \frac{ns^2}{2})$ where $s^2 = \frac{1}{n}(D - H)^t W^{-1}(D - H)$.

proof: From eq. (11) the likelihood function for σ^2 is obtained as

$$L(\sigma^2 \,|\, \boldsymbol{m}, \Sigma, b, D) \propto \frac{1}{\sigma^n} exp\Big\{-\frac{1}{2\sigma^2}(D - H)^t W^{-1}(D - H)\Big\} \quad \text{(A7)}$$

and using the prior $\pi_\sigma(\sigma^2) \propto \frac{1}{\sigma^2}$ the posterior density for σ^2 is then given by

$$p(\sigma^2 \,|\, \boldsymbol{m}, \Sigma, b, D) \propto \frac{1}{\sigma^{n+2}} exp\Big\{-\frac{1}{2\sigma^2}(D - H)^t W^{-1}(D - H)\Big\} \quad \text{(A8)}$$

that is,

$$p(\sigma^2 \mid \boldsymbol{m}, \Sigma, b, D) = c\frac{1}{\sigma^{n+2}} exp\Big\{-\frac{nS^2}{2\sigma^2}\Big\}. \qquad \text{(A9)}$$

where $s^2 = \frac{1}{n}(D-H)^t W^{-1}(D-H)$ and $c^{-1} = \int_0^\infty \frac{1}{\sigma^{n+2}} exp\Big\{-\frac{nS^2}{2\sigma^2}\Big\} d\sigma^2 = \Gamma(\frac{n}{2})\Big(\frac{nS^2}{2}\Big)^{-\frac{n}{2}}$. Therefore,

$$p(\sigma^2 \mid \boldsymbol{m}, \Sigma, b, D) = \frac{\Big(\frac{ns^2}{2}\Big)^{\frac{n}{2}}}{\Gamma(\frac{n}{2})} \frac{1}{\sigma^{n+2}} exp\Big\{-\frac{ns^2}{2\sigma^2}\Big\}. \qquad \text{(A10)}$$

A3. Posterior distribution for marginal prior densities

In this Section we derive in full detail the posterior distribution of the expert's marginal densities $f_i(\theta_i) = \int_{\Re^{k-1}} f(\boldsymbol{\theta}) d\boldsymbol{\theta}_{-i}$.

Now, since the posterior distribution $f(\cdot) \mid D, \alpha$ is a Gaussian process distribution then the posterior distribution of $\int_{\Re^{k-1}} f(\boldsymbol{\theta}) d\theta_{-i}$ is also a Gaussian process distribution with mean function given by

$$E\big[f_i(\theta_i) \mid D, \boldsymbol{\alpha}\big] = \int_{\Re^{k-1}} E\big[f(\boldsymbol{\theta}) \mid D, \boldsymbol{\alpha}\big] d\theta_{-i} = g_i(\theta_i) + \int_{\Re^{k-1}} \boldsymbol{t}(\boldsymbol{\theta})^t W^{-1}(D-H) d\boldsymbol{\theta}_{-i}. \qquad \text{(A11)}$$

Because the vector $W^{-1}(D-H)$ does not depend on $\boldsymbol{\theta}$ then $\boldsymbol{t}(\boldsymbol{\theta})^t W^{-1}(D - H)$ is linear function of $\boldsymbol{t}(\boldsymbol{\theta})^t$ and consequently,

$$E\big[f_i(\theta_i) \mid D, \boldsymbol{\alpha}\big] = g_i(\theta_i) + \boldsymbol{t}_i(\theta_i)^t W^{-1}(D-H) \qquad \text{(A12)}$$

where $\boldsymbol{t}_i(\theta_i) = \int_{\Re^{k-1}} \boldsymbol{t}(\boldsymbol{\theta})^t d\boldsymbol{\theta}_{-i}$. Now, from (20), $\sigma^2 \boldsymbol{t}(\boldsymbol{\theta}) = cov\big[f(\boldsymbol{\theta}), D|\boldsymbol{\alpha}\big]$ is a vector with elements

$$cov\big[f(\boldsymbol{\theta}), P_A \mid \boldsymbol{\alpha}\big] = \sigma^2 \Big(\frac{b}{b+2}\Big)^{\frac{k}{2}} \int_A N_{2k}\Big((\boldsymbol{\theta}, \boldsymbol{\phi}) \mid M, S\Big) d\boldsymbol{\phi} \qquad \text{(A13)}$$

