Estimating the distribution of the preposterior mean using moment matching

Anna Heath, Ioanna Manolopoulou and Gianluca Baio

Department of Statistical Science, University College London, London

Abstract

Preposterior analysis covers a wide range of approaches in many different disciplines and relates to any analysis concerned with understanding the properties of a future posterior distribution *before* relevant data have been collected. Embedding preposterior analysis in a decision making context implies that we are interested in the hypothetical values of the posterior mean utility for different decision options, known as the distribution of the preposterior mean. The distribution of the preposterior mean utility can be summarised with the Expected Value of Sample Information (EVSI) and used to aid funding decisions for future data collection exercises, which is especially important in health economic decision making. We present a method for estimating the distribution of the preposterior mean utility with a small number of hypothetical data samples using moment matching and samples from the prior of the utility. We discuss why this methodology is likely to accurately approximate the distribution of interest. We also suggest some settings where the moment matching method will be less accurate, particularly for small sample sizes of the future data collection exercise. We then illustrate the success of this methodology by calculating the EVSI in four varied examples, including a practical example from the health economic literature.

1   Introduction

Preposterior analysis encompasses a large suite of approaches that are concerned with estimating the properties of a posterior distribution *before* relevant data have been collected. These approaches are used in many different domains, from model calibration (Arendt et al., 2016; Jiang et al., 2015), to model checking (Ben-Zvi et al., 1988) and experimental design (Chaloner and Larntz, 1989; Erkanli and Soyer, 2000; Huang and Wu, 2008; Weaver et al., 2016). Using these analyses for Bayesian experimental design involves determining a data collection exercise that "optimises" in some sense a property of the posterior distribution — for example, minimising the posterior variance.

Another interesting application of preposterior analysis is to embed it within a formal decision making process — how would a future data set affect the decision? This naturally leads to the *value of information* (VoI) framework (Howard, 1966), which compares the decision based on current evidence to a decision made with "more information". According to the precepts of decision theory (Raiffa and Schlaifer, 1961), the value of each possible decision $t = 0, \ldots, T$ is quantified using a utility function that is typically conditional on some underlying model parameters $\boldsymbol{\theta}$, which are subject to uncertainty.

The optimal decision given the current level of uncertainty in $\boldsymbol{\theta}$ is the decision option associated with the highest expected utility. From the decision-theoretic point of view, the identification of the maximum expected utility is all that is required to reach the optimal decision given the current state of knowledge available to the decision-maker (Lindley, 2006). In general, "more information", typically gained from an additional data collection exercise, will decrease uncertainty for the decision makers and may even change the optimal decision. If this is the case and a sample does change the optimal decision, then it has a *value* to the decision maker as it prevents them from wasting resources on a non-optimal decision.

As the decision making process is concerned solely with the *expected* utility of the different decision options, our principal interest is the preposterior *mean*, conditional on a future data collection exercise. The analysis of the preposterior mean utility to obtain the "value" of a data collection exercise was first introduced by Schlaifer (1959) and extended by Raiffa and Schlaifer (1961) under the heading of *Expected Value of Sample Information* (EVSI).

While the concept of the EVSI is relatively old, it has rarely been used in formal decision making due to the immense computational burden required to estimate it using nested Monte Carlo simulations (Brennan et al., 2007). These nested simulations are required as, in general, hypothetical posterior means must be estimated by simulation for a large number of different future samples. To combat this, methods have been developed to take advantage of conjugate families and thus remove the need for nested simulations (Schlaifer, 1959; Ades et al., 2004; Brennan et al., 2007). However, these methods are limited in scope and have consequently prevented the usage of EVSI as a tool for decision making. Therefore, the aim of the current work is to approximate the distribution of the preposterior mean for a general model structure in order to calculate the EVSI for a broader set of possible decisions.

In this paper, we approximate the distribution of the preposterior mean using moment matching. To achieve this, we demonstrate that it is possible to accurately estimate the mean and variance of the preposterior mean using a small number of future posterior samples. We then suggest that the distribution of the preposterior mean is *similar* to the prior distribution for the quantity of interest provided the sample size of the data collection exercise is suitably large and can therefore be transformed to approximate the distribution of the preposterior mean over all future samples. This transformation translates and scales the prior distribution so it "matches" the moments of the preposterior mean.

We discuss different scenarios where this matching is successful and unsuccessful at approximating the EVSI. We compare our methodology with analytic results, where they are available, simulation based approaches and a new EVSI estimation procedure which is based on non-parametric regression and sufficient statistics (Strong et al., 2015). Specifically, we demonstrate that in these examples, our moment matching approach is successful in most settings where the sample size of the future data collection exercise is sufficiently large.

Throughout the paper, we focus on applications in health economic evaluation as VoI analysis is an increasingly popular method of assessing the uncertainty in health economic decisions (Felli and Hazen, 1998, 1999; Claxton, 1999; Claxton et al., 2001; Ades et al., 2004; Brennan and Kharroubi, 2005; Briggs et al., 2006; Fenwick et al., 2006), with preposterior analysis and the calculation of the EVSI becoming more widespread (Welton et al., 2014; Welton and Thom, 2015). In this light, §2 introduces the health economic context of the examples along with some notation and key statistical concepts. In §3, we discuss the theoretical grounding of our estimation method for the distribution of the preposterior mean utility before discussing how to find the approximate distribution by simulations in §4. Finally, we demonstrate the success of our method using different examples in §5.

2   Notation and Concepts

The discussion of our methodology begins with an introduction of the concepts and notation that will be used throughout the paper. In general, health economic models are characterised by a large number of parameters $\boldsymbol{\theta}$ whose distributions are based on the current evidence base from literature reviews, possible clinical trials or meta analyses. Due to the complexity of these models, health economic analysis normally uses a simulation based approach (Baio and Dawid, 2011; Baio, 2012; Andronis et al., 2009), in which S values of the model parameters are simulated, e.g. via MCMC, to fully characterise the uncertainty in the economic analysis under current information.

Typically, these parameter distributions will be informed by past data, e.g. in the form of a previously conducted trial, and would be denoted as $p(\boldsymbol{\theta} \mid \mathcal{D})$, where $\mathcal{D}$ indicates existing data used to estimate the parameters $\boldsymbol{\theta}$. However, as we are considering collecting a future data set, this data $\mathcal{D}$ is obtained *prior* to our investigation. Therefore, throughout, simulations from $p(\boldsymbol{\theta} \mid \mathcal{D})$, denoted $\boldsymbol{\theta}_s$, will be referred to as simulations from the *prior* for the parameters. We will also drop the dependency on the past data $\mathcal{D}$ for notational simplicity so $p(\boldsymbol{\theta})$ represents the prior for the parameters.

In health economic evaluations, $p(\boldsymbol{\theta})$ informs distributions for the cost and effectiveness of a treatment (c, e). These measures reflect both uncertainty in the parameter values and individual level variability, e.g. different responses to a specific drug. Commonly, the *utility* of a treatment is calculated using the *monetary net benefit* (Stinnett and Mullahy, 1998) which assigns a value for

each decision, or treatment option, $t = 0, \ldots, T$, defined by:

$$U_t(\boldsymbol{\theta}) = k\mathrm{E}[e \mid \boldsymbol{\theta}; t] - \mathrm{E}[c \mid \boldsymbol{\theta}; t],$$

where the expectation is taken over *individual* level uncertainty only. In the above expression, k is the *willingness to pay*, which is used to put the cost and effectiveness measures on the same scale, i.e. in terms of the amount of money that the decision maker is willing to pay to increment the benefit by one unit.

Typically, under the simulation based approach, we have access to a vector of values $[U_t(\boldsymbol{\theta}_1), \ldots, U_t(\boldsymbol{\theta}_S)]$ which is a sample from the *prior* distribution of the net benefit for each decision option. Uncertainty in this distribution is driven directly by the current level of parameter uncertainty, quantified using the prior distribution for the parameters $p(\boldsymbol{\theta})$.

For notational simplicity throughout the paper, we denote the utility, typically referred to as the net benefit, as $\mathrm{NB}_t^{\boldsymbol{\theta}}$, where the superscript $\boldsymbol{\theta}$ indicates that the utility is a *function* of $\boldsymbol{\theta}$. We will also continue to use S as the *prior* simulation size. Finally, the notation $p(\cdot)$ will be used to indicate any *density* function. The argument of the function then determines which density is being discussed. Therefore, the functions $p(\boldsymbol{\theta})$ and $p(\mathrm{NB}_t^{\boldsymbol{\theta}})$ are different and give the density function of the random variables $\boldsymbol{\theta}$ and $\mathrm{NB}_t^{\boldsymbol{\theta}}$ respectively.

2.1   Sampling Strategy

We are interested in calculating the EVSI for a specific data collection exercise. Throughout this paper we will use $\boldsymbol{X}$ to denote theoretical future samples from that data collection exercise. For example, $\boldsymbol{X}$ could be the number of people who survive an infection after 10 days in a future clinical trial. In a standard Bayesian analysis, where $\boldsymbol{X}$ is a realised sample, the posterior distribution of the parameters is obtained by combining the prior $p(\boldsymbol{\theta})$ and the model for the data $p(\boldsymbol{X} \mid \boldsymbol{\theta})$.

