Summary

When a computer model is used to inform a decision, it is important to investigate any uncertainty in the model and determine how that uncertainty may impact on the decision. In Probabilistic Sensitivity Analysis, model users can investigate how various uncertain model inputs contribute to the uncertainty in the model output. However, much of the literature only focusses on output uncertainty as measured by variance; the decision problem itself is often ignored, even though uncertainty as measured by variance may not equate to uncertainty about the optimum decision. Consequently, traditional variance-based measures of input parameter importance may not correctly describe the importance of each input. We review a decision theoretic framework for conducting sensitivity analysis which addresses this problem. It is noted that computation of these decision-theoretic measures can be impractical for complex computer models, and so efficient computational tools using Gaussian process are also presented. An illustration is given in the field of medical decision making, and the Gaussian process approach is compared with conventional Monte Carlo sampling.

KEY WORDS: Bayesian quadrature, computer experiment, Gaussian process, expected value of perfect information.

*Jeremy Oakley is lecturer at the Department of Probability and Statistics, University of Sheffield, The Hicks Building, Hounsfield Road, Sheffield, S3 7RH, UK. Email: j.oakley@sheffield.ac.uk
1 Introduction

Complex computer models of physical processes are used in many scientific fields, and model users are increasingly appreciating the importance of investigating uncertainty in model predictions. One important source of model output uncertainty is uncertainty in the model inputs. Denoting the computer model by a function \( y = \eta(x) \), where \( y \) is the output, and \( x = (x_1, \ldots, x_r) \) is the vector of input parameters, we suppose that the model user wishes to know the output \( Y = \eta(X) \) for an input vector \( X = (X_1, \ldots, X_r) \) corresponding to some particular situation of interest, but that there is uncertainty about the true value of \( X \). The model user may then wish to establish how individual elements of \( X \) contribute to the uncertainty in \( Y \), so that the important input parameters in the model can then be identified. This requires a ‘global sensitivity analysis’ to be conducted, which involves investigating how the output \( y \) varies as an input \( x_i \) varies across some range.

In the computer experiments literature, the most common approach for conducting a global sensitivity analysis involves the use of variance-based measures of importance (see for example McKay, 1997; Chan et al., 2000; Saltelli and Tarantola, 2002; Oakley and O’Hagan, 2004). Variance-based measures are concerned with scalar outputs (or individual elements within vector outputs) and describe what proportion of the variance of \( Y \) can be attributed to an uncertain input variable \( X_i \) (or subsets of \( \{X_1, \ldots, X_r\} \)). Whilst a variance-based sensitivity analysis will clearly be of interest to the model user, it typically fails to take into account the ultimate purpose of the model. It is almost always the case that the model will have been developed in order to help the user make a decision. It then follows that the aim of a sensitivity analysis should be to investigate how uncertainty in the inputs affects the user’s optimal decision, and corresponding loss/utility. From this point on we refer to the model user as the decision maker.

In the field of health economics, global sensitivity analysis is conducted within the framework of decision theory, based on the concept of the expected value of perfect information. We review this framework for sensitivity analysis (which includes variance-based measures as a special case). Computationally, decision-theoretic sensitivity analysis can be very demanding and Monte Carlo approaches can be infeasible for complex models. We present methods for efficient computation using Gaussian process emulators, a popular technique for dealing with complex computer models.

Another deficiency in the practical use of existing global sensitivity analysis
methods is the implicit assumption that the possibility exists of learning the parameter precisely; variance-based measures give the expected reduction in variance if a parameter were known with certainty. More realistically, a model user may only have the option of reducing their uncertainty about an unknown input though obtaining new data. The same decision-theoretic framework can be employed to consider the expected value of sample information.

In the next section we introduce an example involving an economic model of different treatment strategies for gastroesophageal reflux disease (GERD). In section 3 we review the framework for decision-theoretic sensitivity analysis, and discuss the value of variance-based measures from this perspective. Computational methods for complex models using Gaussian processes are presented in section 4 and these are demonstrated on the GERD model in section 5.

2 Example: GERD model

For illustration, we will consider an example from the field of health economics. The decision problem typically in question is to choose one treatment from a range of alternatives on the basis of which will be the most cost-effective for the patient population. In assessing the cost-effectiveness of a proposed treatment, a common approach is to construct an economic model of the treatment process. The health status and resources used by a patient will be modelled over some period of time following treatment. Model input parameters will be related to the effectiveness of the treatment and resources used by patients. Given a set of values for all the input parameters, the model will state an overall measure of the efficacy of the treatment and the mean cost of treatment, perhaps incorporating costs such as additional hospital visits depending on the success of the treatment etc. These two outputs can then be combined into a single measure of cost-effectiveness, known as the net-benefit of the treatment, with the net-benefit interpreted as the decision-maker’s utility for that treatment. Uncertainty about the true values of the input parameters will induce uncertainty about the true cost-effectiveness of each treatment.

The example model used here is an economic model of the treatment process of patients with gastroesophageal reflux disease and was presented in Goeree et al. (1999). The model was designed to compare treatment costs and outcomes of various different drug treatment strategies for patients with the disease over a one year period. In this example, we suppose that a decision has to be made regarding the
adoption of one of three treatment strategies:

1. Acute treatment with proton pump inhibitors (PPIs) for 8 weeks, then continuous maintenance treatment with PPIs at the same dose.