where $M = (\boldsymbol{m}, \boldsymbol{m})$ and $S = \begin{bmatrix} \frac{b+1}{b+2}\Sigma & \frac{1}{b+2}\Sigma \\ \frac{1}{b+2}\Sigma & \frac{b+1}{b+2}\Sigma \end{bmatrix} = V \otimes \Sigma$ with $V = \begin{bmatrix} \frac{1+b}{2+b} & \frac{1}{2+b} \\ \frac{1}{2+b} & \frac{1+b}{2+b} \end{bmatrix}$. Therefore, $\boldsymbol{t}_i(\theta_i)$ has elements given by

$$t_i^A(\theta_i) = \Big(\frac{b}{b+2}\Big)^{\frac{k}{2}} \int_A \Big[\int_{\Re^{k-1}} N_{2k}\big((\boldsymbol{\theta}, \boldsymbol{\phi})|\, M, S\big) d\boldsymbol{\theta}_{-i}\Big] d\boldsymbol{\phi}. \tag{A14}$$

The interior integral in (A13) is the marginal density of the random vector $(\theta_i, \boldsymbol{\phi})$ distributed as $N_{k+1}\big(M_i, S_i\big)$ where $M_i = (m_i, \boldsymbol{m})$, $S_i = \begin{bmatrix} \frac{1+b}{2+b}\sigma_i^2 & \frac{1}{2+b}\Sigma_i^t \\ \frac{1}{2+b}\Sigma_i & \frac{1+b}{2+b}\Sigma \end{bmatrix}$ and $\Sigma_i^t = (\sigma_{1i}\,, \sigma_{2i},\, \ldots, \sigma_i^2, ..., \sigma_{ki})$. Consequently, the elements of vector $\boldsymbol{t}_i(\theta_i)$ will be given by

$$t_i^A(\theta_i) = \Big(\frac{b}{b+2}\Big)^{\frac{k}{2}} \int_A N_{k+1}\big((\theta_i, \boldsymbol{\phi})|\, M_i, S_i\big) d\boldsymbol{\phi}. \tag{A15}$$

Due to $(\theta_i, \boldsymbol{\phi}) \sim N_{k+1}\big(M_i, S_i\big)$ then $\theta_i \sim N_1\big(m_i\,, \frac{1+b}{2+b}\sigma_i^2\big)$ and $\boldsymbol{\phi}|\theta_i \sim N_k\big(\boldsymbol{m}(\theta_i), S_i^*\big)$ with $\boldsymbol{m}(\theta_i) = \boldsymbol{m} + \frac{1}{1+b}\frac{1}{\sigma_i^2}(\theta_i - m_i)\Sigma_i$ and $S_i^* = \frac{1+b}{2+b}\Big(\Sigma - \frac{1}{(1+b)^2}\frac{1}{\sigma_i^2}\Sigma_i\Sigma_i^t\Big)$.

Thus, by writing $N_{k+1}\big(M_i, S_i\big)$ as a product of these marginal and conditional densities we have

$$t_i^A(\theta_i) = \Big(\frac{b}{b+2}\Big)^{\frac{k}{2}} N_1\Big(\theta_i|\, m_i\,, \frac{1+b}{2+b}\sigma_i^2\Big) \int_A N_k\big(\boldsymbol{\phi}|\, \boldsymbol{m}(\theta_i), S_i^*\big) d\boldsymbol{\phi}. \tag{A16}$$

The covariance function of the GP is derived as follows:

$$Cov\big[f_i(\theta_i), f_i(\phi_i)\,|\, D, \boldsymbol{\alpha}\big] = \int_{\Re^{k-1}} \int_{\Re^{k-1}} Cov\big[f(\boldsymbol{\theta}), f(\boldsymbol{\phi})\,|\, D, \boldsymbol{\alpha}\big] d\boldsymbol{\theta}_{-i}\, d\boldsymbol{\phi}_{-i}, \tag{A17}$$

and from (4) and (22) results in

$$Cov\big[f_i(\theta_i), f_i(\phi_i)\,|D, \boldsymbol{\alpha}\big] = \sigma^2 \int_{\Re^{k-1}} g(\boldsymbol{\theta})g(\boldsymbol{\phi})c(\boldsymbol{\theta},\boldsymbol{\phi})d\boldsymbol{\theta}_{-i}d\boldsymbol{\phi}_{-i} - $$

$$- \sigma^2 \int_{\Re^{k-1}} \boldsymbol{t}(\boldsymbol{\theta})^t W^{-1}\boldsymbol{t}(\boldsymbol{\phi})d\boldsymbol{\theta}_{-i}\, d\boldsymbol{\phi}_{-i}. \tag{A18}$$

