Supplementary Material: Bayesian Non-Linear Regression to Estimate the Expected Value of Sample Information using Moment Matching Across Different Sample Sizes

Anna Heath*, Ioanna Manolopoulou and Gianluca Baio

[a]*Department of Statistical Science, University College London, Gower Street, London, WC1E 6BT*

1. Non-linear functional form

As stated in the paper, the functional form of the non-linear regression

$$f(N) = \sigma^2_{\boldsymbol{\phi}} \frac{N}{N+h} \tag{1}$$

is the exact relationship between N and $\sigma^2_{\boldsymbol{X}}(N)$ in normal-normal conjugate models with only one underlying parameter. We briefly demonstrate this relationship in this section.

Assume that we have a parameter of interest θ which is modelled with a normal distribution

$$\theta \sim Normal(\mu, \varepsilon^2).$$

We consider collecting additional information to inform this parameter with a sampling distribution conditional on θ

$$X_i \sim Normal(\theta, \varepsilon^2_X),$$

for $i = 1, \ldots, N$. As this is a normal-normal conjugate model with known variance, we know that the posterior of θ is given by

$$\theta \mid \boldsymbol{X} \sim Normal\left(\frac{\varepsilon^2_X \mu + N\varepsilon^2 \bar{\boldsymbol{X}}}{\varepsilon^2_X + N\varepsilon^2}, \frac{\varepsilon^2 \varepsilon^2_X}{\varepsilon^2_X + N\varepsilon^2}\right),$$

where $\bar{\boldsymbol{X}} = \frac{1}{N}\sum_{i=1}^{N} X_i$.

Recall that the variance of the preposterior mean for θ for a specific N is calculated as follows:

$$\begin{aligned} \sigma^2_{\boldsymbol{X}} = \mathrm{Var}_{\boldsymbol{X}}\left[\mathrm{E}_{\boldsymbol{\theta}|\boldsymbol{X}}\left[\theta\right]\right] =& \mathrm{Var}\left[\theta\right] - \mathrm{E}_{\boldsymbol{X}}\left[\mathrm{Var}_{\theta|\boldsymbol{X}}\left[\theta\right]\right] & (2)\\ =& \varepsilon^2 - \mathrm{E}_{\boldsymbol{X}}\left[\frac{\varepsilon^2 \varepsilon^2_X}{\varepsilon^2_X + N\varepsilon^2}\right]. & (3) \end{aligned}$$

In this setting, the posterior variance of θ does not change depending on the value of the observed sample $\boldsymbol{X}$ but only depends on the known variance ε_X. Therefore,

$$\sigma^2_{\boldsymbol{X}} = \varepsilon^2 - \frac{\varepsilon^2 \varepsilon^2_X}{\varepsilon^2_X + N\varepsilon^2} = \frac{\cancel{\varepsilon^2 \varepsilon^2_X} + \varepsilon^2 N \varepsilon^2 \cancel{- \varepsilon^2 \varepsilon^2_X}}{\varepsilon^2_X + N\varepsilon^2}.$$

To get the equation in (1), we then divide by ε^2 in the numerator and denominator giving,

$$\sigma^2_{\boldsymbol{X}} = \varepsilon^2 \frac{N}{\frac{\varepsilon^2_X}{\varepsilon^2} + N},$$

and finally set $h = \frac{\varepsilon^2_X}{\varepsilon^2}$.

*Corresponding Author

2. Multi-Decision Models

In general, there is little theoretical difference between the moment matching method for multi-decision models and the moment matching method in the dual-decision setting that has been presented in the paper. The method still proceeds by calculating the posterior variance of the incremental net benefit across different potential simulated datasets. These calculated variances are then used to rescale the expectation of $\text{INB}^{\boldsymbol{\phi}}$. However, as we have more than two decisions, the incremental net benefit is defined in a slightly different manner.

In the multi-decision setting, we choose a reference treatment, say T, and then calculate the incremental net benefit of all the treatments with respect to this reference treatment, i.e.

$$\text{INB}_t^{\boldsymbol{\theta}} = \text{NB}_t^{\boldsymbol{\theta}} - \text{NB}_T^{\boldsymbol{\theta}}.$$

Therefore, the incremental net benefit $\text{INB}^{\boldsymbol{\theta}}$ is a multivariate random vector:

$$\text{INB}^{\boldsymbol{\theta}} = \begin{pmatrix} \text{INB}_1^{\boldsymbol{\theta}} \\ \text{INB}_2^{\boldsymbol{\theta}} \\ \vdots \\ \text{INB}_{T-1}^{\boldsymbol{\theta}} \end{pmatrix}. \tag{4}$$

This means that the variance of $\text{INB}^{\boldsymbol{\theta}}$ is a variance-covariance matrix rather than a scalar. Thus, we denote it $\Sigma_{\boldsymbol{\theta}}$. If the EVSI calculation is being performed in `R`, then the variance-covariance matrix is computed when using the `var()` function so this extension adds no complexity to the moment matching procedure.