As $\boldsymbol{X}$ is a random variable, the conditional distribution $p(\boldsymbol{X} \mid \boldsymbol{\theta})$ gives the relationship between the future sample and the parameters. The marginal distribution for the future samples is given by the prior predictive distribution – again noting that *prior* in this setting means "prior to the new sample" and may indeed be conditional on past data sets:

$$p(\boldsymbol{X}) = \int_{\Theta} p(\boldsymbol{X} \mid \boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}.$$

In most settings, the data collection exercise will consist of data about N different individuals. For example we would record the survival time of N different patients in the clinical trial. Throughout the paper, N will be referred to as the *sample* size and is reserved for discussion about the future data collection exercises. This contrasts directly with *simulation* size S which is reserved for discussion about simulating from the prior or posterior for $\boldsymbol{\theta}$. This distinction allows us to discuss S simulations of sample size N which relates to S simulations from the prior predictive distribution $p(\boldsymbol{X})$ of a data collection exercise with sample size N.

At this point, it is also important to note that the design of the future data collection exercise is irrelevant to the methodology we will present. The design of a future study can have limitless variations including, but certainly not limited to, variation in sample size. For example, data collection exercises could vary in distribution, levels of missingness and follow-up time, to name but a few.

2.2   The distribution of the preposterior mean

The *distribution of the preposterior mean* is the distribution over the possible future values for the posterior mean *before* the data have been collected. In a standard Bayesian analysis, performed after the data have become available and observed to the value $\boldsymbol{x}$, the posterior mean will simply be a number (or vector for multivariate distributions), with the value of the mean conditional on those data $\boldsymbol{x}$. Therefore, the distribution of the posterior mean *before* collecting the data is a random variable for a given $p(\boldsymbol{X})$.

As in standard Bayesian analysis, the posterior mean is given by

$$\mu_t^{\boldsymbol{X}} := \mathrm{E}_{\boldsymbol{\theta}|\boldsymbol{X}}\left[\mathrm{NB}_t^{\boldsymbol{\theta}}\right] = \int_{\boldsymbol{\Theta}} \mathrm{NB}_t^{\boldsymbol{\theta}}\, p(\boldsymbol{\theta} \mid \boldsymbol{X})d\boldsymbol{\theta} = \int_{\boldsymbol{\Theta}} \mathrm{NB}_t^{\boldsymbol{\theta}}\, \frac{p(\boldsymbol{X} \mid \boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\boldsymbol{X})}d\boldsymbol{\theta}.$$

The only difference for preposterior analysis is that $\boldsymbol{X}$ is a random variable rather than an observed data set $\boldsymbol{x}$. Throughout the paper, we use the notation $\mu_t^{\boldsymbol{X}}$ to denote the preposterior mean, for decision $t = 0, \ldots, T$ highlighting that it is a function of $\boldsymbol{X}$.

2.3   Expected Value of Sample Information

As discussed in §1, the EVSI compares the optimal decision given the current level of uncertainty to a decision made with "more information". The decision making process under current information is solely concerned with finding the treatment option with the maximum expected net benefit. In other words, the "value" of the current decision making process is simply

$$\max_t \mathrm{E}_{\boldsymbol{\theta}}\left[\mathrm{NB}_t^{\boldsymbol{\theta}}\right],$$

i.e. the value of the "optimal" decision based on the current information about $\boldsymbol{\theta}$.

If the future sample had been collected, then the optimal decision based on that sample would be the treatment option with the maximum expected net benefit

$$\max_t \mathrm{E}_{\boldsymbol{\theta}|\boldsymbol{X}=\boldsymbol{x}}\left[\mathrm{NB}_t^{\boldsymbol{\theta}}\right] = \max_t \mu_t^{\boldsymbol{x}}.$$

However, as the data have not been observed yet, the expectation over all possible future samples is taken to give the average value of the decision made with the additional information in the sample. The EVSI is then given by

$$\mathrm{EVSI} = \mathrm{E}_{\boldsymbol{X}}\left[\max_t \mu_t^{\boldsymbol{X}}\right] - \max_t \mathrm{E}_{\boldsymbol{\theta}}\left[\mathrm{NB}_t^{\boldsymbol{\theta}}\right].$$

Notice that this calculation involves finding the treatment with the largest posterior mean net benefit for each sample. Therefore, to calculate this value, it is necessary to find the distribution of the joint preposterior mean across all the alternative treatment options.

2.4   Examples of the distribution of the preposterior mean

To illustrate these concepts we introduce a simple example using the Beta-Binomial conjugate family. Assume that a new drug is available and is associated with a probability θ of curing a particular disease. As this drug is new, it is assumed that there is very limited evidence on its effectiveness. This assumption could be expressed, in a very simplistic way, by modelling $\theta \sim Beta(1,1)$. We note, however, that it is likely that some information would be available, in practical settings (e.g. from small trials), so a more informative prior could be used instead. The effectiveness measure is whether the disease has been cured, meaning that the population level effectiveness is θ. Assume further that the drug costs c, where c is known, and that the willingness to pay is some constant k.

The other possible treatment option is to do nothing. This has no cost and no effectiveness as this (non-life-threatening) disease does not improve without drug intervention. This implies that the two net benefit values are

$$\text{NB}_0^\theta = 0 \quad \text{and} \quad \text{NB}_1^\theta = k\theta - c.$$

The future experiment is to give N people the drug and observe how many are cured. This can be expressed using a binomial distribution for the future sample $X \mid \theta \sim Bin(N, \theta)$. The prior predictive distribution in this setting is given by

$$p(X) = \int_0^1 \binom{N}{X} \phi^X (1-\phi)^{N-X} d\phi = \mathcal{B}(X+1, N-X+1)\frac{N!}{X!(N-X)!}$$
$$= \frac{X!(N-X)!N!}{(N+1)!X!(N-X)!} = \frac{1}{N+1},$$

where $\mathcal{B}(\cdot, \cdot)$ is the Beta function. This implies that all samples for X are equally likely, due to our choice of prior.

Once we know the prior predictive distribution, the distribution of the pre-posterior mean for the two treatment options is determined by calculating the posterior mean (conditional on X) for both the net benefit functions. Obviously, the posterior mean for NB_0^θ is $\mu_0^X = 0$ and therefore the distribution of the preposterior mean is simply a point mass at 0. However, the posterior mean for NB_1^θ does depend on the future value of the random variable X:

$$\mu_1^X = \text{E}_{\theta|X}\left[\text{NB}_1^\theta\right] = \int_0^1 (k\theta - c)\, p(\theta \mid X) d\theta$$
$$= k\frac{1+X}{2+N} - c,$$

as $\theta \mid X \sim Beta(1 + X, 1 + N)$. Therefore, the distribution of μ_1^X is conditional on the uniform prior predictive distribution for X, which in turn is conditional on our uniform prior for θ. This means that the distribution of the preposterior mean is uniform over all possible value for μ_1^X, calculated as a function of the $N + 1$ possible X values.

The EVSI can then be used to summarise $\mu_0^{\mathbf{X}}$ and $\mu_1^{\mathbf{X}}$ and, dependent on the values of k, c and N, give a upper limit to the value for the future data collection exercise. We note that this trial has value as the future sample may change our optimal treatment as a future sample indicates that $t = 1$ (the new drug) is optimal if $\dfrac{1 + X}{2 + N} > \dfrac{c}{k}$ and that $t = 0$ (doing nothing) is optimal otherwise. As an example the EVSI, for $k = 20\,000$, $c = 10\,000$ and $N = 5$, can be calculated exactly as:

$$\text{EVSI} = \left(0 + 0 + 0 + \frac{10\,000}{7} + \frac{30\,000}{7} + \frac{50\,000}{7}\right)\frac{1}{6} - 0 = \frac{15\,000}{7} = 2\,142,$$

which is then compared with the cost of a trial with 5 participants to determine whether that trial would be worth funding.

2.4.1 Distribution of preposterior mean for Exponential-Gamma conjugacy

In this second example, a Gamma prior is assumed for the parameter of interest $\theta \sim Gamma(\alpha, \beta)$. The data collection exercise is then assumed to be N independent observations from an exponential distribution conditional on θ, $X_j \sim Exp(\theta)$ with $j = 1, \ldots, N$. We consider the distribution of the preposterior mean for different values of N, where the two net benefit functions are:

$$\text{NB}_0^\theta = c_0 \quad \text{and} \quad \text{NB}_1^\theta = k\theta - c_1.$$

Figure 1 presents the distribution of the preposterior mean for NB_1^θ for $\alpha = 5$, $\beta = 1$, $k = 200$, $c_0 = 900$ and $c_1 = 100$.

