VALUE IN HEALTH 16 (2013) 438–448

Available online at www.sciencedirect.com

SciVerse ScienceDirect

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

Need for Speed: An Efficient Algorithm for Calculation of Single-Parameter Expected Value of Partial Perfect Information

Mohsen Sadatsafavi, MD,PhD[1,2,3,], Nick Bansback, PhD[4], Zafar Zafari, MSc[2], Mehdi Najafzadeh, MA, PhD[2,5], Carlo Marra, PharmD, PhD[2]*

[1]*Department of Medicine, University of British Columbia, Vancouver, BC, Canada;* [2]*Collaboration for Outcomes Research and Evaluation, Faculty of Pharmaceutical Sciences, University of British Columbia, Vancouver, BC, Canada;* [3]*Centre for Clinical Epidemiology and Evaluation, Vancouver Coastal Health Institute, University of British Columbia, Vancouver, BC, Canada;* [4]*School of Population and Public Health, University of British Columbia, Vancouver, BC, Canada;* [5]*Mongan Institute for Health Policy, Massachusetts General Hospital, Harvard University, Boston, MA, USA*

ABSTRACT

Background: The expected value of partial perfect information (EVPPI) is a theoretically justifiable and informative measure of uncertainty in decision-analytic cost-effectiveness models, but its calculation is computationally intensive because it generally requires two-level Monte Carlo simulation. We introduce an efficient, one-level simulation method for the calculation of single-parameter EVPPI. **Objective:** We show that under mild regularity assumptions, the expectation-maximization-expectation sequence in EVPPI calculation can be transformed into an expectation-maximization-maximization sequence. By doing so, calculations can be performed in a single-step expectation by using data generated for probabilistic sensitivity analysis. We prove that the proposed estimator of EVPPI converges in probability to the true EVPPI. **Methods and Results:** The performance of the new method was empirically demonstrated by using three exemplary decision models. Our proposed method seems to achieve remarkably higher accuracy than the two-level method with a fraction of its computation costs, though the achievement in accuracy was not uniform and varied across the parameters of the models. Software is provided to calculate single-parameter EVPPI based on the probabilistic sensitivity analysis data. **Conclusions:** The new method, though applicable only to single-parameter EVPPI, is fast, accurate, and easy to implement. Further research is needed to evaluate the performance of this method in more complex scenarios and to extend such a concept to similar measures of decision uncertainty.

Keywords: Bayesian analysis, decision uncertainty, medical decision making, value of information..

Background

A robust, unbiased, and easy-to-implement method for calculating the expected value of the outcomes and quantifying uncertainty in cost-effectiveness analyses (CEAs) is to perform a Monte Carlo simulation. In model-based CEAs, this is done by randomly drawing from the distribution of uncertain parameters and calculating cost and effectiveness outcomes. Generally, this method in the health economics literature is referred to as probabilistic sensitivity analysis (PSA).

The expected value of partial perfect information (EVPPI) is the expected benefit by completely resolving uncertainty around a subset of evidence [1]. The EVPPI can be used as a generic measure to compare the relative importance of uncertainty in parameters of a decision model. Population EVPPI sets an analytical upper limit on the budget of future research aimed at obtaining more information on those parameters.

Unfortunately, calculation of the EVPPI is often computationally intensive because it generally requires a two-level nested Monte Carlo expectation [1]. Some economic models are multilinear on input parameters; for such models, the inner expectation can be performed without the need for Monte Carlo simulation [2]. Many practical models, however, are complex functional of their input parameters and therefore the two-level Monte Carlo simulation becomes necessary for the EVPPI calculation. Another challenge in using the two-level Monte Carlo method arises in models with correlated inputs because this method might necessitate sampling from complex conditional distributions. Alternative methods for EVPPI calculation have been proposed; however, they either are based on parametric assumptions or require considerable expertise for implementation [3–5].

This article presents a novel and simple method for calculating single-parameter EVPPI. The main advantages of the present method are its computational efficiency, that it only relies on the data generated through the PSA, and that one set of simulations is enough to generate EVPPIs for each uncertain parameter of the model.

We begin by defining the mathematical formulation of the EVPPI. Next, we outline the concepts underlying the new method and accordingly propose a generic one-level estimator for the

* *Address correspondence to:* Mohsen Sadatsafavi, Faculty of Medicine, University of British Columbia, 2146 East Mall, Vancouver, BC, Canada V6T 1Z3.