To find $\text{INB}^{\boldsymbol{\phi}}$ using non-parametric regression, a regression curve should be fitted for incremental net benefit to give a multivariate vector:

$$\text{INB}^{\boldsymbol{\phi}} = \begin{pmatrix} \text{INB}_1^{\boldsymbol{\phi}} \\ \text{INB}_2^{\boldsymbol{\phi}} \\ \vdots \\ \text{INB}_{T-1}^{\boldsymbol{\phi}} \end{pmatrix}, \tag{5}$$

with a variance covariance matrix denoted $\Sigma_{\boldsymbol{\phi}}$. The standard moment matching method would proceed as normal, i.e. for each potential sample $\boldsymbol{X}_q$, we would calculate the variance of $\text{INB}^{\boldsymbol{\theta}}$, the only difference in the multi-decision setting is that the variance is a variance-covariance matrix which we denote Σ_q.

In the multi-decision setting, as before, we are aiming to estimate the distribution of $\mu^{\boldsymbol{X}} = \text{E}_{\boldsymbol{\theta}|\boldsymbol{X}}\left(\text{INB}^{\boldsymbol{\theta}}\right)$. However, this is now a multivariate distribution with a variance-covariance matrix, denoted $\Sigma_{\boldsymbol{X}}$. Each element of this matrix, $\sigma_{\boldsymbol{X}}^{ij}$ for $i = 1, \ldots, T-1$ and $j = 1, \ldots, T-1$, is calculated in a similar manner as the dual decision setting,

$$\sigma_{\boldsymbol{X}}^{ij} = \sigma_{\boldsymbol{\theta}}^{ij} - \frac{1}{Q}\sum_{q=1}^{Q} \sigma_q^{ij},$$

where $\sigma_{\boldsymbol{\theta}}^{ij}$ is the element of the i-th row and j-th column of the $\Sigma_{\boldsymbol{\theta}}$ matrix and σ_q^{ij} is the same element of the Σ_q matrices. This is therefore, the same formula as the dual-decision setting but separately for each element of the variance-covariance matrix.

To rescale the distribution of $\text{INB}^{\boldsymbol{\phi}}$, we use the same formula as in the standard method, but, rather than dividing by the standard deviation we must multiply by the inverse matrix square root. Matrix square roots and inverses are well-defined and can easily be found in `R` using the `expm` package. Therefore, to rescale the simulated PSA vectors for $\text{INB}^{\boldsymbol{\phi}}$, denoted $\text{INB}^{\boldsymbol{\phi}_s}$, $s = 1, \ldots, S$, we use the following formula:

$$\eta_s^{\boldsymbol{X}} = (\text{INB}^{\boldsymbol{\phi}_s} - \mu)\Sigma_{\boldsymbol{\phi}}^{-\frac{1}{2}}\Sigma_{\boldsymbol{X}}^{\frac{1}{2}} + \mu.$$

The EVSI is then calculated by taking the row-wise maximum of each of the $\eta_s^{\boldsymbol{X}}$ vectors and 0 and then taking the mean of these maximums.

Using non-linear regression to calculate the EVSI in multi-decision problems involves an extension to the standard method. The non-linear model defined in the main paper is a scalar function and we must estimate a variance-covariance matrix across different sample sizes. Therefore, we must extend this regression model to a matrix function. In practice, we suggest that the non-linear regression function is extended by fitting the non-linear model from main paper

$$f(N) = \sigma^2_{\boldsymbol{\phi}} \frac{N}{N+h}$$

separately for each unique element of the variance-covariance matrix. Essentially, this involves estimating the element in the i-th row and j-th column of the variance matrix $\Sigma_{\boldsymbol{X}}(N)$ for a sample size N, denoted $\sigma^{ij}_{\boldsymbol{X}}(N)$ as

$$f^{ij}(N) = \sigma^{ij}_{\boldsymbol{X}}(N) = \sigma^{ij}_{\boldsymbol{\phi}} \frac{N}{N+h^{ij}},$$

where $\sigma^{ij}_{\boldsymbol{\phi}}$ is the i-th, j-th element of the $\Sigma_{\boldsymbol{\phi}}$ matrix. It is possible to demonstrate in a decision model with three treatment options that these functions approximately estimate the variance of both incremental net benefits and their covariance in normal-normal conjugate settings.