Note that the distribution of the preposterior mean gets closer to the prior as the sample size increases for the data collection exercise. The distribution of the preposterior mean also has a larger variance as the sample size increases. These two properties are at odds with the intuition that the distribution of the mean would get more concentrated and close to a normal as the sample size increases. Nonetheless, these counterintuitive results hold because the "strength" of the data increases as the sample size increases and so the posterior mean can deviate further from the prior mean implying that the variance of the preposterior mean does increase as the sample size increases. In addition to this, at the current state of knowledge, i.e. *before* the data collection has taken place, it is not possible to learn anything additional about the parameters. Therefore, if a future data collection exercise gives exact information about the parameter location then the distribution of possible posterior means is equal to distribution of possible parameter values i.e. the prior distribution. To clarify these ideas further, a normal-normal conjugate example is given in the supplementary material.
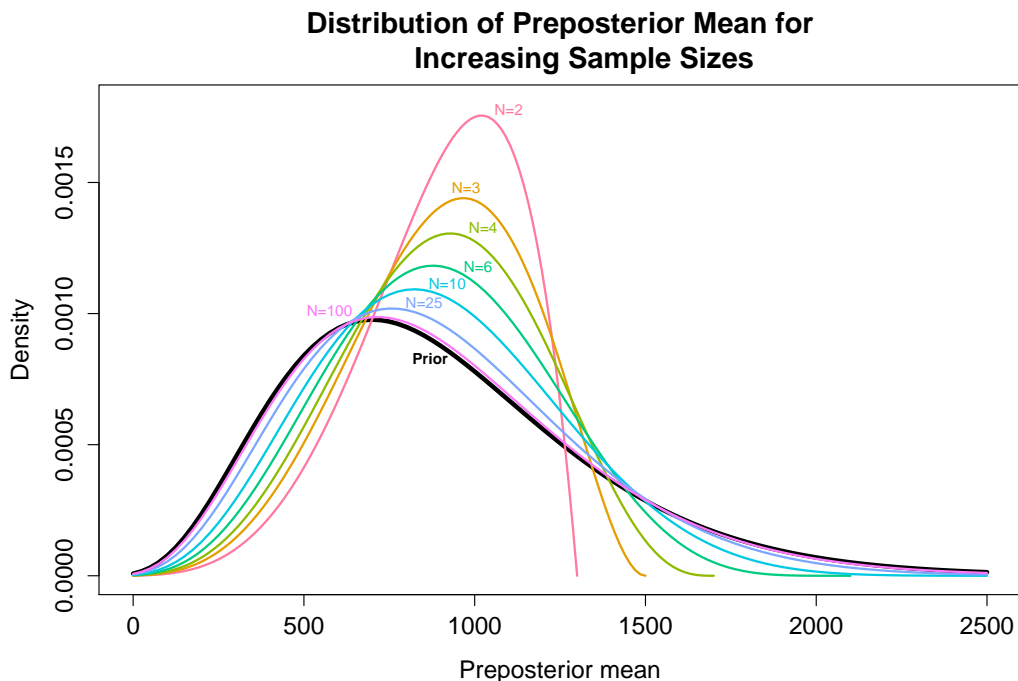
Figure 1: The distribution of the exact preposterior mean for different samples sizes using Exponential-Gamma conjugacy, with the prior for the net benefit marked in black.

3 Estimating the preposterior mean distribution

Historically, it has been suggested that the distribution of the preposterior mean should be estimated by Monte Carlo simulation (Pratt et al., 1995), specifically in the health economic literature (Brennan et al., 2007; Ades et al., 2004). This involves simulating a large number of draws from $p(\boldsymbol{X})$. For each sample, the posterior is then updated, either using conjugate models or by MCMC simulation, and used to estimate the posterior mean. Our methodology presented below reduces the number of simulations required from $p(\boldsymbol{X})$ by exploiting the information available in the prior.

Throughout, we have been concerned solely with the *distribution* of the preposterior mean which, in §2.4, was computed by finding the prior predictive distribution $p(\boldsymbol{X})$ and the functional relationship between the preposterior mean and the sample $\boldsymbol{X}$. However, if the prior predictive distribution is not known, as in most practical situations, then a known functional form of the $\mu_t^{\boldsymbol{X}}$ cannot be used to determine the distribution of the preposterior mean. Therefore, while other EVSI estimation methods have focused on estimating a functional form for the expected net benefit *conditional* on the future sample (Strong

8

et al., 2015; Ades et al., 2004), we focus solely on estimating the *distribution* of the preposterior mean. Therefore, we are simply concerned with estimating a probability density, for which there is a large wealth of statistical theory.

3.1 Expectation and Variance for the preposterior mean

To approximate the probability density of the preposterior mean, we begin by estimating its mean and variance. In this analysis, the interest lies with the expectation and variance *conditional* on the value of $\boldsymbol{X}$, implying that standard formulæ for conditional iterated expectation can be used to calculate both the expectation and variance of the preposterior mean.

Therefore the mean of the distribution of the preposterior mean is given by

$$\mathrm{E}_{\boldsymbol{X}}\left[\mu_t^{\boldsymbol{X}}\right] = \mathrm{E}_{\boldsymbol{X}}\left[\mathrm{E}_{\boldsymbol{\theta}|\boldsymbol{X}}\left[\mathrm{NB}_t^{\boldsymbol{\theta}}\right]\right] = \mathrm{E}_{\boldsymbol{\theta}}\left[\mathrm{NB}_t^{\boldsymbol{\theta}}\right],$$

which implies that the expectation of the preposterior mean is equal to the prior mean. Thus, unsurprisingly, preposterior analysis does not give any additional information about the net benefit. On average, over all the expected samples (which are conditional on the prior beliefs), the expected net benefit is the same.

The variance of the preposterior mean has a more complex formula but can also be re-expressed using iterated expectation as

$$\mathrm{Var}_{\boldsymbol{X}}\left[\mu_t^{\boldsymbol{X}}\right] = \mathrm{Var}_{\boldsymbol{X}}\left[\mathrm{E}_{\boldsymbol{\theta}|\boldsymbol{X}}\left[\mathrm{NB}_t^{\boldsymbol{\theta}}\right]\right] = \mathrm{Var}_{\boldsymbol{\theta}}\left[\mathrm{NB}_t^{\boldsymbol{\theta}}\right] - \mathrm{E}_{\boldsymbol{X}}\left[\mathrm{Var}_{\boldsymbol{\theta}|\boldsymbol{X}}\left[\mathrm{NB}_t^{\boldsymbol{\theta}}\right]\right].$$

This means that the variance of the preposterior mean is equal to the variance of the prior distribution minus the expectation, over all the possible samples $\boldsymbol{X}$, of the posterior variance. Therefore, to calculate the variance of the preposterior mean distribution practically, the average posterior variance over all possible samples $\boldsymbol{X}$ must be estimated. However, §4.1 demonstrates that the average posterior variance can be estimated using a significantly reduced number of posterior samples compared to calculating the EVSI by simulation.

3.2 Moment Matching

Moment matching is a common method of performing parameter inference within a model but has recently been applied in the context of estimating an unknown density (Çetinkaya and Thiele, 2016; Feng et al., 2015). In general, an unknown distribution can be accurately characterised by a large set of moments. However, it can also be approximated using a known distribution and "matching" a small number of the moments. This means that an alternative family of distributions is chosen and then parameters are found to determine a distribution in this family with the same moments as the distribution of interest. In the simplest setting, this involves approximating the distribution of the preposterior mean by a Gaussian with the mean and variance calculated using the formulæ in (1).

However, this is unlikely to be sufficiently accurate for our purposes as the EVSI is strongly influenced by the tails of the distribution of the preposterior

mean. This is due to the fact that the optimal decision is most likely to be different from the current optimal decision in the tails as our current decision is, by definition, optimal for the majority of the prior mass. Therefore, the EVSI estimate will be significantly improved if the distribution of the preposterior mean is approximated using moment matching with an alternative distribution that is closer to that of the preposterior mean. In fact, we suggest that the prior distribution for the net benefit is similar enough to the true distribution of the preposterior mean to give a good approximation for the EVSI, specifically for larger sample sizes N. This is because, while a specific future sample would give additional information, the preposterior analysis (before the data are collected) cannot give any information in addition to that contained in the prior. At this point, it is worth reiterating that typically the *prior* is actually a posterior distribution conditional on data and is, therefore, more likely to contain useful information about the parameters and the net benefits.

In this sense, the unknown distribution of the preposterior mean net benefit is approximated by a distribution with the correct expectation and variance but all other distributional properties such as the skewness are determined from the prior $p(\text{NB}_t^{\boldsymbol{\theta}})$. This idea can be generalised to situations where the sample $\boldsymbol{X}$ is dependent on a subset of the underlying model parameters $\boldsymbol{\theta}$ (see §4.3) and although we acknowledge difficulties in some specific cases (§5.1), it is successful in many settings (§5).

3.2.1   Linear transformation to moment match

Practically, to "moment match" with the prior, the distribution of the preposterior mean is estimated by a shifted and rescaled version of the prior. This implies that a linear transformation of $\text{NB}_t^{\boldsymbol{\theta}}$ must be found such that $a\,\text{NB}_t^{\boldsymbol{\theta}} + b$ has the same mean and variance as the distribution of the preposterior mean:

$$\text{E}_{\boldsymbol{\theta}}\left[a\,\text{NB}_t^{\boldsymbol{\theta}} + b\right] = \text{E}_{\boldsymbol{\theta}}\left[\mu_t^{\boldsymbol{X}}\right] \Rightarrow a\text{E}_{\boldsymbol{\theta}}\left[\text{NB}_t^{\boldsymbol{\theta}}\right] + b = \text{E}_{\boldsymbol{\theta}}\left[\text{NB}_t^{\boldsymbol{\theta}}\right]$$

$$\text{Var}_{\boldsymbol{\theta}}[a\,\text{NB}_t^{\boldsymbol{\theta}} + b] = \text{Var}[\mu_t^{\boldsymbol{X}}] \Rightarrow a^2\text{Var}_{\boldsymbol{\theta}}[\text{NB}_t^{\boldsymbol{\theta}}] = \sigma^2,$$

where σ^2 is the variance of the preposterior mean distribution that can be written as a function of the prior variance and the expected posterior variance. Solving for a and b yields

$$a = \sqrt{\frac{\text{Var}_{\boldsymbol{X}}\left[\mu_t^{\boldsymbol{X}}\right]}{\text{Var}_{\boldsymbol{\theta}}\left[\text{NB}_t^{\boldsymbol{\theta}}\right]}} = \frac{\sigma}{\sqrt{\text{Var}_{\boldsymbol{\theta}}\left[\text{NB}_t^{\boldsymbol{\theta}}\right]}} \quad \text{and} \quad b = \text{E}_{\boldsymbol{\theta}}\left[\text{NB}_t^{\boldsymbol{\theta}}\right](1 - a), \quad (1)$$

which depend on the prior expectation, prior variance and expected posterior variance for the net benefit.