2. Acute treatment with PPIs for 8 weeks, then continuous maintenance treatment with hydrogen receptor antagonists (H2RAs).

3. Acute treatment with proton pump inhibitors PPIs for 8 weeks, then continuous maintenance treatment with PPIs at a lower dose.

In the scenario that we are considering, there are twenty-three uncertain inputs, relating to quantities such as the proportion of patients that will experience a recurrence of the symptoms with each treatment, and resources used by patients such as mean number of visits to a general practitioner. We denote the true values of these inputs by $X$, with distribution $G$. The distributions of all these uncertain inputs are given in Briggs et al. (2002).

This model is not computationally expensive, and so sensitivity analysis can be conducted using Monte Carlo methods and compared with the computational methods presented in this paper. Examples of genuinely computationally expensive models within health economic modelling, where Monte Carlo sensitivity analysis is not feasible, can be found in Stevenson et al. (2004) and Tappenden et al. (2004).

3 Sensitivity analysis using value of information

We now review some standard results from decision theory and explain how they can be used to conduct global sensitivity analysis. The key concept is the notion of the expected value of perfect information (EVPI), which is discussed, for example, in Bernardo and Smith (1994).

We suppose that the decision-maker has to choose a decision $d$ from a set of known decision options $D$, and that his or her utility for decision $d$ will depend on the value of the model output $Y = \eta(X)$ where $X = \{X_1, \ldots, X_r\}$ are the true values of the unknown inputs. In the GERD example, $D$ is the set of the three available treatment strategies, and $Y$ is vector of the net-benefits of the three treatments.

Writing the decision-maker’s utility function as $U\{d, \eta(X)\}$, the expected utility given no new information is

$$U^* = \max_{d \in D} E_X[U\{d, \eta(X)\}].$$

(1)
Now suppose the decision-maker wishes to consider the importance of an uncertain input parameter $X_i$. Ignoring any costs for learning the value of $X_i$, the expected value to the decision-maker of learning $X_i$ before making a decision is given by

$$E_{X_i} \left[ \max_{d \in D} E_{X|X_i} [U\{d, \eta(X)\}] \right] - U^*.$$  

(2)

In the health economics literature, the expression in (2) is known as the partial EVPI for the input parameter $X_i$ and has been advocated as a measure of parameter importance by Felli and Hazen (1998) and Claxton (1999). In global sensitivity analysis, we can see how the partial EVPI relates the importance of an uncertain input parameter $X_i$ directly to the decision problem.

### 3.1 Variance-based sensitivity analysis as a special case of EVPI

In variance-based sensitivity analysis, for scalar $Y$, the importance of an input parameter is given by the term

$$Var_{X_i}\{E(Y|X_i)\},$$  

(3)

(though this is sometimes normalised by the variance of $Y$). Since

$$Var_{X_i}\{E(Y|X_i)\} = Var(Y) - E_{X_i}\{Var(Y|X_i)\},$$

we can see that if the decision problem were to estimate $Y$, and the decision-maker had a quadratic loss function for the size of the estimation error, i.e. $U\{d, \eta(X)\} = -(\eta(X) - d)^2$ for $d \in D = R^1$, the variance-based measure given in (3) is equal to the partial EVPI for $X_i$. Hence although variance-based measures can be seen as a useful exploratory tool, they are not necessarily suitable for all practical decision problems. A quadratic loss function may not be appropriate, so that a variance-based approach will not be quantifying input importance in the correct manner. We illustrate this with two simple examples.

A common scenario for a decision-maker is one in which the uncertain event of interest is whether or not a variable $Y$ lies above or below a particular threshold. Suppose for example that $D = \{d_1, d_2\}$, with $d_1$ preferable to $d_2$ if $Y > 0$. Now consider the two alternative density functions for $Y$ plotted in figure 1. Uncertainty about $Y$ as measured by variance is appreciably greater in the second plot, yet uncertainty about the best decision is greater in the first.
Now suppose, under the same decision scenario, we have \( Y = X_1 X_2 \), with \( X_1 \sim N(1, 1) \) and \( \log X_2 \sim N(0, 1) \). A variance-based sensitivity analysis would give us \( \text{Var}_X \{ E(Y|X_1) \} = 2.7 \), and \( \text{Var}_X \{ E(Y|X_2) \} = 4.6 \), indicating \( X_2 \) is ‘more important’ than \( X_1 \). However, for the purpose of establishing if \( Y > 0 \), the variable \( X_1 \) is clearly more important than \( X_2 \), as the event \( Y > 0 \) is determined by the sign of \( X_1 \) only.

![Figure 1: Two alternative density functions for an uncertain variable \( Y \). Uncertainty as described by variance does not equate decision uncertainty if the optimal decision depends on the event \( \{ Y > 0 \} \).](image)

3.2 Expected value of sample information

A criticism regarding interpretation of partial EVPIs is that the decision-maker is unlikely to be able to eliminate uncertainty about any input parameter by obtaining ‘perfect information’. A more realistic scenario is one in which the decision-maker has the possibility of reducing uncertainty about input parameters by obtaining more data. The data may be informative about a single input parameter \( X_i \) or any subset of \( \{ X_1, \ldots, X_r \} \). In the GERD example, a new trial might be conducted to obtain more data on the efficacy of a particular treatment strategy. The same decision-theoretic argument can be used to quantify the value of this new data.
Denoting the future set of data by $Z$, the value of collecting this data is given by

$$E_Z \left[ \max_{d \in D} E_{X|Z} [U\{d, \eta(X)\}] \right] - U^*,$$

which is known as the expected value of sample information (EVSI).

