Decision-theoretic Sensitivity Analysis using Value of Information

Jeremy Oakley
Department of Probability and Statistics
University of Sheffield, UK
j.oakley@sheffield.ac.uk

June 13, 2003

When a computer model is to be used to guide a decision, it is important for the decision-maker to acknowledge and investigate the uncertainty in the model. Typically, there will be uncertainty surrounding the true values of the input parameters in the model that should be used for the decision problem in question, and this then induces uncertainty in the output of the model. If the decision-maker considers their probability distribution for each unknown input in the model, they can then derive their probability distribution for the model output. The combination of their output distribution and an appropriate utility/loss function can then guide their decision.

In some cases, it may be possible to learn more about some or all of the uncertain input parameters before a final decision is made. In this case, it is then desirable to assess the importance of each uncertain input parameter in the model. Quantifying parameter importance is known as global or probabilistic sensitivity analysis. A measure of parameter importance that has been advocated previously is the variance-based measure. Variance-based measures consider the contribution of each uncertain input parameter to the variance of the model output. However, uncertainty about the model output as characterised by its variance is not necessarily equivalent to uncertainty about the optimum decision. Consequently, using variance-based measures to establish parameter importance in decision problems can in some cases produce misleading results, even as far as ranking the parameters in the wrong order of importance.

An alternative measure of parameter importance can be derived within the framework of utility theory. The idea is to determine whether different values of a particular input parameter lead to different optimum decisions, and if so, how much the expected utility/loss under alternative optimum decisions varies. Specifically, the expected utilty of learning the true numerical value of an uncertain input parameter before the decision is made can be calculated. This quantity is known as the partial expected value of perfect information (partial EVPI), and precisely quantifies the importance of an uncertain input variable. When the specific purpose of the model is to guide a decision within a clearly defined utility/loss structure, we advocate the partial EVPI as the single correct measure of an uncertain parameter's importance.

In most practical situations, the decision-maker will not be able to learn the true value of an uncertain input parameter precisely, even if they desire to do so. The more likely possibility is that they may have the option of collecting more data to reduce their uncertainty about the unknown parameter. The expected value of perfect information framework can be extended to consider the expected value of collecting this data before making the decision; this is known as the expected value of sample information (EVSI). EVSI measures can then be used for deriving optimal sample sizes.

Both partial EVPIs and EVSIs can be computed using Monte Carlo methods. Unfortunately, to obtain these measures accurately, the model needs to be run a very large number of times (potentially millions) at different sets of input parameter values. For computationally expensive computer models, evaluating these measures may then require prohibitively lengthy computing times. However, in many cases it will be possible to exploit a feature of the computer model to dramatically speed up the computation; the function mapping inputs to output is often a smooth function. If the model is run at a particular set of input values and the output is observed, we will then also have information about the likely output at neighbouring sets of input parameter values.

When the time needed for a single run of the model is non-trivial, it can be highly advantageous to construct an *emulator*, a statistical approximation to the original computer model based on a fairly small number of different runs of that model. The emulator can then be used to give a fast approximation to the computer model regardless of the complexity of the model. An emulator is a regression model, and any regression technique can be employed. Our preferred option is the Gaussian process model. The Gaussian process emulator is a non-parametric approach that with the exception of continuity, makes no other assumptions about the functional form of the computer model. Gaussian processes have been used successfully before for efficient computation in other areas of sensitivity and uncertainty analysis. It will be demonstrated that the Gaussian process approach is of the order of 1000 times more efficient than Monte Carlo methods in terms of numbers of model runs, for computing partial EVPIs and EVSIs.

An application is given in the field of health economics. Given limited financial resources to spend on healthcare, increasing emphasis is being placed on the cost-effectiveness of treatments in addition to their clinical effectiveness. To estimate the cost-effectiveness of a proposed treatment, information is drawn together from a variety of sources and represented in an economic models. A decision-maker will use then use the output of the model to help decide whether or not to approve the new treatment. There is always uncertainty regarding the values of the input parameters needed for the model; for example, it will not be known exactly how effective the treatment is, or what financial resources the patients on the treatment will use. An uncertainty analysis will be essential to make a rational decision regarding whether or not to adopt the treatment in the light of input uncertainty in the model. There will also be particular interest in conducting a probabilistic sensitivity analysis when using the model. It will often be possible to obtain more data regarding some of the model parameters, and hence reduce input uncertainty. Additionally, a certain class of models, known as patient simulation models, require an extensive simulation to produce the output for a single choice of input parameters. These models can be very computationally expensive, requiring in some cases in excess of an hour per run. In this scenario, emulator methods are essential for computation of EVPIs and EVSIs.