Journal of Health Economics 29 (2010) 468–477

Contents lists available at ScienceDirect

Journal of Health Economics

journal homepage: www.elsevier.com/locate/econbase

Simulation sample sizes for Monte Carlo partial EVPI calculations

Jeremy E. Oakley[a,*], Alan Brennan[b], Paul Tappenden[b], Jim Chilcott[b]

[a] Department of Probability and Statistics, University of Sheffield, The Hicks Building, Hounsfield Road, Sheffield, S3 7RH, UK
[b] School of Health and Related Research, University of Sheffield, UK

ARTICLE INFO

Article history:
Received 22 February 2007
Received in revised form
10 November 2009
Accepted 10 March 2010
Available online 17 March 2010

JEL classification:
I18
C110
C150

Keywords:
Economic model
Expected value of perfect information
Monte Carlo estimation
Bayesian decision theory

ABSTRACT

Partial expected value of perfect information (EVPI) quantifies the value of removing uncertainty about unknown parameters in a decision model. EVPIs can be computed via Monte Carlo methods. An outer loop samples values of the parameters of interest, and an inner loop samples the remaining parameters from their conditional distribution. This nested Monte Carlo approach can result in biased estimates if small numbers of inner samples are used and can require a large number of model runs for accurate partial EVPI estimates. We present a simple algorithm to estimate the EVPI bias and confidence interval width for a specified number of inner and outer samples. The algorithm uses a relatively small number of model runs (we suggest approximately 600), is quick to compute, and can help determine how many outer and inner iterations are needed for a desired level of accuracy. We test our algorithm using three case studies.

1. Introduction

It is common practice to use decision models to estimate the expected net benefit of alternative strategy options open to a decision maker (Raiffa, 1968; Brennan and Akehurst, 2000). Invariably in health economic cost-effectiveness models, the input parameters' true values are not known with certainty. A probabilistic sensitivity analysis will then be required to investigate the consequences of this input parameter uncertainty (Briggs and Gray, 1999). Simple Monte Carlo propagation of input uncertainty through the model can provide an estimate of the distribution of net benefit, thus giving its expected value, and the probability that the incremental net benefit will be positive (Van Hout et al., 1994).

More detailed analysis can compute the partial expected value of perfect information (partial EVPI), that is, the value to the decision maker of learning the true value of the uncertain parameter input before deciding whether to adopt the new treatment (Raiffa, 1968; Claxton and Posnett, 1996). Partial EVPIs are recommended because they quantify the importance of different parameters using decision-theoretic arguments (Claxton, 1999; Meltzer, 2001), and thus can quantify the societal value of further data collection to help design and prioritise research projects (Claxton and Thompson, 2001; Chilcott et al., 2003a). Unfortunately, evaluating partial EVPIs is computationally demanding. Generally, a two level Monte Carlo procedure is needed, requiring many runs of the economic model. A developing literature has described this more and more clearly (Felli and Hazen, 1998, 2003; Brennan et al., 2002, 2007; Ades et al., 2004; Yokota and Thompson, 2004; Koerkamp et al., 2006; Brennan and Kharroubi, 2007b). The procedure begins with an outer loop sampling values of the parameters of interest, and for each of these, an inner loop sampling the remaining parameters from their conditional distribution (Brennan et al., 2002; Ades et al., 2004).

Increasingly large inner and outer sample sizes will produce increasingly accurate estimates of partial EVPI. However, the nested Monte Carlo approach can result in biased estimates if the inner loop sample size is small (Brennan and Kharroubi, 2007a), regardless of the outer loop sample size. Most studies recommend 'large enough' inner and outer samples, e.g. 1000 or 10,000 in order to produce 'accurate' partial EVPI estimates, but none discuss in detail the choice of sample sizes in relation to bias or confidence interval widths.

In this paper we present an algorithm for determining the inner and outer loop sample sizes needed to estimate a partial EVPI to a required level of accuracy. We first review the theory of estimating partial EVPIs via Monte Carlo. We propose using normal approximations for conditional expected net benefits to estimate the bias

* Corresponding author. Tel.: +44 114 2223853.
E-mail address: j.oakley@sheffield.ac.uk (J.E. Oakley).