### 3.3 Computation using Monte Carlo

Partial EVPI’s can be estimated using Monte Carlo integration. The difficulty arises with the first term in (2), which we write as

$$\int m(X_i) dG(X_i),$$

with $G(X_i)$ the distribution of $X_i$ and

$$m(X_i) = \max_{d \in D} E_{X|X_i} \{U(d, \eta(X))\}.$$

The integral (4) is one-dimensional and so numerical integration can be used. However, evaluating the integrand $m(X_i)$ requires Monte Carlo integration, but in this case, a large sample size may be needed to estimate reliably the maximised expected utility. Given a sample $x_1, \ldots, x_N$ from the conditional distribution of $X|X_i$, we would estimate $m(X_i)$ by

$$\hat{m}(X_i) = \max_{d \in D} \left[ \frac{1}{N} \sum_{j=1}^{N} U\{d, \eta(x_j)\} \right],$$

and although $\frac{1}{N} \sum_{j=1}^{N} U\{d, \eta(x_j)\}$ is an unbiased estimate of $E_{X|X_i} \{U(d, \eta(X))\}$, it can be seen that $\hat{m}(X_i)$ is a biased estimate of $m(X_i)$ with $E\{\hat{m}(X_i)\} \geq m(X_i)$, as the expectation of the maximum is greater than (or equal to) the maximum of the expectations. To achieve negligible bias, $N$ may need to be large. This can then result in $\eta(x)$ being evaluated a substantial number of times at different values of $x$. [REF - BLINDED] present a means of estimating a value of $N$ to achieve an acceptably small bias using simple Monte Carlo sampling, and in the case study presented found this to be of the order of 1000. Efficiency gains can be made using Latin hypercube sampling (McKay, 1997), though large numbers of model runs may still be required. Similar difficulties are faced when computing EVSIs, and the problem can be exacerbated when the density $G(x)$ is not conjugate to the likelihood function $f(Z|X)$ for the future sample of data $Z$. 

7
4 Gaussian process emulators

For computationally expensive models, a substantial saving in computational effort can be made through the use of a Gaussian process emulator. An emulator is a statistical representation of a computer model which can then be used to provide a fast surrogate. The emulator approach was proposed by Sacks et al. (1989), and examples of its use can be found in Currin et al. (1991), Kennedy and O’Hagan (2001), Oakley and O’Hagan (2002) and Kennedy et al. (2006). We outline the main results here, but for a detailed treatment see Santner et al. (2003) and Rasmussen and Williams (2006).

In the Gaussian process emulator, for any collection of input values \{x_1, \ldots, x_n\}, we suppose that the corresponding set of outputs \{\eta(x_1), \ldots, \eta(x_n)\} have a multivariate normal distribution. The Gaussian process is characterised by the mean of \eta(x) and the covariance function for \eta(x) and \eta(x').

The mean of \eta(x) is given by

$$E\{\eta(x)|\beta\} = h(x)^T \beta,$$

conditional on \beta. The vector \(h(\cdot)\) consists of \(q\) known regression functions of \(x\), and \(\beta\) is a vector of coefficients. The choice of \(h(\cdot)\) is arbitrary, though it should be chosen to incorporate any beliefs we might have about the form of \(\eta(\cdot)\). The covariance between \(\eta(x)\) and \(\eta(x')\) is given by

$$Cov\{\eta(x), \eta(x')|\sigma^2\} = \sigma^2 c(x, x'),$$

conditional on \(\sigma^2\), where \(c(x, x')\) is a function which decreases as |\(x - x'\)| increases, and also satisfies \(c(x, x) = 1 \forall x\). The function \(c(\ldots)\) must ensure that the covariance matrix of any set of outputs \{\(y_1 = \eta(x_1), \ldots, y_n = \eta(x_n)\)\} is positive semi definite. A typical choice is

$$c(x, x') = \exp\{-(x - x')^T B (x - x')\}, \quad (6)$$

where \(B = \text{diag}\{b_i\}\) a diagonal matrix of (positive) roughness parameters (note that various authors use different names and parameterisations, e.g. \(1/\sqrt{b_i}\) is defined as the range in Cressie (1993), as the correlation length in Santner et al. (2003) and as the characteristic length-scale in Rasmussen and Williams (2006)). Conventionally, a weak prior of \(\beta\) and \(\sigma^2\) in the form \(p(\beta, \sigma^2) \propto \sigma^{-2}\) is used. In Oakley (2002) a means of including proper prior information about the function \(\eta(\cdot)\) is presented,
through the use of the conjugate prior, the normal inverse gamma distribution. We have

$$p(\beta, \sigma^2) \propto (\sigma^2)^{-\frac{1}{2}(v+q+2)} \exp\left[-\left\{ (\beta - z)^T V^{-1} (\beta - z) + a \right\} / (2\sigma^2) \right].$$

(7)

(Recall that $q$ is the number of regressors in the mean function).