Because the prior covariance can be written as $g(\boldsymbol{\theta})g(\boldsymbol{\phi})c(\boldsymbol{\theta},\boldsymbol{\phi})=$ $\left(\frac{b}{b+2}\right)^{\frac{k}{2}} N_{2k}\Big((\boldsymbol{\theta},\boldsymbol{\phi})\,|\,M,S\Big)$, the first integral will be

$$\int_{\Re^{k-1}}\int_{\Re^{k-1}} g(\boldsymbol{\theta})g(\boldsymbol{\phi})c(\boldsymbol{\theta},\phi)d\boldsymbol{\theta}_{-i}d\phi_{-i} = \left(\frac{b}{b+2}\right)^{\frac{k}{2}} \int_{\Re^{k-1}}\int_{\Re^{k-1}} N_{2k}\Big((\boldsymbol{\theta},\boldsymbol{\phi})\,|\,M,S\Big)d\boldsymbol{\theta}_{-i}d\boldsymbol{\phi}_{-i}$$

that is the marginal of $(\theta_i\,,\,\phi_i)$. Therefore,

$$\int_{\Re^{k-1}}\int_{\Re^{k-1}} g(\boldsymbol{\theta})g(\boldsymbol{\phi})c(\boldsymbol{\theta},\boldsymbol{\phi})d\boldsymbol{\theta}_{-i}d\boldsymbol{\phi}_{-i} = \left(\frac{b}{b+2}\right)^{\frac{k}{2}} N_2\Big((\theta_i,\phi_i)\,|\,\boldsymbol{m}_i,\sigma_i^2 V\Big), \tag{A19}$$

where $\boldsymbol{m}_i = (m_i, m_i)$ and $V = \begin{bmatrix} \frac{1+b}{2+b} & \frac{1}{2+b} \\ \frac{1}{2+b} & \frac{1+b}{2+b} \end{bmatrix}$.

Writing this bivariate normal density as the product of marginal density of θ_i and conditional density of ϕ_i given θ_i the resulting integral is

$$\int_{\Re^{k-1}}\int_{\Re^{k-1}} g(\boldsymbol{\theta})g(\boldsymbol{\phi})c(\boldsymbol{\theta},\boldsymbol{\phi})d\boldsymbol{\theta}_{-i}d\boldsymbol{\phi}_{-i} = \left(\frac{b}{b+2}\right)^{\frac{k}{2}} N_1\Big(\theta_i|\, m_i, \frac{b+1}{b+2}\sigma_i^2\Big) N_1\Big($$

$$\phi_i|\frac{1}{b+1}(\theta_i + bm_i)\,,\,\frac{b}{b+1}\sigma_i^2\Big). \tag{A20}$$

Again, because $g(\boldsymbol{\theta})g(\boldsymbol{\phi})c(\boldsymbol{\theta},\boldsymbol{\phi})$ can be written as $\left(\frac{b}{b+2}\right)^{\frac{k}{2}} N_{2k}\Big((\boldsymbol{\theta},\boldsymbol{\phi})\,|\,M,S\Big)$ for $2k$-multivariate normal we can also write the product $g_i(\theta_i)g_i(\phi_i)c_i(\theta_i\,,\phi_i)$ as function of the bivariate normal distribution given in (A18), that is,

$$g_i(\theta_i)g_i(\phi_i)c_i(\theta_i\,,\phi_i) = \left(\frac{b}{b+2}\right)^{\frac{1}{2}} N_2\Big((\theta_i,\phi_i)\,|\,\boldsymbol{m}_i,\sigma_i^2 V\Big). \tag{A21}$$