0167-6296/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhealeco.2010.03.006

and variance of a Monte Carlo estimate for a specified number of inner and outer loops. We set out an algorithm, using a moderate number of model runs, to produce an estimate of the bias and confidence interval widths for partial EVPI estimates, and hence to help determine the inner and outer sample sizes needed. The results of applying and testing the algorithm's performance in three case studies are given, followed by a discussion of how this approach can be applied more generally to support Monte Carlo estimation of partial EVPI.

If the decision model is complex, requiring substantial computation time for each model run, then it may not be feasible to do the number of model runs required for our algorithm, or the algorithm may suggest that an infeasible number of model runs are required for sufficiently accurate partial EVPI estimates. In this case, one can use instead a Gaussian process meta-model, which approximates the economic model, enabling more efficient Monte Carlo sampling, but with the disadvantages that it is far more complex to program initially and is not always feasible for models with large numbers of uncertain input parameters (e.g. in excess of 100) (Oakley, 2009; Oakley and O'Hagan, 2004; Stevenson et al., 2004).

2. Methods

2.1. Evaluating partial EVPI via Monte Carlo sampling

Suppose we have T treatment options, and our economic model computes the net benefit $NB(t, \mathbf{x})$ for treatment $t = 1, \ldots, T$, when provided with input parameters $\mathbf{x}$. We denote the true, uncertain values of the input parameters by $\mathbf{X} = \{X_1, \ldots, X_d\}$, so that the true, uncertain net benefit of treatment t is given by $NB(t, \mathbf{X})$. When considering the partial EVPI of a particular parameter X_i, we use the notation $\mathbf{X}_{-i} = \{X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_d\}$ to denote all the inputs in $\mathbf{X}$ except X_i. For any input X, we use subscripts to denote a particular input parameter (or group of parameters) in the economic model, and a superscript to denote a randomly sampled value of that input. Note that all of the equations presented, and indeed the proposed algorithm, are the same if we are considering a group of parameters $\mathbf{X}_i$ rather than a single scalar parameter.

We use the notation $NB(t, \mathbf{X}) = NB(t, X_i, \mathbf{X}_{-i})$ and for expectations $E_{\mathbf{X}}$, E_{X_i} and $E_{\mathbf{X}_{-i}|X_i}$ denote expectations over the full joint distribution of $\mathbf{X}$, the marginal distribution of X_i, and the conditional distribution of $\mathbf{X}_{-i}|X_i$, respectively.

The partial EVPI for the ith parameter X_i is given by

$$EVPI(X_i) = E_{X_i}\left[\max_t E_{\mathbf{X}_{-i}|X_i}\{NB(t, X_i, \mathbf{X}_{-i})\}\right] - \max_t E_{\mathbf{X}}\{NB(t, \mathbf{X})\}. \quad (1)$$

The second term in the RHS of (1) can be estimated by Monte Carlo. We sample $\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(N)}$ from the distribution of $\mathbf{X}$, evaluate $NB(t, \mathbf{X}^{(n)})$ for $t = 1, \ldots, T$ and $n = 1, \ldots, N$, and then estimate $NB^* = \max_t E_{\mathbf{X}}\{NB(t, \mathbf{X})\}$ by

$$\widehat{NB}^* = \max_t \frac{1}{N}\sum_{n=1}^{N} NB(t, \mathbf{X}^{(n)}).$$

We do not consider the choice of N in this paper. In practice, it is usually feasible to have N sufficiently large such that $\widehat{NB}^*$ is an accurate estimate of NB^*, and can be used in the partial EVPI estimate of any parameter or group of parameters.

In this paper we concentrate on estimating the first term in the RHS of (1). We define the maximum conditional expected net benefit given X_i as

$$m(X_i) = \max_t E_{\mathbf{X}_{-i}|X_i}\{NB(t, X_i, \mathbf{X}_{-i})\},$$

and hence, the first term on the RHS of Eq. (1) can be written as $E_{X_i}\{m(X_i)\}$.