It is fairly common practice to ignore uncertainty about $B$, and simply condition on a point estimate once the code $\eta(.)$ has been run, for example the maximum likelihood estimate. Here, we also allow for uncertainty in $B$, and consider two different prior distributions for each diagonal element of $B$ (with the diagonal elements independent): an exponential distribution with mean 1, and a $U[0, 100]$ distribution, where each input parameter has been scaled to lie in $[0, 1]$. For smooth functions we expect the elements of $B$ to be small, so that the exponential distribution is a better representation of prior beliefs than the uniform, but we also consider the uniform as a means of checking for robustness.

4.1 The posterior Gaussian process

The output of $\eta(.)$ is observed at $n$ design points, $x_1, \ldots, x_n$ to obtain data $y$. We opt for a space-filling design using the maximin Latin hypercube approach suggested in Mitchell and Morris (1995). For the purpose of generating the design, we assume a uniform distribution for each input, so that the design points are spread evenly throughout the input region of interest. The bounds are chosen fairly conservatively based on each input’s original distribution: we use $E(X_i) - 3.5\sqrt{Var(X_i)}$ and $E(X_i) + 3.5\sqrt{Var(X_i)}$. This is intended to give suitable information about $\beta$, and reduce the need to extrapolate when predicting $\eta(.)$ at untested inputs.

Given the prior in (7) it can be shown that

$$\frac{\eta(x) - m^*(x)}{\hat{\sigma}\sqrt{c^*(x, x)}} | y, B \sim t_{v+n},$$

(8)

where

$$m^*(x) = h(x)^T \hat{\beta} + t(x)^T A^{-1}(y - H\hat{\beta}),$$

(9)

$$c^*(x, x') = c(x, x') - t(x)^T A^{-1} t(x') + (h(x)^T - t(x)^T A^{-1} H)(H^T A^{-1} H)^{-1}(h(x')^T - t(x')^T A^{-1} H)^T.$$  

(10)

$$t(x)^T = (c(x, x_1), \ldots, c(x, x_n)),$$

(11)

$$H^T = (h(x_1), \ldots, h(x_n)).$$
\[
A = \begin{pmatrix}
1 & c(x_1, x_2) & \cdots & c(x_1, x_n) \\
c(x_2, x_1) & 1 & \ddots & \\
\vdots & \ddots & \ddots & 1 \\
c(x_n, x_1) & \cdots & c(x_n, x_{n-1}) & 1
\end{pmatrix},
\]

\[
\hat{\beta} = V^*(V^{-1}z + H^TA^{-1}y),
\]

\[
\hat{\sigma}^2 = \frac{\{a + z^TV^{-1}z + y^TA^{-1}y - \hat{\beta}^T(V^*)^{-1}\hat{\beta}\}}{(n + v)},
\]

\[
V^* = (V^{-1} + H^TA^{-1}H)^{-1}
\]

\[
y^T = (\eta(x_1), \ldots, \eta(x_n)).
\]

It is not possible to remove analytically the conditioning on \(B\), and we choose to deal with uncertainty in \(B\) by using MCMC to obtain a sample from the posterior distribution of \(B|y\).

### 4.2 Emulators for vector-output computer models

In the GERD example, the computer model produces a vector \(\eta(x) = \{\eta_1(x), \eta_2(x), \eta_3(x)\}^T\) of three outputs corresponding to the net benefits of the three treatment strategies. As there may be correlations between the elements of \(\eta(x)\), we could consider a multivariate version of the Gaussian process emulator, as in Conti and O’Hagan (2007):

\[
\eta(\cdot)|\Sigma, R \sim N_3(m(\cdot), c(\cdot, \cdot)\Sigma),
\]

with \(c(\cdot, \cdot)\) defined as in (6), and \(\Sigma\) the covariance matrix of the outputs at any single input vector \(x\). Note this gives a separable covariance structure

\[
Cov\{\eta_i(x), \eta_j(x')\} = \exp\{- (x - x')^T B(x - x')\}\Sigma_{ij},
\]

which forces each output to have the same matrix \(B\) of roughness parameters. More flexible, non-separable covariance structures are used in the geostatistics literature (see, for example, Ver Hoef and Barry, 1998; Gelfand et al., 2004), but with a considerably smaller input dimension (typically with two inputs; in the GERD model we have twenty-three).

For simplicity, we choose to emulate the three outputs independently here. Although we are potentially ignoring correlations between outputs, this allows us to have different roughness matrices \(B\) for each output. A further difficulty in using the covariance structure in (12) for the GERD model is that not all inputs are common to each output.
4.3 Evaluating partial EVPIs with Gaussian processes and Bayesian quadrature

The exact manner in which the Gaussian process emulator can be used to speed up computation will depend on the nature of the decision problem at hand. At the simplest level, one could use the posterior mean $m^*(x)$ as a fast surrogate for $\eta(x)$, and then proceed using Monte Carlo. Here we consider a particular class of decision problem, in which a decision-maker has to choose between a small set of possible actions, and show how the emulator can be used to estimate partial EVPIs particularly efficiently using Bayesian quadrature.