Indeed, just consider $k = 1$, $\Sigma = \sigma_i^2$ and $\boldsymbol{m} = m_i$ for M and S given in Section 4.4. Therefore, from (A18) and (A20) the first integral in (A17) is

$$\int_{\Re^{k-1}} \int_{\Re^{k-1}} g(\boldsymbol{\theta}) g(\boldsymbol{\phi}) c(\boldsymbol{\theta}, \boldsymbol{\phi}) d\boldsymbol{\theta}_{-i} d\boldsymbol{\phi}_{-i} = \left(\frac{b}{b+2}\right)^{\frac{k-1}{2}} g_i(\theta_i) g_i(\phi_i) c_i(\theta_i\,, \phi_i). \tag{A22}$$

The second integral in (A17) is easily obtained as

$$\int_{\Re^{k-1}} \int_{\Re^{k-1}} \Big[\int_{\Re^{k-1}} \int_{\Re^{k-1}} \boldsymbol{t}(\boldsymbol{\theta})^t W^{-1} \boldsymbol{t}(\boldsymbol{\phi}) d\boldsymbol{\theta}_{-i} \Big] d\boldsymbol{\phi}_{-i} \int_{\Re^{k-1}} \int_{\Re^{k-1}} \boldsymbol{t}_i(\theta_i)^t W^{-1} \boldsymbol{t}(\boldsymbol{\phi}) d\boldsymbol{\phi}_{-i} = \tag{A23}$$

$$= \boldsymbol{t}_i(\theta_i)^t W^{-1} \boldsymbol{t}_i(\phi_i).$$

Finally,

$$Cov\big[f_i(\theta_i)\,, f_i(\phi_i) \,|\, D, \alpha\big] = \sigma^2 \Big[C_i(\theta_i, \phi_i) - \boldsymbol{t}_i(\theta_i)^t W^{-1} \boldsymbol{t}_i(\phi_i)\Big]. \tag{A24}$$

A4. Posterior distribution for expert's distribution functions

Since the analyst's posterior distribution of $f(\cdot)$ is a Gaussian process then the posterior distribution of $F(\boldsymbol{x}) = \int_{-\infty}^{\boldsymbol{x}} f(\boldsymbol{\theta}) d\boldsymbol{\theta}$ is also a Gaussian process with mean function

$$E\big[F(\boldsymbol{x}) \,|\, D, \boldsymbol{\alpha}\big] = \int_{-\infty}^{\boldsymbol{x}} E\big[f(\boldsymbol{\theta}) \,|\, D, \boldsymbol{\alpha}\big] d\boldsymbol{\theta} = \int_{-\infty}^{\boldsymbol{x}} g(\boldsymbol{\theta}) d\boldsymbol{\theta} + \int_{-\infty}^{\boldsymbol{x}} \boldsymbol{t}(\boldsymbol{\theta})^t W^{-1} (D - H) d\boldsymbol{\theta}$$
$$= \Phi_k(\boldsymbol{x} \,|\, \boldsymbol{m}, \Sigma) + \int_{-\infty}^{\boldsymbol{x}} \boldsymbol{t}(\boldsymbol{\theta})^t W^{-1} (D - H) d\boldsymbol{\theta} = \Phi_k(\boldsymbol{x} \,|\, \boldsymbol{m}, \Sigma) + \boldsymbol{t}^*(\boldsymbol{x})^t W^{-1} (D - H) \tag{A25}$$

where the vector $\boldsymbol{t}^*(\boldsymbol{x}) = \int_{-\infty}^{\boldsymbol{x}} \boldsymbol{t}(\boldsymbol{\theta}) d\boldsymbol{\theta}$ has elements obtained from (A12) as

$$\left(\frac{b}{b+2}\right)^{\frac{k}{2}} \int_{-\infty}^{\boldsymbol{x}} \int_A N_{2k}\Big((\boldsymbol{\theta}, \boldsymbol{\phi}) \,|\, M, S\Big) d\boldsymbol{\theta} d\boldsymbol{\phi}. \tag{A26}$$

The covariance function is derived as

$$Cov\big[F(\boldsymbol{x}), F(\boldsymbol{y})|D, \boldsymbol{\alpha}\big] = \int_{-\infty}^{\boldsymbol{x}}\int_{-\infty}^{\boldsymbol{y}} Cov\big[f(\boldsymbol{\theta}), f(\boldsymbol{\phi})|\, D, \boldsymbol{\alpha}\big]d\boldsymbol{\theta}d\boldsymbol{\phi} = \sigma^2\Big[\int_{-\infty}^{\boldsymbol{x}}\int_{-\infty}^{\boldsymbol{y}}$$
$$g(\boldsymbol{\theta})g(\boldsymbol{\phi})c(\boldsymbol{\theta},\boldsymbol{\phi}) - \int_{-\infty}^{\boldsymbol{x}}\int_{-\infty}^{\boldsymbol{y}} \boldsymbol{t}(\boldsymbol{\theta})^t W^{-1}\boldsymbol{t}(\boldsymbol{\phi})\Big]d\boldsymbol{\theta}d\boldsymbol{\phi}. \tag{A27}$$