Interestingly, these constants allow for a relatively simple interpretation of the approximation of the density of the preposterior mean. The constant a can be thought of as the proportion of the variance in $\text{NB}_t^{\boldsymbol{\theta}}$ that is *explained* by the

future sample $\boldsymbol{X}$. This means that the more closely the sample $\boldsymbol{X}$ reflects the underlying $\boldsymbol{\theta}$ values, the higher the value of a.

The constant b in (1), however, is the prior (and preposterior) mean weighted by one minus this *explained variance*. This weight is directly related to how closely the sample reflects the underlying values of $\boldsymbol{\theta}$. Thus, the density of the preposterior mean is estimated as a convex combination of the prior for the net benefit and the mean of the net benefit.

As the sample size in the data collection exercise increases, then the sample $\boldsymbol{X}$ (or some summary measure of $\boldsymbol{X}$) reflects the underlying value of $\boldsymbol{\theta}$ more closely. In turn, this implies that the density of the preposterior mean reflects the prior for the net benefit more and more closely. This is property is required for any approximation to the distribution of the preposterior mean, as seen in Figure 1.

Given the linear transformation in (1), it is possible to write down our approximation for the density of the preposterior mean directly as a function of the density of the prior $p(\mathrm{NB}^{\boldsymbol{\theta}})$

$$p(\mu^{\boldsymbol{X}}) \approx \frac{1}{a} p\left(\frac{\mu^{\boldsymbol{X}} - b}{a}\right).$$

Note that this density does not approximate the function $\mu^{\boldsymbol{X}}$ itself but the *density* $p(\mu^{\boldsymbol{X}})$.

In general, samples from the prior of the net benefit are available or easy to obtain. Thus, approximating the density of the preposterior mean is simply a matter of estimating the constants a and b. However, as the prior mean and variance can be calculated from the available prior samples, the approximation of these two parameters reduces to estimating the expected variance of the preposterior mean. We return to this consideration in §4 but continue now with some validation of the proposed method.

3.2.2   Why does it work?

In order to provide more insight as to when it is possible to approximate the density of the preposterior mean using this method, we consider here two special cases. At one extreme, assume that $\boldsymbol{X}$ is independent of the underlying model parameters: $p(\boldsymbol{X} \mid \boldsymbol{\theta}) = p(\boldsymbol{X})$. Evidently, this setting would never occur as decision makers only consider data collection that would aid the decision making process. Nevertheless, if the sample is independent of the model parameters then the distribution of the preposterior mean is a point mass at the prior mean

$$\mathrm{E}_{\boldsymbol{\theta}|\boldsymbol{X}}\left[\mathrm{NB}_t^{\boldsymbol{\theta}}\right] = \mathrm{E}_{\boldsymbol{\theta}}\left[\mathrm{NB}_t^{\boldsymbol{\theta}}\right],$$

by the condition of independence.

Using the definition for a and b from (1), note that

$$a = \sqrt{\frac{\mathrm{Var}\left[\mathrm{E}_{\boldsymbol{\theta}|\boldsymbol{X}}\left[\mathrm{NB}_t^{\boldsymbol{\theta}}\right]\right]}{\mathrm{Var}_{\boldsymbol{\theta}}\left[\mathrm{NB}_t^{\boldsymbol{\theta}}\right]}} = 0$$

and

$$b = \mathrm{E}_{\boldsymbol{\theta}}\left[\mathrm{NB}_t^{\boldsymbol{\theta}}\right](1-a) = \mathrm{E}_{\boldsymbol{\theta}}\left[\mathrm{NB}_t^{\boldsymbol{\theta}}\right],$$

which means that the approximation for the density of the preposterior mean is also equal to the prior mean and therefore exact for all model structures when the sample and the model parameters are independent. While, practically, this is a relatively unimportant result, it does indicate that our approximation is roughly accurate when the variance of the preposterior mean is small.

At the other end of the scale, it is possible to show that our approximation using moment matching is exact when the sample is deterministically linked to the model parameters, i.e. $\boldsymbol{X} = h(\boldsymbol{\theta})$ for some $h(\cdot)$. In this setting the *conditional* mean for the net benefit is equal to the net benefit since, if the value for $\boldsymbol{X}$ is known, then the exact $\mathrm{NB}_t^{\boldsymbol{\theta}}$ value is also known. In a similar manner to above it can be shown that $a = 1$ and $b = 0$ so the approximation for the density of the preposterior mean is equal to $\mathrm{NB}_t^{\boldsymbol{\theta}}$ which, again, is the exact distribution for the preposterior mean.

Therefore, as the variance of the preposterior mean increases, the approximation becomes exact. This has a practical implication since, provided the posterior is consistent, this approximation is accurate for large sample sizes. This is because, as the sample size N increases, the distribution of the preposterior mean reflects the prior more closely (Figure 1) as the sample contains more information about the underlying values for $\boldsymbol{\theta}$.

To extend these ideas, this approximation is accurate for moderate N when the posterior mean net benefit is a weighted average between the prior mean and a data summary

$$\mathrm{E}_{\boldsymbol{\theta}|\boldsymbol{X}}\left[\mathrm{NB}_t^{\boldsymbol{\theta}}\right] = c\,\mathrm{E}_{\boldsymbol{\theta}}\left[\mathrm{NB}_t^{\boldsymbol{\theta}}\right] + d\,g(\boldsymbol{X}),$$

where c and d are constants and $g(\cdot)$ is an arbitrarily complex function of the data which must have a similar density to $\mathrm{NB}_t^{\boldsymbol{\theta}}$.

For all conjugate settings in the exponential family the posterior mean can be written as weighted average of the prior mean and a data summary (Diaconis and Ylvisaker, 1979). Therefore, it is sufficient to consider whether the prior predictive distribution of the data summary has a similar distribution to the prior. In the simplest setting, it is possible to demonstrate that this is true in the normal-normal setting, as seen in the supplementary material, and therefore, our approximation will be accurate when the prior for the net benefit is approximately normal, coupled with an approximately normal distribution for $g(\boldsymbol{X})$. In §5, we demonstrate that the approximation can give biased estimates in non-normal settings. However, the bias is minimal for realistic sample sizes and decreases further as the sample size N increases since the preposterior distribution approaches the prior and the variance of the preposterior mean tends to the prior variance.

12

4 Approximating the preposterior mean distribution by simulation

As previously discussed, using the moment matching methodology reduces approximating the distribution of the preposterior mean to estimating the constants a and b from §3.2, under the assumption (usually true, in the context of health economic evaluation) that simulations from the distribution for $\mathrm{NB}_t^{\boldsymbol{\theta}}$ under current information are available. These constants are based on the mean and variance of the prior for the net benefit (i.e. prior to the future sample $\boldsymbol{X}$), and the expected posterior variance over all possible samples $\boldsymbol{X}$. Therefore, the following section is concerned with estimating the expected posterior variance using a *small* number of posterior samples. This reduces the number of posterior samples needed compared to the nested Monte Carlo simulation and therefore reduces the computational time required to approximate the distribution of the preposterior mean.

4.1 Estimating the variance of the preposterior mean

To begin, it may seem that estimating the expected posterior variance over different possible samples of $\boldsymbol{X}$ by Monte Carlo simulation would not save computational time compared to estimating the distribution of the preposterior mean by finding the posterior mean for different samples. However, in general, the posterior variance is relatively *stable* implying that the posterior variance changes relatively little across the different future samples $\boldsymbol{X}$ compared to the posterior mean. This stability is most extreme in the normal-normal conjugate setting where the posterior variance is independent of the posterior mean and dependent simply on the variance of the sample $\boldsymbol{X}$, not its location. Therefore, the posterior variance is the same for each future sample $\boldsymbol{X}$ for a fixed sample size N, implying that only one posterior sample is required to estimate the *expected* posterior variance which is then used to estimate the constant b (1).

With substantial departures from normality, the posterior variance is no longer independent of the location of the samples, but, as shown in §5.3, a small number of posterior samples (around 20-50) can be used to estimate the expected posterior variance, even in highly non-normal settings. This is because quadrature can be employed to reduce the number of hypothetical posterior variance estimates that are needed for this estimation.

Specifically, quadrature is employed to calculate

$$\mathrm{E}_{\boldsymbol{X}}\left[\mathrm{Var}_{\boldsymbol{\theta}|\boldsymbol{X}}\left[\mathrm{NB}_t^{\boldsymbol{\theta}}\right]\right] = \mathrm{E}_{\boldsymbol{\theta}}\left[\mathrm{E}_{\boldsymbol{X}|\boldsymbol{\theta}}\left[\mathrm{Var}_{\boldsymbol{\theta}|\boldsymbol{X}}\left[\mathrm{NB}_t^{\boldsymbol{\theta}}\right]\right]\right],$$

where the two outer expectations on the RHS allow us to take an expectation over the prior predictive distribution without direct sampling. We use quadrature to estimate the outer expectation with respect to $\boldsymbol{\theta}$ and Monte Carlo simulation across the different quadrature points for the inner expectation with respect to $\boldsymbol{X} \mid \boldsymbol{\theta}$.

In most data collection exercises, $\boldsymbol{X}$ will only be directly conditional on a small number of parameters as in general researchers are only directly interested in 1 or 2 parameters when designing a trial. Therefore, the outer expectation, with respect to $\boldsymbol{\theta}$ above, will typically be over a uni- or bi-variate vector, in which case quadrature is simple to implement. In §4.3, we further the discussion of situations where the sample is only dependent on a sub set of the model parameters.

In general, to perform quadrature, Q evenly spaced values in the domain of interest are selected and then used to simulate from $\boldsymbol{X} \mid \boldsymbol{\theta}$. These samples are then used to update the posterior and calculate the variance. As Monte Carlo simulations are used to estimate the inner expectation, we recommend in excess of 20 simulations to avoid the dependence on specific samples. Notice here that there is a clear trade-off between accuracy of the estimate for the variance of the preposterior mean and the computational time required to obtain this estimate.