Typically, we cannot evaluate $m(X_i)$ analytically, and so we estimate it using Monte Carlo. We randomly sample J values of $\mathbf{X}_{-i}$ from the conditional distribution of $\mathbf{X}_{-i}|X_i$ to obtain $\mathbf{X}_{-i}^{(1)}, \ldots, \mathbf{X}_{-i}^{(J)}$. We run the economic model for each of the J sampled inputs to obtain $NB(t, X_i, \mathbf{X}_{-i}^{(j)})$ for $t = 1, \ldots, T$ and $j = 1, \ldots, J$, and estimate $m(X_i)$ by

$$\hat{m}(X_i) = \max_t \frac{1}{J}\sum_{j=1}^{J} NB(t, X_i, \mathbf{X}_{-i}^{(j)}). \quad (2)$$

We now approximate $E_{X_i}\{m(X_i)\}$ by $E_{X_i}\{\hat{m}(X_i)\}$, and estimate $E_{X_i}\{\hat{m}(X_i)\}$ by randomly sampling K values $X_i^{(1)}, \ldots, X_i^{(K)}$ from the distribution of X_i and computing the estimator

$$\hat{E}_{X_i}\{\hat{m}(X_i)\} = \frac{1}{K}\sum_{k=1}^{K} \hat{m}\{X_i^{(k)}\}. \quad (3)$$

We refer to the process of obtaining the maximum conditional net benefit estimator $\hat{m}\{X_i\}$ for a single given X_i using J samples from the distribution of $\mathbf{X}_{-i}|X_i$ as the *inner level* Monte Carlo procedure. The process of calculating (3) (given the values $\hat{m}(X_{i,1}), \ldots, \hat{m}(X_{i,K})$) using the K samples $X_i^{(1)}, \ldots, X_i^{(K)}$ is the *outer level* Monte Carlo procedure. The Monte Carlo estimate for the partial EVPI for parameter X_i is therefore

$$\widehat{EVPI}(X_i) = \hat{E}_{X_i}\{\hat{m}(X_i)\} - \widehat{NB}^* = \frac{1}{K}\sum_{k=1}^{K}$$

$$\left[\max_t \left\{\frac{1}{J}\sum_{j=1}^{J} NB(t, X_i^{(k)}, \mathbf{X}_{-i}^{(j,k)})\right\}\right] - \max_t \frac{1}{N}\sum_{n=1}^{N} NB(t, \mathbf{X}^{(n)}), \quad (4)$$

where $\mathbf{X}_{-i}^{(j,k)}$ is the jth sample from the distribution of $\mathbf{X}_{-i}|X_i = X_i^{(k)}$. The inner and outer level sample sizes, J and K, respectively, mean that the total number of runs of the economic model required for the partial EVPI estimate is $J \times K$ (given $\widehat{NB}^*$). The objective in this paper is to determine what values of J and K should be used in order to obtain a sufficiently accurate estimate of the partial EVPI.

2.2. Uncertainty and bias in Monte Carlo estimates of partial EVPI

Any Monte Carlo estimate of an expectation is subject to uncertainty due to random sampling, and the larger the number of samples, the more this uncertainty is reduced. We assume that N is sufficiently large such that uncertainty in NB^* is small relative to uncertainty in $E_{X_i}\{m(X_i)\}$. If K is sufficiently large then a normal approximation will apply, and a 95% confidence interval for $EVPI(X_i)$ is given by

$$\left(\widehat{EVPI}(X_i) - 1.96\sqrt{\frac{Var\{\hat{m}(X_i)\}}{K}}, \quad \widehat{EVPI}(X_i) + 1.96\sqrt{\frac{Var\{\hat{m}(X_i)\}}{K}}\right). \quad (5)$$

This interval appears to suggest that to obtain a sufficiently accurate estimate of the partial EVPI, we just need K to be large, as the width of the confidence interval will decrease as K increases. Unfortunately, this is not the case because, as we now show, $\hat{m}(X_i)$ is an upwards biased estimator of $m(X_i)$, and so $\widehat{EVPI}(X_i)$ is a biased estimator of $EVPI(X_i)$. The bias is independent of K, but it depends