As the covariance matrix is symmetric, $f^{ij}(N) = f^{ji}(N)$ and so we fit

$$\frac{(T-1)(T-2)}{2}$$

regression models separately to calculate the EVSI across different sample sizes, where T is the number of possible interventions. Each curve will produce an estimate for the h^{ij} parameter and these distributions can be combined to find the posterior distribution of the variance-covariance matrix for $\Sigma_{\boldsymbol{X}}(N)$ for each sample size N under consideration.

Finally, recall that for the dual-decision setting presented in the paper, the distribution of $\sigma^2_{\boldsymbol{X}}(N)$ was approximated by a low-dimensional summary for each sample size considered. In the multi-decision setting, this low-dimensional representation of the is more challenging to conceive. One suggestion is to summarise each posterior for h^{ij} by finding the median. It is then possible to create a variance-covariance matrix which contains the median value of h^{ij} from each posterior combined with a sample size of interest. This "median" variance-covariance matrix could then used to rescale $\text{INB}^{\boldsymbol{\phi}}$ using the formula above. The EVSI is then calculated from the rescaled $\text{INB}^{\boldsymbol{\phi}}$ values using standard formuæfor the EVSI in multi-decision settings.

Repeating the creation of matrices for alternative quantiles of the posterior distribution would give a low-dimensional summary of the possible EVSI values. For example, you could create a variance-covariance matrix that contains all the 2.5^{th} quantiles from each posterior for h^{ij} and then use this to calculate the EVSI. Coupled with a similar matrix created with the 97.5^{th} quantiles would give the EVSI for a 95% credible interval.

3. Calculating the EVSI using Moment Matching for the Brennan and Kharroubi Example

This model has frequently been used to assess calculation methods for the EVSI. It was first developed by Brennan and Kharroubi [3] and modified by Menzies [5] to compare two treatments used to treat a hypothetical disease. For each drug, a patient can respond to the treatment, experience side effects or visit hospital for a certain length of time. A utility value is assigned to each of these possible outcomes and costs are associated with the drugs and hospital stays.

All the parameters are assumed to be normal with the mean and standard deviation given in Table 1. The studies are also assumed to have normal distributions, with the standard deviations given in Table 1. In this example, it is assumed that $\theta_5, \theta_7, \theta_{14}$ and θ_{16} are correlated with correlation coefficient 0.6 and the parameters θ_6 and θ_{15} are also correlated with a correlation coefficient 0.6 and independent of the other set of parameters.

The net benefits for each treatment are calculated as a deterministic function of these parameters

$$\text{NB}_1 = \lambda(\theta_5\theta_6\theta_7 + \theta_8\theta_9\theta_{10}) - (\theta_1 + \theta_2\theta_3\theta_4),$$

Parameter	Mean		Standard Deviation (SD)		Data SD	
	$t=0$	$t=0$	$t=0$	$t=1$	$t=0$	$t=1$
Drug Cost (θ_1, θ_{11})	\$10 000	\$15 000	\$10	\$10	-	-
Probability of Hospitalisation (θ_2, θ_{12})	0.1	0.08	0.02	0.02	-	-
Days in Hospital (θ_3, θ_{13})	5.2	6.1	1	1	-	-
Hospital Cost per Day (θ_4)	\$4 000	\$4 000	\$2 000	\$2 000	-	-
Probability of Responding (θ_5, θ_{14})	0.7	0.8	0.1	0.1	0.2	0.2
Utility Change due to Response (θ_6, θ_{15})	0.3	0.3	0.1	0.05	0.2	0.2
Duration of Response (years) (θ_7, θ_{16})	3	3	0.5	1	1	2
Probability of Side Effects (θ_8, θ_{17})	0.25	0.2	0.1	0.05	-	-
Utility Change due to Side Effects (θ_9, θ_{18})	-0.1	-0.1	0.02	0.02	-	-
Duration of Side Effects (years) (θ_{10}, θ_{19})	0.5	0.5	0.2	0.2	-	-