4.2 Calculating the EVSI for a specific set of treatment options

Thus far, we demonstrated how to estimate the distribution of the preposterior mean for the net benefit. However, to calculate the EVSI, we need to compute the joint distribution of the preposterior means across the different treatments to find the dominant treatment. In general, this requires the estimation of the posterior variance-covariance matrix for the net benefits for the different treatments.

While this adds little theoretical complexity, this EVSI estimation method is more stable if we work directly with the incremental net benefit (INB); defined as the difference between two treatment options, e.g. $\mathrm{INB}^{\boldsymbol{\theta}} = \mathrm{NB}_1^{\boldsymbol{\theta}} - \mathrm{NB}_0^{\boldsymbol{\theta}}$. We can then find the optimal treatment by comparing the $\mathrm{INB}^{\boldsymbol{\theta}}$ with 0; if $\mathrm{INB}^{\boldsymbol{\theta}} > 0$ then treatment 1 ($t = 1$) is optimal and if it is negative $t = 0$ is optimal. Therefore, if only two treatment options are available and the distribution of the preposterior mean of the INB is estimated using moment matching, then the EVSI can be calculated using

$$\mathrm{E}_{\boldsymbol{X}} \left[\max \left\{ 0, \mathrm{E}_{\boldsymbol{\theta} \mid \boldsymbol{X}} \left[\mathrm{INB}^{\boldsymbol{\theta}} \right] \right\} \right] - \max \left\{ 0, \mathrm{E}_{\boldsymbol{\theta}} \left[\mathrm{INB}^{\boldsymbol{\theta}} \right] \right\},$$

where $\mu^{\boldsymbol{X}} = \mathrm{E}_{\boldsymbol{\theta} \mid \boldsymbol{X}} \left[\mathrm{INB}^{\boldsymbol{\theta}} \right]$ is only based on scalar mean and variance values rather than a mean vector and a variance-covariance matrix.

When more than two options are under consideration, it is also preferable to work with the INB as it reduces the size of the variance-covariance matrix, i.e. the EVSI can be calculated based on the variance-covariance matrix for the distribution of $\mathrm{NB}_1^{\boldsymbol{\theta}} - \mathrm{NB}_0^{\boldsymbol{\theta}}$ and $\mathrm{NB}_2^{\boldsymbol{\theta}} - \mathrm{NB}_0^{\boldsymbol{\theta}}$ rather than $\mathrm{NB}_0^{\boldsymbol{\theta}}$, $\mathrm{NB}_1^{\boldsymbol{\theta}}$ and $\mathrm{NB}_2^{\boldsymbol{\theta}}$. This reduces the number of parameters that need to be estimated as there are only 3 unique elements in the variance-covariance matrix rather than 6.

4.3 Nuisance Parameters

In realistic health economic models it is very unlikely that the proposed data collection exercise will depend directly on all the model parameters $\boldsymbol{\theta}$. For example, a clinical trial may only give information about the drug effectiveness and not the societal costs of the disease. In these settings, consider that the parameter vector $\boldsymbol{\theta}$ can be split into 2 components $\boldsymbol{\theta} = (\boldsymbol{\phi}, \boldsymbol{\psi})$ where the sample $\boldsymbol{X}$ is directly dependent of the parameters $\boldsymbol{\phi}$, while $\boldsymbol{\psi}$ are nuisance parameters.

In general, there is no guarantee that the shape of the prior distribution for the INB conditional on all the model parameters $\boldsymbol{\theta}$ will be the same as the distribution of the INB conditional on $\boldsymbol{\phi}$ with all the uncertainty due to the $\boldsymbol{\psi}$ marginalised out. To demonstrate the phenomenon, we introduce a simple two parameter model: $\phi \sim Be(1, 4)$ and $\psi \sim N(-0.5, 1)$ where $\mathrm{NB}_0^{\boldsymbol{\theta}} = 10\,000\psi - 4\,000$; $\mathrm{NB}_1^{\boldsymbol{\theta}} = 10\,000\phi - 6\,500$; and $\mathrm{INB}^{\boldsymbol{\theta}} = 10\,000(\phi - \psi) - 2\,500$.
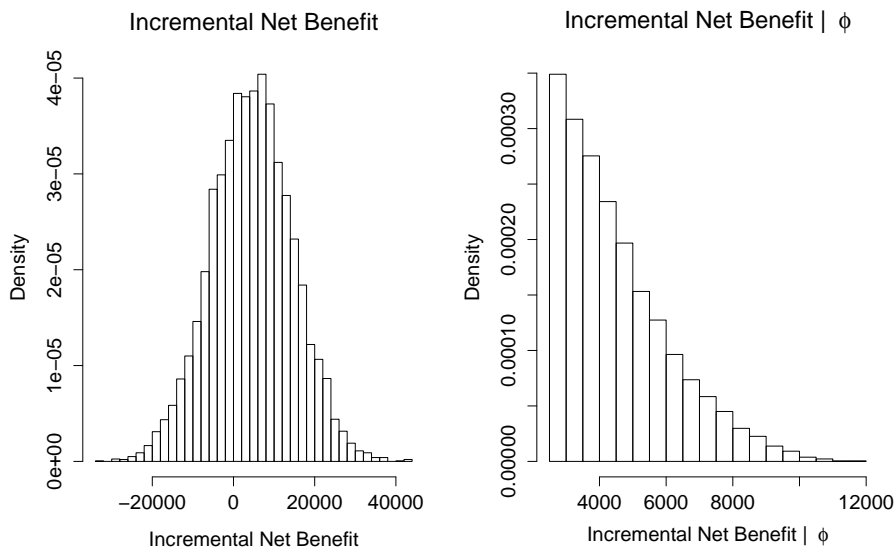

Figure 2: The distribution of the incremental net benefit conditional on both the model parameters $\boldsymbol{\theta}$ (LHS) and conditional on the parameter of interest ϕ (RHS).

Figure 2 (LHS) shows the prior for the INB which has a similar shape to a normal distribution — particularly in the tails. However, as the INB is linear in ϕ, the distribution of the INB conditional only on ϕ would be a shifted and scaled Beta distribution which has a very different shape, as seen in Figure 2 (RHS). Therefore, if the prior for the INB was used to approximate the distribution of the preposterior mean where $\boldsymbol{X}$ only gives information about ϕ, the mean and variance would be correct but the shape would be approximately normal which would lead to wildly inaccurate estimates for the EVSI.

To identify a more appropriate shape for the distribution of the preposterior mean, the uncertainty due to $\boldsymbol{\psi}$ should be marginalised out so the nuisance parameters do not impact the shape of the approximate distribution of the preposterior mean:

$$\int_{\boldsymbol{\Psi}} \text{INB}(\boldsymbol{\phi}, \boldsymbol{\psi}) p(\boldsymbol{\phi}, \boldsymbol{\psi}) d\boldsymbol{\psi} = \text{E}_{\boldsymbol{\psi}|\boldsymbol{\phi}}[\text{INB}(\boldsymbol{\phi}, \boldsymbol{\psi})]. \tag{2}$$

In general, this marginalisation is a computationally intensive procedure. However, this expectation should be available as part of a standard Value of Information (VoI) analysis (Tuffaha et al., 2016). As the EVSI is computationally intensive to calculate, it should be preceded in a full VoI analysis with an assessment of the value of resolving *all* the uncertainty in $\boldsymbol{\phi}$ – known as the Expected Value of *Partial* Perfect Information (EVPPI). Only if the value of resolving *all* the uncertainty in $\boldsymbol{\phi}$ is sufficiently large, i.e. a trial aimed at resolving the uncertainty in $\boldsymbol{\phi}$ would be significantly less expensive than the EVPPI, should the EVSI for a specific data collection exercise be considered. This is because, the EVSI is concerned with a specific data collection strategy which is typically time-consuming to design. Only once we know there is some value in learning the parameter should we be concerned with whether a specific trial should go ahead and what is the "optimal" design. More importantly, the EVPPI is based on the expectation in (2) and therefore, the samples from the conditional distribution of the INB should already have been calculated.

However, if these values are not available then Strong et al. (2014) and Heath et al. (2016) offer computationally efficient procedures for marginalising out uncertainty due to $\boldsymbol{\psi}$ based on non-parametric regression. These methods were developed to estimate the EVPPI and therefore are likely to be familiar to researchers who would be calculating the EVSI.

5 Examples

The approximation for the distribution of the preposterior mean is now used to estimate the EVSI in some standard scenarios. This section begins with three "toy" examples that exploit conjugacy to find both analytic results and computationally efficient algorithms to calculate both the true distribution for the preposterior mean and the true EVSI. These examples are used to explore some difficulties associated with estimating the EVSI using this methodology and demonstrate situations when it is suitable. Firstly, the Beta-Binomial example from §2.4 is extended to demonstrate the difficulties associated with using a continuous approximation for discrete samples. Secondly, Exponential-Gamma conjugacy is exploited to demonstrate this methodology where the data summary does not have the same distribution as the prior. Finally, Normal-Normal conjugacy is exploited to explore the variance estimation procedure §4.1 when the net benefit function is highly non-normal.

To conclude this section a decision tree model developed in Ades et al. (2004) is used to explore the use of non-parametric regression to marginalise out uncer-

tainty due to the nuisance parameters presented in §4.3. The new methodology presented here is also contrasted with another EVSI calculation method, developed in Strong et al. (2015), which requires explicit knowledge of summary statistics and is based on non-parametric regression.