We now suppose that the set $D$ is finite and contains $s$ possible decisions, where $s$ is small. To link the output of the computer code to the decision-maker’s utility we suppose further that the computer code $\eta(x)$ returns a vector of outputs $\{\eta_1(x), \ldots, \eta_s(x)\}$, with

$$\eta_d(x) = U\{d, \eta(x)\},$$

for $d = 1, \ldots, s$, so in the GERD model we have $s = 3$, with $\eta_d(X)$ the true, unknown population mean net-benefit or utility of treatment strategy $d$.

We have a separate Gaussian process emulator for each output $\eta_i(x)$, and define $y_i, \beta_i, \sigma^2_i, H_i$ etc as in section 4, now corresponding to the individual output $\eta_i(x)$. In this section, we present all results conditional on $B_i$, for $i = 1, \ldots, s$. We will return to the issue of uncertainty in each $B_i$ in section 4.5.

Consider first evaluating the expected utility of each decision. In our framework we are also treating the function $\eta_d$ as unknown as well as $X$, and so when determining the expected utility of a decision, the expectation must be taken with respect to both $X$ and $\eta_d$:

$$E_{\eta_d}[E_X[U\{d, \eta(X)\}]|y_d, B_d] = E_{\eta_d}[E_X\{\eta_d(X)\}|y_d, B_d].$$

So when the function $\eta_d$ is also unknown, the partial EVPI of $X_i$ is given by

$$E_{X_i} \left[ \max_{d \in D} E_{\eta_d}[E_X\{\eta_d(X)\}|y_d, B_d] \right] - \max_{d \in D} E_{\eta_d}[E_X\{\eta_d(X)\}|y_d, B_d]. \quad (13)$$

Considering first the second term in 13, we need to find $E_{\eta_d}[E_X\{\eta_d(X)\}|B_d, y_d]$ under the Gaussian process model for $\eta_d(.)$. If $\eta_d(.)$ is a Gaussian process, the integral $E_X\{\eta_d(X)\}$ has a normal distribution. This result is central to Bayesian quadrature, described in O’Hagan (1991), and so we use Bayesian quadrature to further speed up the computation of partial EVPIs. It is straightforward to show...
that

\[ E_{\eta_d}[E_X\{\eta_d(X)\}|y_d, B_d] = R_d\hat{\beta}_d + T_dA_d^{-1}(y_d - H_d\hat{\beta}_d), \]  

(14)

with

\[ R_d = \int h_d(x)^T dG(x), \]  

(15)

\[ T_d = \int t_d(x)^T dG(x), \]  

(16)

and \( G(x) \) the distribution of \( X \). For normally distributed inputs and certain choices of \( h_d(\cdot) \) and \( c_d(\cdot, \cdot) \), these integrals can be evaluated analytically. If this is not possible then numerical or Monte Carlo integration can be used, without severe computational effort. (For independent inputs and standard choices of \( c_d(x, x') \) and \( h_d(x) \) these can typically be written as the product of \( d \) one-dimensional integrals).

To compute the partial EVPI, we also need to be able to estimate the term \( E_{\eta_d}[E_X\{\eta_d(X)\}|y_d, B_d] \). We can again use (14) with a slight modification:

\[ E_{\eta_d}[E_X\{\eta_d(X)\}|y_d, B_d] = R_{d,i}\hat{\beta}_{d,i} + T_{d,i}A_{d,i}^{-1}(y_d - H_d\hat{\beta}_d), \]  

(17)

with

\[ R_{d,i} = \int h_d(x)^T dG(x|X_i), \]  

(18)

\[ T_{d,i} = \int t_d(x)^T dG(x|X_i), \]  

(19)

with \( G(x|X_i) \) the conditional distribution of \( X \) given \( X_i \). Note that once (15) and (16) have been evaluated, then for independent inputs we can evaluate (18) and (19) almost instantaneously. A typical choice for \( h_d(x) \) is \( h_d(x) = (1, x_1, \ldots, x_r) \). We can then derive \( R_{d,i} \) from \( R_d \) by replacing the appropriate element of \( R_d \) by \( X_i \). If we denote \( x_j \) to be the \( j \)th design point at which we have run the model, then the \( j \)th element of \( T_d \) is given by

\[ \int \exp\{- (x - x_j)^T B_d(x - x_j)\} dG(x) = \prod_{i=1}^r \int \exp\{- b_{d,i}(x_i - x_{j,i})^2\} dG(x_i), \]

where \( b_{d,i} \) is the \( i \)th element on the diagonal of \( B_d \). To derive \( T_{d,i} \) from \( T_d \), we simply replace the appropriate integral in the \( j \)th element of \( T_d \) by \( \exp\{- b_{d,i}(X_i - x_{j,i})^2\} \).