By using similar arguments to those used in the proof of marginal densities, in Section A3, for both integrals above we arrive at

$$Cov\big[F(\boldsymbol{x})\,,\,F(\boldsymbol{y})\,|\,D, \boldsymbol{\alpha}\big] = \sigma^2\,\Big[\Big(\frac{b}{b+2}\Big)^{\frac{k}{2}}\int_{-\infty}^{\boldsymbol{x}}\int_{-\infty}^{\boldsymbol{y}} N_{2k}\Big((\boldsymbol{\theta},\boldsymbol{\phi})|M,S\Big)d\boldsymbol{\theta}d\boldsymbol{\phi} -$$
$$-\,\boldsymbol{t}^*(\boldsymbol{x})^t W^{-1}\boldsymbol{t}^*(\boldsymbol{y})\Big] = \sigma^2\,\Big[\Big(\frac{b}{b+2}\Big)^{\frac{k}{2}}\Phi_{2k}\Big((\boldsymbol{x},\boldsymbol{y})\,|\,M,S\Big) - \boldsymbol{t}^*(\boldsymbol{x})^t W^{-1}\boldsymbol{t}^*(\boldsymbol{y})\Big]. \tag{A28}$$

A5. Posterior distribution for moments of θ

Since $f_i(\theta_i)$ given D has a Gaussian process distribution and $h(\theta_i)=\theta_i^r$ is a known function of θ_i, then $h(\theta_i)f_i(\theta_i)$ is also a Gaussian process, conditional on $\boldsymbol{\alpha}$. Besides, the integral $\int_{-\infty}^{+\infty} h(\theta_i)f_i(\theta_i)d\theta_i$ is a defined linear functional of $h(\theta_i)f_i(\theta_i)$ and therefore $\mu_i^{(r)}$ is a normally distributed random variable. The analyst's posterior mean of $\mu_i^{(r)}$ is derived as:

$$E\big[\mu_i^{(r)}\,|\,D, \boldsymbol{\alpha}\big] = \int_{-\infty}^{+\infty}\theta_i^r E\big[f_i(\theta_i)\,|\,\boldsymbol{\alpha}\big]d\theta_i, \tag{A29}$$

where the marginal mean function is given in (A14). Thus, (A28) becomes

$$E\big[\mu_i^{(r)}\,|\,D, \boldsymbol{\alpha}\big] = \int_{-\infty}^{+\infty}\theta_i^r g_i(\theta_i)d\theta_i + \int_{-\infty}^{+\infty}\theta_i^r \boldsymbol{t}_i(\theta_i)^t W^{-1}(D-H)d\theta_i =$$
$$= \int_{-\infty}^{+\infty}\theta_i^r g_i(\theta_i)d\theta_i + [\boldsymbol{t}_i^{(r)}]^t W^{-1}(D-H), \tag{A30}$$

where $\boldsymbol{t}_i^{(r)} = \int_{-\infty}^{+\infty} \theta_i^r \boldsymbol{t}_i(\theta_i) d\theta_i$ is a vector.

From equation (A14) the elements of vector $\boldsymbol{t}_i^{(r)}$ will be given by

$$\left(\frac{b}{b+2}\right)^{\frac{k}{2}} \int_{-\infty}^{+\infty} \theta_i^r \Big[\int_A N_{k+1}\big((\theta_i, \boldsymbol{\phi})|\, M_i, S_i\big) d\boldsymbol{\phi}\Big] d\theta_i. \qquad \text{(A31)}$$