5.1 Discrete Samples with Beta-Binomial Conjugacy

Revisiting the first example presented in §2.4, the parameter θ is modelled using a vague Beta prior $\theta \sim Beta(1,1)$ and the data have a binomial distribution $X \mid \theta \sim Bin(N, \theta)$. The two net benefit functions are then $\mathrm{NB}_0^\theta = 0$ and $\mathrm{NB}_1^\theta = k\theta - c$.

In this setting, the approximation of the distribution of the preposterior mean could be poor, as the data collection exercise is discrete, implying that the true distribution of the preposterior mean is also discrete, while the prior for NB_1^θ is continuous. This phenomenon can be seen most clearly when the binomial sample size $N = 1$ and there are 2 equally likely possible samples, $X = 0$ and $X = 1$. This implies that there are two equally likely possible preposterior means; $\mu_1^0 = \frac{k}{3} - c$ or $\mu_1^1 = \frac{2k}{3} - c$. Clearly, this distribution will never be well approximated by a shifted and rescaled beta distribution.

To investigate when such a continuous approximation is suitably accurate, we estimate the EVSI for different possible sample sizes N. Due to the conjugate structure, it is possible to calculate both the EVSI and the variance of the preposterior mean analytically. The true variance of the preposterior mean is then used to calculate a and b from §3.2. These are then used to approximate the EVSI by shifting and rescaling the prior for the net benefit:

$$\mathrm{EVSI} \approx \frac{1}{10\,000} \sum_{s=1}^{10\,000} \max\left\{0, a\,\mathrm{NB}_1^{\theta_s} + b\right\},$$

where θ_s is the s-th simulated value from the prior for θ (in this case, we use a *simulation* size of $10\,000$).

While the variance of the preposterior mean is available analytically, the estimator for the EVSI is still subject to variability due to the specific prior simulation of θ. Therefore, $10\,000$ different simulations of size $10\,000$ were taken from the prior for θ and used to calculate the EVSI with our method to find the sampling distribution of the EVSI estimator. This distribution should be centred on the true analytic value for the EVSI.

Figure 3 shows the sampling distribution of the EVSI estimator for different sample sizes for X, with a red line marking the sample specific EVSI for each sample size N. The top LHS shows that the EVSI estimator for $N = 1$ has a significant downward bias as the sampling distribution does not include the true EVSI value of 1667. This clearly indicates that the weighted prior distribution is not a suitable approximation for the distribution of the preposterior mean in this setting, as expected. However, as the sample size increases, the bias decreases and for a sample size of only 10 it is negligible, as the distribution of the EVSI is centred at the true value.
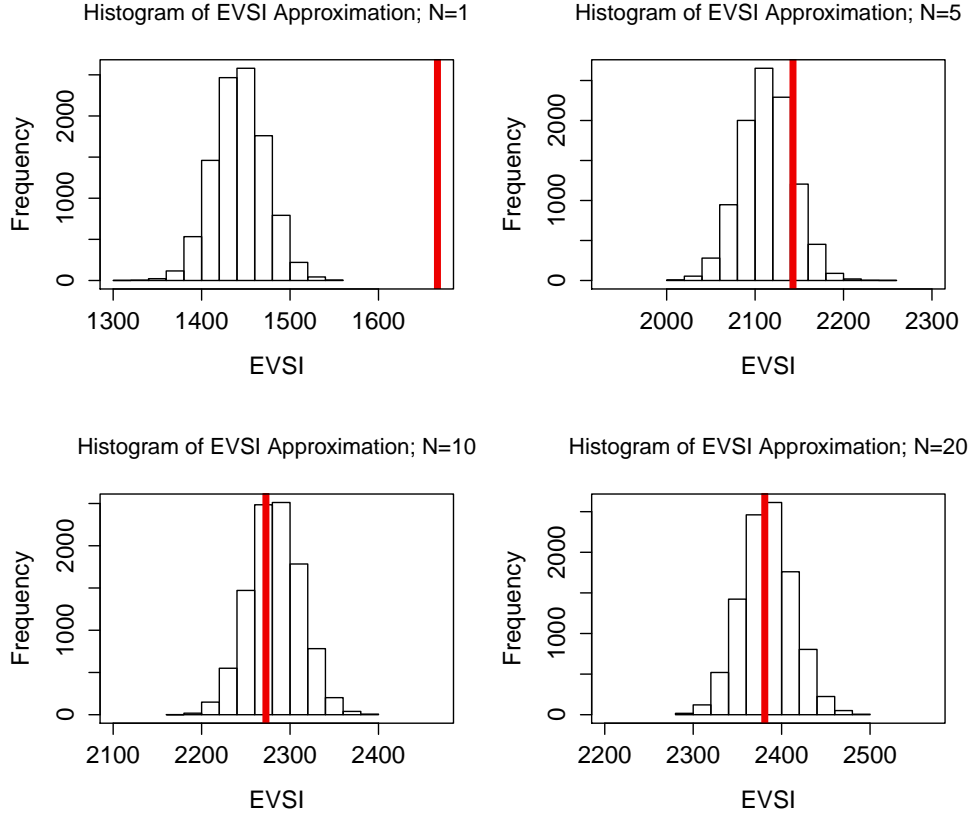
17

Figure 3: The distribution of the EVSI estimator over 10 000 different simulations from the prior of θ for 4 different sample sizes for $\boldsymbol{X}$ for the Beta-Binomial conjugate model. The red line represents the analytical value of the EVSI.

This analysis indicates that even if the distribution of the preposterior mean is discrete, it can be well approximated when the sample size for X is sufficiently large, > 10 in this example. In general, if the estimation of the EVSI is the primary interest, then a smaller sample size may be permitted, as we rarely require that the EVSI is estimated to a high degree of accuracy. This is because the EVSI is compared with trial costs which are rarely known with certainty. Additionally, the EVSI is based on incorrect model assumptions meaning that even if the EVSI is highly accurate given the model, it will be approximate when applied in practice.

5.2 Non-linear mean function with Exponential-Gamma conjugacy

We now revisit the second example in §2.4 where a Gamma prior is assumed for the parameter of interest $\theta \sim Gamma(\alpha, \beta)$, the data collection exercise is assumed to be N independent observations from an exponential distribution $X_j \sim Exp(\theta)$, $j = 1, \dots, N$ and the two net benefit functions are:

$$\text{NB}_0^\theta = c_0 \quad \text{and} \quad \text{NB}_1^\theta = k\theta - c_1.$$

Throughout §5.2, we present the results for $\alpha = 5$, $\beta = 1$, $k = 200$ and $c_0 = 900$ and $c_1 = 100$ as in Figure 1.

Using Gamma-Exponential conjugacy, it is trivial to show that the preposterior mean is equal to

$$\mu_1^{\boldsymbol{X}} = \text{E}_{\theta|\boldsymbol{X}}\left[\text{NB}_1^\theta\right] = k\frac{\alpha + N}{\beta + \sum_{i=1}^{N} X_i} - c_1,$$

which means that both the variance of the preposterior mean and the EVSI can be found analytically, as for the previous example. Therefore, any difficulties in estimating the EVSI are because the weighted prior distribution is not a suitable approximation to the distribution of the preposterior mean. As this is a conjugate model, this misspecification is because the data summary does not have the same distribution as the prior.

Figure 4 shows the sampling distribution of the EVSI values, over different prior samples, for different values of N. Clearly, moment matching with the prior gives a biased EVSI estimate for small samples. However, this bias is at most 4% of the total EVSI value, meaning that it can still be used, as the estimate is sufficiently accurate for decision making. It is recommended, however, to remember that the EVSI estimate is slightly biased for small sample sizes and therefore care should be taken interpreting the EVSI for these small samples. As N increases, this bias becomes negligible as the distribution of the preposterior mean tends exactly to the prior as the sample size increases, see Figure 1.

5.3 Estimating the variance of the preposterior mean

In this example we investigate the estimation procedure for the variance of the preposterior mean given in §4.1. This estimation procedure is highly effective when the prior for the $\text{INB}^{\boldsymbol{\theta}}$ is roughly normal. Therefore, to test this procedure we use a model where the prior for the $\text{INB}^{\boldsymbol{\theta}}$ is highly non-normal.

Counterintuitively, it is possible to exploit normal-normal conjugacy to investigate this estimation procedure by setting $\text{INB}^\theta = \theta^2 - 5$ where θ is normal *a priori* with mean 0 and precision 0.2: $\theta \sim N(0, 5)$. The data collection is then assumed to be 10 independent observations $X_j \sim N(\theta, 1)$ for $j = 1, \dots, 10$. Conjugacy can now be used to calculate a value for the posterior for θ efficiently, while inducing a highly non-normal prior for the INB^θ.