We have detailed how we compute

\[ \max_{d \in D} E_{\eta_d}[E_X\{\eta_d(X)\}|y_d, B_d]. \]  

(20)

Finally, we still need to take the expectation of (20) with respect to \( X_i \) to get the partial EVPI. Since this expectation is a one-dimensional integral, we evaluate it numerically using Simpson’s rule.
4.4 Estimating the expected value of sample information

In principle, extensions to EVSI calculations are straightforward. We are now concerned with estimating

$$E_Z[\max_{d \in D} E_{\eta_d} \{E_X|Z[U\{d, \eta(X)\}]|y_d, B_d]\}],$$

where $Z$ consists of sample data of a particular size. Given $Z$, the same procedure as in section 4.3 can be employed, modifying equation (17) to give

$$E_{\eta_d}[E_X|Z\{\eta_d(X)\}|y_d, B_d] = R_d(Z)\hat{\beta}_d + T_d(Z)A^{-1}_d(y_d - H_d\hat{\beta}_d),$$

with

$$R_d(Z) = \int h_d(x)^T dG(x|Z)$$
$$T_d(Z) = \int t_d(x)^T dG(x|Z),$$

with $G(x|Z)$ the distribution of $X|Z$. The ease of this procedure will depend on the form of the distribution $G(X)$ and the likelihood function $f(Z|X)$. In the simplest case, the data $Z$ may inform us about a single uncertain parameter $X_i$, with $G(X_i)$ conjugate to $f(Z|X_i)$. If, alternatively, updating from $G(X)$ to $G(X|Z)$ requires MCMC, evaluating an EVSI may be computationally very intensive regardless of the complexity of $\eta(x)$; it may be necessary to obtain MCMC samples many times for different values of $Z$, but we do not consider this further in this paper.

4.5 Quantifying uncertainty in partial EVPIs and EVSI

Technically, it is not appropriate to discuss uncertainty in a partial EVPI or EVSI, as a utility of any status cannot be an uncertain quantity. However, there are two issues to consider regarding uncertainty about $\eta(.)$ and its relationship with the value of any partial EVPI or EVSI. Firstly, the various expectations (with respect to $\eta(.)$) considered in the previous two sections are all presented conditional on $B_d$. However, as $B_d$ is unknown, we require the marginal distribution of $\eta_d(.)$ unconditional on $B_d$, and so there is the problem of obtaining this marginal distribution. Secondly, the value of any partial EVPI or EVSI will depend on the decision-maker’s current beliefs about $\eta(.)$. The decision-maker is uncertain about $\eta(.)$ because the model is computationally expensive and has only been run at a small number of inputs. Hence it is important to establish whether further runs of the model are likely to
result in different utilities for learning any input (or collecting more data about the inputs). Here, we consider a simulation approach that addresses these two issues simultaneously.

Suppose first that \( \eta(\cdot) \) is a known function. We write the partial EVPI as

\[
E_X \left[ \max_{d \in D} E_{X_{-i}|X_i} \{ \eta_d(X)|\eta(.) \} \right] - \max_{d \in D} [E_X \{ \eta_d(X)|\eta(.) \}].
\] (21)

Simulation can be used to explore how the partial EVPI varies as \( \eta(.) \) varies, following a variant on a procedure described in Oakley and O’Hagan (2002). We first obtain a sample \( B_{d,1}, \ldots, B_{d,N} \) from the posterior distribution of \( B_{d}|y_d \), using MCMC sampling. We then choose additional simulation design points \( x'_{1}, \ldots, x'_{n'} \) and generate new data, denoted by \( y_{d,j} \) from the joint posterior distribution of \( \eta_d(x'_1), \ldots, \eta_d(x'_{n'})|y_d, B_{d,j} \), for \( j = 1, \ldots, N \). This is a multivariate \( t \) distribution, with mean vector and covariance matrix obtained from (9) and (10) respectively.

The intention is that conditional on both \( y_d \) and \( y_{d,j} \), uncertainty about \( \eta_d(.) \) is sufficiently small such that \( E\{ \eta_d(.)|y_d, y_{d,j}, B_{d,j} \} \) can be treated as an approximate draw from the posterior distribution of \( \eta_d(.)|y_d \). Suitable simulation design points are chosen by first generating a large Latin hypercube sample of candidate inputs, then by sequentially selecting \( x'_1 \) to be the candidate input with the largest variance, \( x'_2 \) to be the candidate input with the largest variance conditional on both \( y_d \) and \( \eta_d(x'_1) \), and so on.

We now evaluate the \( j \)-th sampled partial EVPI value as

\[
E_X \left[ \max_{d \in D} E_{\eta_d} \{ E_{X_{-i}|X_i} \{ \eta_d(X)|y_d, y_{d,j}, B_{d,j} \} \} \right] - \max_{d \in D} E_{\eta_d} [E_X \{ \eta_d(X)|y_d, y_{d,j}, B_{d,j} \}].
\]

This gives us a sample from the distribution of (21), unconditional on any \( B_{d,i} \), thought of as a function of the uncertain \( \eta(.) \).