The density $N_{k+1}\big((\theta_i, \boldsymbol{\phi})|\, M_i, S_i\big)$ can be written as a product of marginal density of $\boldsymbol{\phi} \sim N_k\big(\boldsymbol{m}, \frac{1+b}{2+b}\Sigma\big)$ and the conditional density $\theta_i \mid \boldsymbol{\phi} \sim N_1\big(m_i^*, \sigma_i^{*2}\big)$ where $m_i^* = m_i + \frac{2+b}{1+b}\Sigma_i^t \Sigma^{-1}(\boldsymbol{\phi} - \boldsymbol{m})$ and $\sigma_i^{*2} = \sigma_i^2 - \frac{2+b}{1+b}\Sigma_i^t \Sigma^{-1}\Sigma_i$. Therefore, (A30) becomes

$$\left(\frac{b}{b+2}\right)^{\frac{k}{2}} \int_A \Big[\int_{-\infty}^{+\infty} \theta_i^r N_1\big(\theta_i|\, m_i^*, \sigma_i^{*2}\big) d\theta_i\Big] N_k\big(\boldsymbol{\phi}|\, \boldsymbol{m}, \frac{1+b}{2+b}\Sigma\big) d\, \boldsymbol{\phi} = \qquad \text{(A32)}$$

where $\psi(\boldsymbol{\phi}) = \int_{-\infty}^{+\infty} \theta_i^r N_1\big(\theta_i|\, m_i^*, \sigma_i^{*2}\big) d\theta_i$.

And finally, the posterior variance of marginal moments is obtained as

$$var\big[\mu_i^{(r)} \mid D, \boldsymbol{\alpha}\big] = Cov\big[\mu_i^{(r)}, \mu_i^{(r)} \mid D, \boldsymbol{\alpha}\big] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \theta_i^r \phi_i^r Cov\big[f_i(\theta_i)\,, f_i(\phi_i) \mid D, \boldsymbol{\alpha}\big] d\theta_i d\phi_i. \qquad \text{(A33)}$$

From eqs. (36) and (38) the variance is

$$var\big[\mu_i^{(r)} \mid D, \boldsymbol{\alpha}\big] = \sigma^2\Big[\Big(\frac{b}{b+2}\Big)^{\frac{k-1}{2}} \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \theta_i^r \phi_i^r g_i(\theta_i) g_i(\phi_i) c_i(\theta_i\,, \phi_i) d\theta_i d\phi_i - - \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \theta_i^r \phi_i^r \boldsymbol{t}_i(\theta_i)^t W^{-1} \boldsymbol{t}_i(\phi_i) d\theta_i d\phi_i\Big]. \qquad \text{(A34)}$$

The first integral in (A33) is calculated as

$$\left(\frac{b}{b+2}\right)^{\frac{k-1}{2}} \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \theta_i^r \phi_i^r g_i(\theta_i) g_i(\phi_i) c_i(\theta_i\,, \phi_i) d\theta_i d\phi_i = \left(\frac{b}{b+2}\right)^{\frac{k}{2}} \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \theta_i^r \phi_i^r N_2(\boldsymbol{m}_i, \sigma_i^2 V)\, d\theta_i d\phi_i \qquad \text{(A35)}$$

and the second integral is given by

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\theta_i^r\phi_i^r\big[\boldsymbol{t}_i(\theta_i)\big]^tW^{-1}\boldsymbol{t}_i(\phi_i)d\theta_i d\phi_i=\int_{-\infty}^{+\infty}\phi_i^r\Big[\int_{-\infty}^{+\infty}\theta_i^r\big[\boldsymbol{t}_i(\theta_i)\big]^t d\theta_i\Big]W^{-1}$$
$$\boldsymbol{t}_i(\phi_i)d\phi_i=\big[\boldsymbol{t}_i^{(r)}\big]^tW^{-1}\Big[\int_{-\infty}^{+\infty}\phi_i^r\boldsymbol{t}_i(\phi_i)d\phi_i\Big]=\big[\boldsymbol{t}_i^{(r)}\big]^tW^{-1}\boldsymbol{t}_i^{(r)}. \tag{A36}$$

Therefore,

$$var\big[\mu_i^{(r)}\,|\,D,\alpha\big]=\sigma^2\Big[\Big(\frac{b}{b+2}\Big)^{\frac{k}{2}}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\theta_i^r\phi_i^r N_2(\boldsymbol{m}_i,\sigma_i^2V)\,d\theta_i d\phi_i-\big[\boldsymbol{t}_i^{(r)}\big]^tW^{-1}\boldsymbol{t}_i^{(r)}\Big]. \tag{A37}$$