E-mail: mohsen.sadatsafavi@alumni.ubc.ca.

1098-3015/$36.00 – see front matter Copyright © 2013, International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc.

http://dx.doi.org/10.1016/j.jval.2012.10.018

EVPPI. We prove the convergence in probability of the estimator to the true EVPPI. We elaborate on a visual tool for checking the assumptions underlying the algorithm. We then discuss some challenges in using the estimator for models with more than two decisions and accordingly propose a modified estimator for such models. We compare the numerical and computational performance of the new algorithm with the conventional two-level Monte Carlo simulation method by using three exemplar models.

Methods

Notations

Let Ω denote the set of all uncertain parameters that are represented by probability distributions. We denote the single parameter whose EVPPI is of interest as θ. Let $NB_d(\mathbf{\Omega})$ represent the function that calculates the net benefit (NB) of the dth decision (out of a total of D decisions) associated with a realized value of $\mathbf{\Omega}$. A fundamental assumption in any stochastic CEA is that $NB_d(\mathbf{\Omega})$ has a finite expected value and a finite variance, so that by calculating $NB_d(\mathbf{\Omega})$ with several random samples from $\mathbf{\Omega}$ and averaging the results, one can obtain a convergent estimate for the expected value of cost and effectiveness for each decision. Let $P_\theta(\theta)$ be the (continuous) probability density function of θ. Let θ_L and θ_U denote the lower and upper bounds (either or both can be infinite) of θ.

The PSA is performed by randomly drawing from the distribution of $\mathbf{\Omega}$ and calculating the NB for all D decisions, repeating this process n times. The PSA data can be denoted by $\{\hat{\boldsymbol{\theta}}, \widehat{\mathbf{NB}}\}$, where $\hat{\boldsymbol{\theta}} = \left(\hat{\theta}_i,\ i=1:n,\ \hat{\theta}_1 \le \hat{\theta}_2 \ldots \le \hat{\theta}_n\right)$ is the vector of random draws from the distribution of the parameter of interest (ordered ascendingly, without loss of generality) and $\widehat{\mathbf{NB}} = \left(\widehat{NB}_{i,d}, i=1:n,\ d=1:D\right)$ is the corresponding matrix of NBs (the draws from the other parameters are irrelevant and are omitted in this notation).

We define the function $f(\theta)$ as the expected NB of the dth decision conditional on the parameter of interest being fixed at θ:

$$f_d(\theta) = \mathrm{E}_{\mathbf{\Omega}^-|\theta}(NB_d(\mathbf{\Omega})) \tag{1}$$

where $\mathbf{E}_{\mathbf{\Omega}^-|\theta}$ indicates expectation with respect to all parameters except θ in $\mathbf{\Omega}$. In this case, the EVPPI for the parameter θ can be written as

$$EVPPI_\theta = \mathrm{E}_\theta\{\max_d[f_d(\theta)]\} - \max_d\{\mathrm{E}_\theta[f_d(\theta)]\} \tag{2}$$

The challenging term in the above equation is indeed $\mathrm{E}_\theta\{\max_d[f_d(\theta)]\}$, which involves an expectation (with respect to the distribution of the parameter of interest)-maximization (over the set of all decisions)-expectation (with respect to the distribution of all other parameters) sequence.

Concept

The concept underlying the present method can be summarized as changing the expectation-maximization-expectation sequence to an expectation-maximization-maximization sequence. Define the function $\delta(\theta)$ as

$$\delta(\theta) = \mathrm{argmax}_d\, f_d(\theta). \tag{3}$$

the function that returns the index of the decision that has the highest expected NB at a given value of θ. The heuristic is that if a decision has the maximum expected NB at θ, it probably has the maximum expected NB at the vicinity of θ as well. The δ function is, therefore, for the most realistic scenarios a piecewise constant with finitely many pieces. As such, it is good enough to restrict our attention to the set of such functions in calculating the EVPPI.