There are two practical problems with this method. Firstly, it may not be possible to achieve negligible uncertainty about \( \eta_d(.) \) given both \( y_d \) and \( y_{d,j} \), particularly if the input dimension is large, but we can check convergence of partial EVPI estimates to increasing \( n' \). For the GERD example, we try \( n' = 100, 200, 300 \), and observe only minor changes in posterior inferences. Secondly, we are now required to repeatedly evaluate integrals such as (18) and (19). In the GERD example, we are able to transform the each input so that it is normally distributed, but this will not always be a practical option, and so our method is limited to fairly simple choices of input distributions.
5 GERD model results

The model outputs a net benefit for each treatment, and this net benefit is assumed to be equal to the utility, so that we have $\eta(X) = \{\eta_1(X), \eta_2(X), \eta_3(X)\}$ (although not all elements of $X$ are used to evaluate each $\eta_i(X)$). We construct three separate Gaussian process emulators to represent each function. We choose 200 sets of inputs for each function to cover the sample space as described by the input distributions, based on Latin hypercube samples. After evaluating each function at the 200 input configurations, we fit a Gaussian process emulator to each function, and use MCMC to obtain 20,000 samples from the posterior distribution of $f(B_d|y_d)$. We then sample from the posterior distribution of each partial EVPI. Only 6 of the 23 inputs have non-negligible partial EVPIs, and means and 95% credible intervals for these are reported in table 1, using both the uniform and exponential priors for $B$. We note that these two priors result in very similar intervals.

We now consider validating our estimates. The GERD model is computationally cheap, i.e., given any particular set of values for the input parameters, we can obtain net benefits of each treatment strategy almost instantaneously. We can therefore determine the true partial EVPIs almost exactly, based on massive Monte Carlo samples (several hundred million in this case). These are also reported in table 1. We see that these are all contained within the 95% credible intervals.

For comparison, we also estimate the partial EVPIs of each input variable using a combination of Simpson’s rule and Latin hypercube sampling. Firstly, the expected utility (equation (1)) given no further information is estimated, using Latin hypercube samples of size 200 for each of the three functions. Then, for each uncertain input in the model, the integral in (4) is estimated using Simpson’s rule at 11 values of $X_i$, and the estimator of $m(X_i)$ in (5) is estimated using a Latin hypercube sample of 50 points. This will require 33550 runs of the GERD model. We repeat this estimation procedure 1000 times, and give the mean and 95% intervals for each of the 6 non-negligible partial EVPIs. We note that the intervals are wider than the credible intervals obtained using the GP emulators. With regard to simply estimating (1), the 95% intervals using Latin Hypercube sampling (not reported here) are actually smaller than those obtained with the emulator. The efficiency gain with the emulator is only achieved when conducted the EVPI analysis, as no further runs of the model are required.

Each partial EVPI in this case can be interpreted as the value (in Canadian dollars) per patient of learning the true value of the input parameter before deciding...
which treatment to use for the patient population. Consider parameter $X_{19}$, which gives the probability of recurrence of symptoms within the first 6 months for a patient on H2RAs. This has the highest partial EVPI, estimated at 21 dollars per patient, and so is considered to be the ‘most important’ input in the sensitivity analysis. If, for example, the patient population to be treated numbered 100,000 patients, the decision-maker would deduce that it is only worthwhile eliminating uncertainty about this input parameter if it can be done for a cost of less than 2.1 million dollars. Parameters with negligible partial EVPIs are considered ‘unimportant’, in the sense that decision-maker can be confident that eliminating/reducing uncertainty about these parameters is not likely to change the decision as to the optimum treatment strategy.

5.1 EVSI computation

As we have commented before, it is unlikely that we would be able to eliminate the uncertainty about any parameter, but the uncertainty could be reduced by collecting more data. Continuing with the study of parameter $X_{19}$, we now consider a future trial of size $m$ patients on H2RAs, and imagine that after 6 months, we will have observed $Z$ out the $m$ patients experience a recurrence of symptoms, with $Z|X_{19} \sim \text{Binomial}(m, X_{19})$. The predictive distribution of $Z$ can be obtained analytically, and so, conditional on a sampled value of $B_{d,j}, y_{d,j}$ for $d = 1, 2, 3$, we compute

$$E_Z \left[ \max_d E_{\eta_d} \left[ E_X | Z = j \left\{ \eta_d(X) \right\} | y_{d,j}, B_{d,j} \right] \right] = \sum_{k=0}^{m} \max_d E_{\eta_d} \left[ E_X | Z = j \left\{ \eta_d(X) \right\} | y_{d,j}, B_{d,j} \right] P(Z = k)$$

For various values of $m$, means and 95% credible intervals of the EVSI are reported in table 2 (we do not consider any costs here of collecting the data). True values are again obtained using very large Monte Carlo samples.

For comparison, estimates with confidence intervals (based on 1000 replications) obtained using a combination of Simpson’s rule and Latin hypercube sampling are also reported in table 2, based on a total of 12120 runs (a LHS of size 20 is generated for each possible value of $Z$, for each sample size $m$). We observe that in most cases, the confidence intervals are wider than the corresponding credible intervals. The EVSI values have all been overestimated using the Gaussian process, though we would expect strong correlations between estimates at different values of $m$. 

16
### Table 1: True values, Gaussian process estimates and Simpson/Monte Carlo estimates of the partial EVPIs of the six most influential input variables. The Gaussian process estimates are based on 600 model runs, and the Simpson/LHS estimates are based on 37950 model runs.