Using conjugacy, it is possible to estimate both the EVSI and the variance of the preposterior mean cheaply using Monte Carlo methods (Ades et al., 2004).
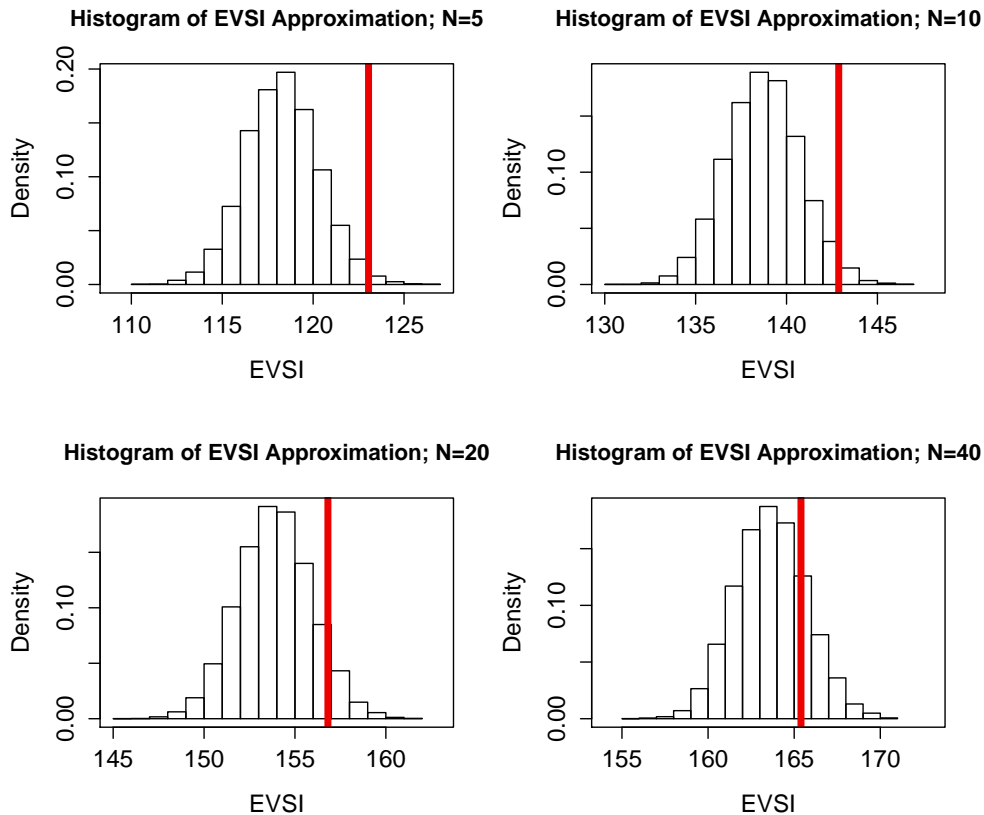
Figure 4: The distribution of the EVSI estimator over 10 000 different simulations from the prior of θ for 4 different samples sizes for $\boldsymbol{X}$ using the Exponential-Gamma conjugate model. The red line represents the analytic value of the EVSI,

Therefore, using 10 000 samples from the prior for the INB$^\theta$, the EVSI is estimated as 2.00 and the variance of the preposterior mean is 35.20.

The estimation method for the variance of the preposterior mean requires Q quadrature points spaced throughout the domain of θ. Practically these are taken as the Q quantiles for θ, i.e. the $S\frac{q}{Q+1}$–th θ values in an ordered sample, with $q = 1, \ldots, Q$.

Figure 5 shows the average estimate, over 500 simulations, for the preposterior variance for increasing values of Q up to $Q = 100$ — this means that 1 000 simulations were taken from 100 different posterior distributions 500 times. The red line in Figure 5 shows the estimated value of the variance of the preposterior distribution calculated using the method from Ades et al. (2004). The dashed lines indicate plus or minus one standard deviation from the mean estimate of the variance of the preposterior distribution for the different sample sizes.

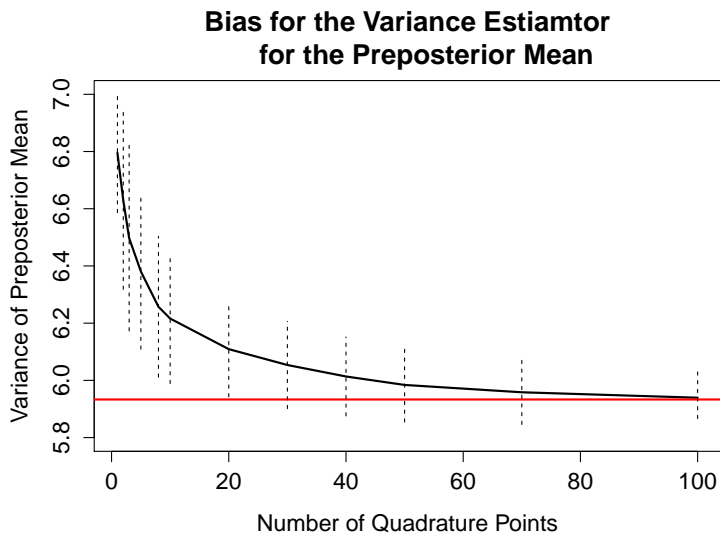**Bias for the Variance Estiamtor for the Preposterior Mean**

Figure 5: The estimate of the variance of the preposterior mean for increasing numbers of quadrature points. The red line gives the variance of the preposterior mean calculated using all the samples in the prior for θ. The dashed lines are the standard errors for the estimates of the variance of the preposterior mean.

In general, our estimation method for the variance of the preposterior mean produces biased estimates for small numbers of quadrature points. However, once the number of quadrature points exceeds 30, the true variance is within one standard deviation of the average estimate of the variance of the preposterior mean. The standard deviation of this estimate also decreases as the number of quadrature points increases.

Table 1: The EVSI estimate for different numbers of posterior samples using the moment matching method.

Number of simulations	1	2	3	5	8	10	∞
Estimate of EVSI	2.30	2.24	2.20	2.16	2.12	2.10	2.00
Percentage Bias	0.15	0.12	0.10	0.08	0.06	0.05	0.00
Number of simulations	20	30	40	50	75	100	∞
Estimate of EVSI	2.06	2.05	2.03	2.02	2.01	2.01	2.00
Percentage Bias	0.03	0.02	0.02	0.01	0.01	0.01	0.00

Table 1 shows the EVSI estimate and its bias when we use the estimates for the variance of the preposterior mean given in Figure 5 and our moment matching methodology. Notice that while all the EVSI values are over-estimated the percentage bias drops below 0.02 for more than 30 quadrature points. There-

21

fore, it seems that a relatively small number of quadrature points are reasonable for estimating the variance of the preposterior mean, even in significantly non-normal settings.

5.4   Ades et al. Decision Tree Model

Ades et al. (2004) develop a simple decision tree model to demonstrate their methods for calculating the EVSI using single step Monte Carlo. This involves assuming that the stochastic model parameters, which represent odds ratios, probabilities and utility measures for the different health states, are independent and have distributions in conjugate families. In addition to this, the net benefit function must be of a certain form in order to calculate the posterior mean without sampling from the posterior distribution of the parameters. If this is not the case, then Taylor series expansions are used to avoid sampling from the posterior. This use of Taylor series expansions means that the results below do not depend on conjugacy.

To compare our methodology with Ades et al. (2004) and Strong et al. (2015), a recent EVSI approximation method that uses sufficient statistics and non-parametric regression, the variance estimation procedure with 30 posterior simulations of 10 000 was performed 1 000 times to give a distribution over the posterior variance. Posterior updating was performed using OpenBUGS through R. 10 000 samples from the prior for the incremental net benefit were used and GAM regression (Hastie and Tibshirani, 1990) using the gam function from the mgcv package (Wood et al., 2016) in R was used to integrate out uncertainty due to ψ. For full details on the model structure see Ades et al. (2004) and appendices therein.

Figure 6 shows the distribution for the EVSI estimate over the different estimates for the variance of the preposterior distribution for 4 different parameter combinations from the Ades et al. (2004) model. The EVSI for 3 of the 4 parameters, θ_1, θ_2 and θ_3, are calculated in Strong et al. (2015) using nested Monte Carlo simulation and their method based on 10^{10} and 10^6 simulations respectively. The values are represented by red (nested Monte Carlo) and blue lines (non-parametric regression) in Figure 6. The 4$^\text{th}$ parameter subset contains two parameters θ_3^T and θ_3^C and was not considered in Ades et al. (2004) or Strong et al. (2015) and therefore the blue line in the lower RHS panel represents the Strong et al. (2015) estimate using 10^6 observations.

Our estimates for the EVSI are in line with both these alternative estimation methods. This suggests that using non-parametric regression to marginalise out uncertainty due to the additional model parameters gives sufficiently accurate estimates for the EVSI. It also suggests that using quadrature over a two dimensional parameter vector does not have an effect on the estimation properties. As this example is fairly representative of many health economic models, we believe that our method can be successfully applied in practise. This allows practitioners to calculate the EVSI is most practical settings without resorting to full nested Monte Carlo simulation which is prohibitively expensive.
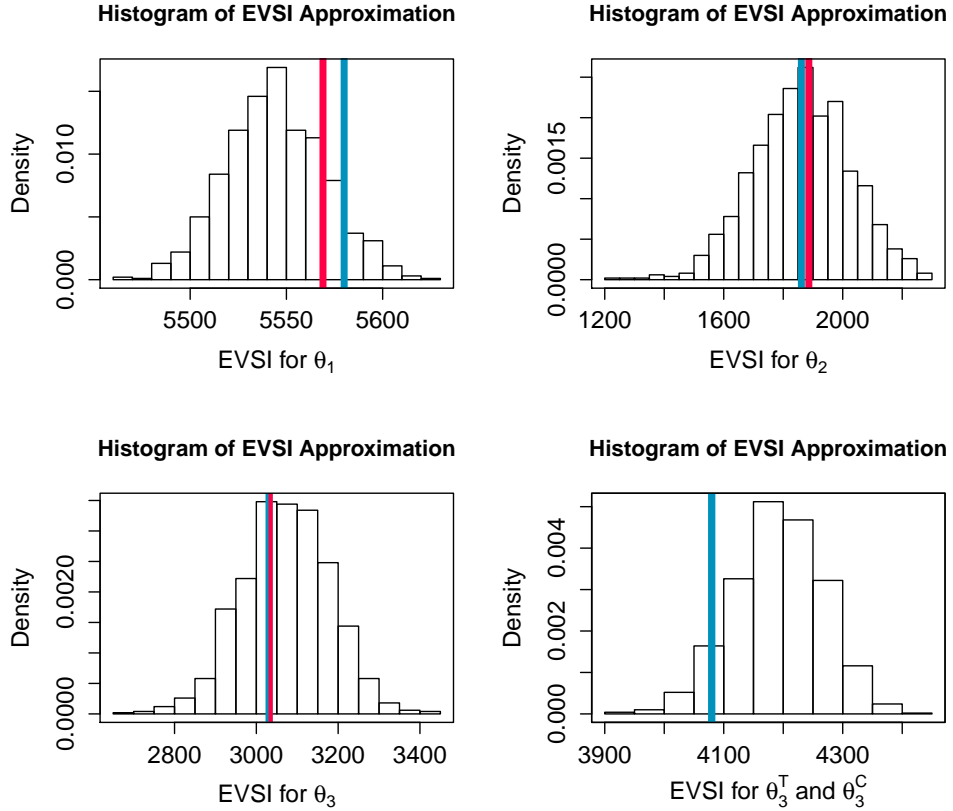
Figure 6: The sampling distribution of the EVSI conditional on the distribution over the different estimates for the variance of the preposterior mean for θ_1, θ_2 and θ_3 for the (Ades et al., 2004) example. The red line represents the EVSI calculated using Monte Carlo methods and 10^{10} simulations. The blue line represents the EVSI estimate obtained using the Strong et al. (2015) method with 10^6 simulations. Both values are taken directly from Strong et al. (2015) except for the bottom RHS graphic.