The Estimator

Restricting our attention to the set of piecewise constant functions proves advantageous: Assume δ consists of $M+1$ pieces created by M discontinuity points. Define the vector $\mathbf{M}$ $(m_i; i=1:M;\ m_1 < m_2 < \cdots < m_M)$ as a real-valued vector with all its elements within the range of θ. Then, the first term on the right-hand side of Eq. 2 can be reexpressed as the solution to the following expectation-maximization-maximization problem:

$$\mathrm{E}_\theta\{\max_d[f_d(\theta)]\} = \max_{m_1,\ldots,m_M}\left\{\sum_{i=0}^{M}\max_d\left[\int_{m_i}^{m_{i+1}} P_\theta(x) f_d(x)dx\right]\right\}, \tag{4}$$

with $m_0 = \theta_L$ and $m_{M+1} = \theta_U$. The reason for using such an equation for EVPPI is that it involves one level of expectation (integration) because within each segment we have $\int P_\theta(x).f_d(x).dx = \mathrm{E}_\theta\{\mathrm{E}_{\mathbf{\Omega}^-|\theta}(NB_d(\mathbf{\Omega}))\} = \mathrm{E}_{\mathbf{\Omega}}(NB_d(\mathbf{\Omega}))$. As such, this estimator has a counterpart in the PSA data: define the "segmentation vector" $\mathbf{L}$ $(l_i; i=1:L; l_1 < l_2 < \ldots < l_L)$ as the $L\times 1$ vector of ascending integers in $1{:}n$. The elements of L correspond to the row indices of elements in $\hat{\boldsymbol{\theta}}$. Define $\psi(\mathbf{L})$ as

$$\psi(\mathbf{L}) = \frac{1}{n}\sum_{i=0}^{L}\max_d\left(\sum_{j=l_i+1}^{l_{i+1}}\widehat{NB}_{j,d}\right), \tag{5}$$

With $l_0 = 0$ and $l_{L+1} = n$, define

$$\widehat{EVPPI} = \max_{l_1,\ldots,l_L}\psi(\mathbf{L}) - \frac{1}{n}.\max_d\sum_{i=1}^{n}\widehat{NB}_{i,d}, \tag{6}$$

as an estimator of the EVPPI from the PSA data. After the Discussion section, we have provided a sketch of a proof that as long as L, the size of the segmentation vector specified by the investigator, is at least as large as M, the number of discontinuity points on $\delta(\theta)$, the above estimator converges in probability to the *EVPPI*. A complete proof is provided in Supplementary Material 1 found at http://dx.doi.org/10.1016/j.jval.2012.10.018. Figure 1 provides an illustration relating PSA data, $f_d(\theta)$ function defined in Equation 1, and $\delta(\theta)$ defined in Eq. 3.

Deciding on the Number of Segments, and a Visual Tool for Checking the Assumptions

Our proposed estimator requires that the investigator specifies the size of the segmentation vector, and the convergence of $\widehat{EVPPI}$ to the true *EVPPI* rests on the critical assumption that the size of the segmentation vector is at least as large as M, the number of discontinuity points on $\delta(\theta)$. M is indeed unknown; although one can choose a very large L to ensure this condition is satisfied, each additional segmentation point can cause overestimation of the *EVPPI* in finite PSA samples. In fact by choosing $L=n$, the size of the PSA data, the maximization will take place at each point of $\hat{\boldsymbol{\theta}}$ such that the estimator will converge to the expected value of perfect information (EVPI), which is the expected gain in resolving uncertainty among all model parameters. Therefore, a parsimonious choice for L is important for avoiding overtly overestimated EVPPIs.

Fortunately, there is a visual tool for assessing the adequacy of the size of a given segmentation vector: for a pair of decisions a and b, define the quantity

$$\hat{S}_{a,b}(\theta) \equiv \frac{1}{n}\sum_{i=1}^{n} I\left(\hat{\theta}_i < \theta\right)\left(\widehat{NB}_{i,a} - \widehat{NB}_{i,b}\right), \tag{7}$$

where $I(.)$ is the indicator function taking the value of 1 when the condition is satisfied and 0 otherwise. $\hat{S}_{a,b}$ is the running cumulative sum of the incremental NBs between decisions a and b from the PSA data after creating an ordered version of the PSA data (such that $\hat{\theta}_1 \le \hat{\theta}_2 \ldots \le \hat{\theta}_n$), which can be calculated very easily in a spreadsheet.