<table>
<thead>
<tr>
<th>uncertain input parameter</th>
<th>true partial EVPI</th>
<th>GP, Uniform prior</th>
<th>GP, Exp prior</th>
<th>Simpson/LHS</th>
</tr>
</thead>
<tbody>
<tr>
<td>hazard for healing on PPIs</td>
<td>1.24 (0.93,1.45)</td>
<td>1.18 (0.92,1.43)</td>
<td>1.17 (0.90,2.70)</td>
<td>1.86 (1.36,3.43)</td>
</tr>
<tr>
<td>no. of symptom weeks after surgery</td>
<td>2.61 (2.07,2.82)</td>
<td>2.44 (2.07,2.83)</td>
<td>2.44 (1.36,3.43)</td>
<td>2.41 (1.36,3.43)</td>
</tr>
<tr>
<td>Recurrence prob. on PPIs (6-12 months)</td>
<td>4.71 (3.98,5.11)</td>
<td>4.55 (3.98,5.10)</td>
<td>4.55 (2.54,4.60)</td>
<td>3.62 (2.54,4.60)</td>
</tr>
<tr>
<td>Recurrence prob. on H2RAs (6-12 months)</td>
<td>2.65 (2.33,3.03)</td>
<td>2.67 (2.32,3.01)</td>
<td>2.66 (0.25,2.18)</td>
<td>1.25 (0.25,2.18)</td>
</tr>
<tr>
<td>Recurrence prob. on low dose PPIs (6-12 months)</td>
<td>3.47 (3.16,3.82)</td>
<td>3.49 (3.16,3.82)</td>
<td>3.49 (2.67,4.69)</td>
<td>3.74 (2.67,4.69)</td>
</tr>
</tbody>
</table>

6 Discussion

Sensitivity analysis based on the concept of value of information allows the model-user or decision-maker to relate the importance of each uncertain input parameter directly to the decision problem at hand, something that is lacking in a traditional variance-based sensitivity analysis. If the decision-maker then wishes to consider reducing their uncertainty about a particular input, this can be explored through considering the EVSI for that input parameter. We have also shown that computation of partial EVPIs and EVSIs for different sample designs can be considerably more efficient using the emulator approach.
<table>
<thead>
<tr>
<th>Sample size</th>
<th>true EVSI</th>
<th>GP, Uniform prior</th>
<th>GP, Exp prior</th>
<th>Simpson/LHS</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>4.61</td>
<td>4.78</td>
<td>4.77</td>
<td>4.56</td>
</tr>
<tr>
<td></td>
<td>(4.35,5.20)</td>
<td>(4.34,5.18)</td>
<td>(2.75,6.23)</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>7.82</td>
<td>8.09</td>
<td>8.07</td>
<td>7.81</td>
</tr>
<tr>
<td></td>
<td>(7.55,8.61)</td>
<td>(7.53,8.59)</td>
<td>(6.46,9.19)</td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>9.90</td>
<td>10.23</td>
<td>10.21</td>
<td>9.90</td>
</tr>
<tr>
<td></td>
<td>(9.64,10.80)</td>
<td>(9.63,10.78)</td>
<td>(8.64,11.05)</td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>11.37</td>
<td>11.75</td>
<td>11.73</td>
<td>11.39</td>
</tr>
<tr>
<td></td>
<td>(11.13,12.35)</td>
<td>(11.11,12.32)</td>
<td>(10.27,12.42)</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>12.47</td>
<td>12.88</td>
<td>12.86</td>
<td>12.49</td>
</tr>
<tr>
<td></td>
<td>(12.25,13.50)</td>
<td>(12.22,13.48)</td>
<td>(11.51,13.52)</td>
<td></td>
</tr>
<tr>
<td>300</td>
<td>16.80</td>
<td>17.39</td>
<td>17.32</td>
<td>16.80</td>
</tr>
<tr>
<td></td>
<td>(16.64,18.02)</td>
<td>(16.62,18.00)</td>
<td>(16.04,17.56)</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: True values, Gaussian process estimates and Simpson/Monte Carlo estimates of the EVSI for parameter $X_{19}$ at various sample sizes. The Gaussian process estimates are based on 600 model runs, and the Simpson/LHS estimates are based on 12120 model runs.

The analysis also illustrates the advantage of the emulator approach in general. Once the emulator has been constructed, the EVPI and EVSI analysis can be conducted without having to do any further runs of the model. The additional computation required is not trivial; MCMC sampling has been used, but for genuinely expensive models this is relatively minor compared to the large numbers of additional runs that would be required for Monte Carlo EVPI estimates.

One extension of this method would be to consider the case when the set of possible decisions $D$ is not finite. One might consider treating the decision $d$ as a model input, and constructing an emulator of $U\{d, \eta(x)\}$. However, this could introduce a complication in that given $X_i$, optimisation of $E[U\{d, \eta(X)\}|X_i]$ with respect to $d$ might not be straightforward. Simulation-based methods for maximising expected utility functions are presented in Muller et al. (2004) and Amzal et al. (2006) in the context of optimal design, but without the additional problem of the utility function being computationally expensive to evaluate. The current methodology is also only suitable for fairly simple choices of input distributions, and further work is needed.
to deal with more complex cases.

We acknowledge that in some applications simply identifying an appropriate utility function will in itself be very difficult, and so there will always be a role for variance-based measures. However, we argue that by considering both the possible actions that could be taken on the basis of model output and the consequences of those actions, model-users will be able to gain a better understanding of the importance of their uncertain input parameters.

7 Acknowledgements

I would like to thank Andrew Briggs, Ron Goeree, Gord Blackhouse and Bernie O’Brien for providing the input distributions for use in the GERD model.

References