Table 1: The parameters for the Brennan and Kharroubi example. The mean and standard deviations for the distributions of the parameters is also given, along with the standard deviation of the data collection exercise aimed at reducing uncertainty in that parameter

$$\text{NB}_2 = \lambda(\theta_{14}\theta_{15}\theta_{16} + \theta_{17}\theta_{18}\theta_{19}) - (\theta_{11} + \theta_{12}\theta_{13}\theta_4),$$

with $\lambda = \$100\,000$. Five alternative data collection exercises are proposed by Menzies and are also considered in this exploration:

1. A clinical trial collecting information on the probability that a patient responds to the two treatment options which informs parameters θ_5 and θ_{14}.
2. A study looking at the utility improvement for responding to the different treatments which informs parameters θ_6 and θ_{15}.
3. A study investigating the duration of response to the therapy (for those who do respond), informing parameters θ_7 and θ_{16}.
4. A study combining the first two studies, i.e. informing $\theta_5, \theta_6, \theta_{14}$ and θ_{15}.
5. A study combining all the previous studies and therefore informing $\theta_5, \theta_6, \theta_7, \theta_{14}, \theta_{15}$ and θ_{16}.

3.1. Analysis for the BK example

To estimate the EVSI for different sample sizes using the moment matching method, the PSA distribution for the incremental net benefit is estimated using 1 million simulations from the parameter distributions. This implies that σ^2 and μ, the variance and mean of the incremental net benefit respectively, are estimated using this full sample. INB^{Φ} are also found using these 1 million simulations, expect for exercise 5 which is based on 6 underlying parameters meaning that the computational demands of estimating INB^{Φ} was too high. These fitted values are, therefore, based on 20 000 simulations and obtained using the `R` package `BCEA` [1, 2].

In line with Menzies [5], sample sizes between $N_{\min} = 10$ and $N_{\max} = 200$ are considered for each of the different exercises outlined above. Throughout the analysis, we set $Q = 50$ which implies that 10 000 simulations are taken from 50 different posterior distributions to calculate the variance of the posterior incremental net benefit for 50 different sample sizes. The distribution for the EVSI is then determined using the method described in §**??** in the main paper.

The results determined using our method are compared with the nested Monte Carlo approach for calculating the EVSI and Menzies' approach which also reweights the PSA simulations for the INB but with an alternative method. These results are taken directly from Menzies [5] and are the most accurate estimates available. The conventional approach required 1 billion model evaluations per sample size compared with 500 000 model evaluations for the moment matching method to estimate the EVSI across the different sample sizes.

3.2. Results for the BK example

Figure 1 shows the EVSI estimates for the BK example. The solid line gives the EVSI calculated with the median of the posterior distribution of $\sigma^2_{\boldsymbol{X}}(N)$, whereas the dashed line is the 75% credible interval and the dotted line the 95% credible interval. The EVSI estimates from the nested Monte Carlo estimator and the Menzies estimator are given by the red dots and the blue crosses respectively. The nested Monte Carlo estimator (representative of the "true" EVSI) is within the 95% credible interval for all exercises except exercise 5 (bottom), where the EVSI is slightly over estimated for small values of N. This small over-estimation may be due to the inaccuracies introduced by estimating INB^{ϕ} using only 20 000 observations, as opposed to the full PSA simulation used for the other examples.

Figure 1 demonstrates that the EVSI is estimated with more relative precision as the EVSI estimate increases. This is because, for small values of the EVSI, the difference between σ^2_q and σ^2 is small and the estimate of the two variances needs to be very accurate in order to estimate the difference. Therefore, this method for the EVSI calculation should be reserved for situations where the underlying parameters have significant value. If the EVSI estimate is too variable to aid decision making, as seen by the confidence bands, then more simulation should be undertaken. In general, extra simulations should be gained by increasing Q, provided the number of posterior simulations is sufficient to characterise the distribution of the "posterior" incremental net benefit [4].

References

[1] G. Baio, A. Berardi, and A. Heath. *BCEA: Bayesian Cost Effectiveness Analysis*, 2016. R package version 2.2-3.

[2] G. Baio, A. Berardi, and A. Heath. *Bayesian Cost Effectiveness Analysis with the `R` package `BCEA`*. Springer, 2017.

[3] A. Brennan and S. Kharroubi. Efficient computation of partial expected value of sample information using bayesian approximation. *Journal of health economics*, 26(1):122–148, 2007.

[4] A. Heath and G. Baio. Calculating the Expected Value of Sample Information Using Efficient Nested Monte Carlo: A Tutorial. *Value in Health*, -(-):–, 2018.

[5] N. Menzies. An efficient estimator for the expected value of sample information. *Medical Decision Making*, 36(3):308–320, 2016.

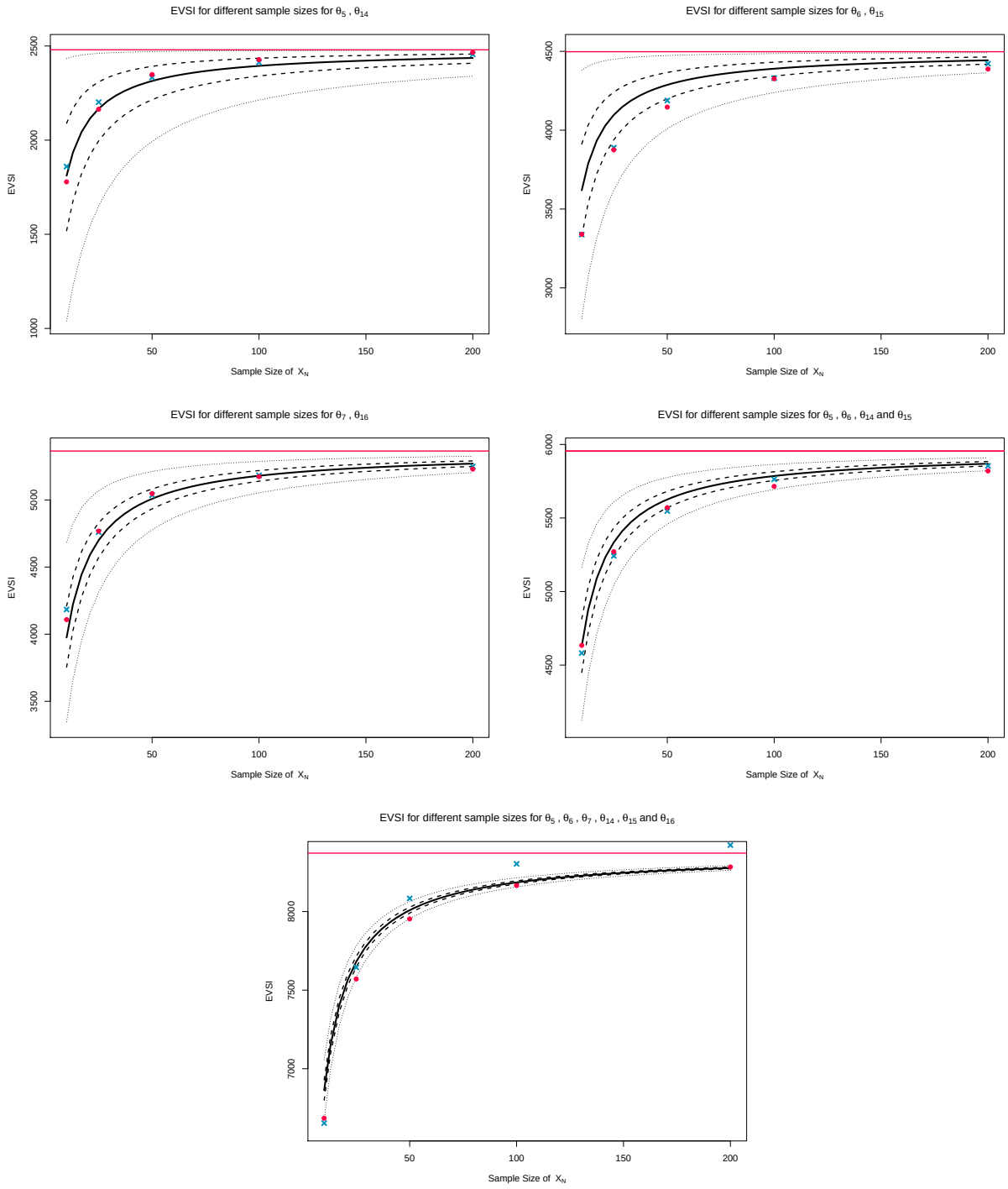

Figure 1: The EVSI estimate calculated with the posterior median of $\sigma^2_{\boldsymbol{X}}$ (solid line) along with 75% (dashed line) and 95% (solid line) posterior credible intervals for the BK example; Study 1 (Top Left), Study 2 (Top Right), Study 3 (Middle Left), Study 4 (Middle Right) and Study 5 (Bottom). These are compared with the nested Monte Carlo estimates (red dots) and the Menzies estimates (blue crosses). The EVPPI for the parameters targeted by the study is shown as the horizontal red line.