6   Acknowledgements

References

Ades, A., Lu, G., and Claxton, K. (2004). "Expected Value of Sample Information Calculations in Medical Decision Modeling." *Medical Decision Making*, 24: 207–227.

Andronis, L., Barton, P., and Bryan, S. (2009). "Sensitivity analysis in economic evaluation: an audit of NICE current practice and a review of its use and value in decision-making."

Arendt, P., Apley, D., and Chen, W. (2016). "A preposterior analysis to predict identifiability in the experimental calibration of computer models." *IIE Transactions*, 48(1): 75–88.

Baio, G. (2012). *Bayesian methods in Health Economics*. CRC Press, Boca Raton, FL.

Baio, G. and Dawid, P. (2011). "Probabilistic sensitivity analysis in health economics." *Statistical methods in medical research*.

Ben-Zvi, M., Berkowitz, B., and Kesler, S. (1988). "Pre-posterior analysis as a tool for data evaluation: application to aquifer contamination." *Water Resources Management*, 2(1): 11–20.

Brennan, A. and Kharroubi, S. (2005). "Efficient Computation of Partial Expected Value of Sample Information Using Bayesian Approximation." Research Report 560/05, Department of Probability and Statistics, University of Sheffield, UK.

Brennan, A., Kharroubi, S., O'Hagan, A., and Chilcott, J. (2007). "Calculating Partial Expected Value of Perfect Information via Monte Carlo Sampling Algorithms." *Medical Decision Making*, 27: 448–470.

Briggs, A., Sculpher, M., and Claxton, K. (2006). *Decision modelling for health economic evaluation*. Oxford University Press, Oxford, UK.

Çetinkaya, E. and Thiele, A. (2016). "A moment matching approach to lognormal portfolio optimization." *Computational Management Science*, 1–20.

Chaloner, K. and Larntz, K. (1989). "Optimal Bayesian design applied to logistic regression experiments." *Journal of Statistical Planning and Inference*, 21(2): 191–208.

Claxton, K. (1999). "Bayesian approaches to the value of information: implications for the regulation of new pharmaceutical." *Health Economics*, 8: 269–274.

Claxton, K., Neumann, P., Araki, S., and Weinstein, M. (2001). "Bayesian Value-Of-Information Analysis." *International Journal of Technology Assessment in Health Care*, 17: 38–55.

Diaconis, P. and Ylvisaker, D. (1979). "Conjugate Priors for Exponential Families." *Ann. Statist.*, 7(2): 269–281.
URL http://dx.doi.org/10.1214/aos/1176344611

Erkanli, A. and Soyer, R. (2000). "Simulation-based designs for accelerated life tests." *Journal of Statistical Planning and Inference*, 90(2): 335–348.

Felli, J. and Hazen, G. (1998). "Sensitivity analysis and the expected value of perfect information." *Medical Decision Making*, 18: 95–109.

— (1999). "A Bayesian approach to sensitivity analysis." *Health Economics*, 8: 263–268.

Feng, L., Korvink, J., and Benner, P. (2015). "A fully adaptive scheme for model order reduction based on moment matching." *IEEE Transactions on Components, Packaging and Manufacturing Technology*, 5(12): 1872–1884.

Fenwick, E., Palmer, S., Claxton, K., Sculpher, M., Abrams, K., and Sutton, A. (2006). "An Iterative Bayesian Approach to Health Technology Assessment: Application to a Policy of Preoperative Optimization for Patients Undergoing Major Elective Surgery." *Medical Decision Making*, 26: 480–496.

Hastie, T. and Tibshirani, R. (1990). *Generalized additive models*, volume 43. CRC Press.

Heath, A., Manolopoulou, I., and Baio, G. (2016). "Estimating the expected value of partial perfect information in health economic evaluations using integrated nested laplace approximation." *Statistics in medicine*.

Howard, R. (1966). "Information Value Theory." In *IEEE Transactions on System Science and Cybernetics*. (1) 22-26: SCC-2.

Huang, Y. and Wu, H. (2008). "Bayesian experimental design for long-term longitudinal HIV dynamic studies." *Journal of statistical planning and inference*, 138(1): 105–113.

Jiang, Z., Apley, D., and Chen, W. (2015). "Surrogate Preposterior Analyses for Predicting and Enhancing Identifiability in Model Calibration." *International Journal for Uncertainty Quantification*, 5(4).

Lindley, D. (2006). *Understanding uncertainty*. John Wiley & Sons.

Pratt, J., Raiffa, H., and Schlaifer, R. (1995). *Introduction to statistical decision theory*. MIT press.

Raiffa, H. and Schlaifer, H. (1961). *Applied Statistical Decision Theory*. Harvard University Press, Boston, MA.

Schlaifer, R. (1959). "Probability and statistics for business decisions."

Stinnett, A. and Mullahy, J. (1998). "Net health benefits a new framework for the analysis of uncertainty in cost-effectiveness analysis." *Medical Decision Making*, 18(2): S68–S80.

Strong, M., Oakley, J., and Brennan, A. (2014). "Estimating Multiparameter Partial Expected Value of Perfect Information from a Probabilistic Sensitivity Analysis Sample A Nonparametric Regression Approach." *Medical Decision Making*, 34(3): 311–326.

Strong, M., Oakley, J., Brennan, A., and Breeze, P. (2015). "Estimating the Expected Value of Sample Information Using the Probabilistic Sensitivity Analysis Sample A Fast Nonparametric Regression-Based Method." *Medical Decision Making*, 0272989X15575286.

Tuffaha, H., Gordon, L., and Scuffham, P. (2016). "Value of Information Analysis Informing Adoption and Research Decisions in a Portfolio of Health Care Interventions." *MDM Policy & Practice*, 1(1): 2381468316642238.

Weaver, B., Williams, B., Anderson-Cook, C., Higdon, D., et al. (2016). "Computational enhancements to Bayesian design of experiments using Gaussian processes." *Bayesian Analysis*, 11(1): 191–213.

Welton, N., Madan, J., Caldwell, D., Peters, T., and Ades, A. (2014). "Expected value of sample information for multi-arm cluster randomized trials with binary outcomes." *Medical Decision Making*, 34(3): 352–365.

Welton, N. and Thom, H. (2015). "Value of Information Weve Got Speed, What More Do We Need?" *Medical Decision Making*, 35(5): 564–566.

Wood, S., Pya, N., Kneib, T., Hornik, K., Lonergan, M., Nilsson, H., Scheipl, F., and Ripley, B. (2016). *mgcv: Mixed GAM Computation Vehicle with GCV/AIC/REML smoothness estimation and GAMMs by REML/PQL*.

A   Normal normal conjugacy example.

This normal-normal example is presented for two reasons. Firstly, it demonstrates that our method is exact in the normal-normal conjugate setting and therefore can be used when the prior distribution and $g(\boldsymbol{X})$ are both sufficiently normal. It is also presented to help clarify some of the thinking presented throughout the paper.

To begin, it is assumed that the variances for the prior of θ and the data collection exercise are known and denoted σ_θ^2 and $\sigma_{\boldsymbol{X}}^2$ respectively. We then assume that the prior for θ is

$$\theta \sim N(\theta_0, \sigma_\theta^2)$$

and the data collection exercise is N independent samples from

$$X_i \sim N(\theta, \sigma_{\boldsymbol{X}}^2).$$

This implies that the sample mean of $\boldsymbol{X}$ has a distribution, conditional θ,

$$\bar{\boldsymbol{X}} \sim N\left(\theta, \frac{\sigma_{\boldsymbol{X}}^2}{N}\right).$$

In a normal-normal setting with known variances, the prior-predictive distribution can be easily calculated with the prior-predictive distribution for $\bar{\boldsymbol{X}}$

$$\bar{\boldsymbol{X}} \sim N\left(\theta_0, \sigma_\theta^2 + \frac{\sigma_{\boldsymbol{X}}^2}{N}\right).$$

We assume that the net benefits in this example is given by

$$\mathrm{NB}_0^\theta = 0 \quad \text{and} \quad \mathrm{NB}_1^\theta = k\theta - c.$$

The preposterior mean for the INB is then

$$\mathrm{E}_{\theta|\boldsymbol{X}}(\mathrm{INB}^\theta) = k\left(\frac{\sigma_{\boldsymbol{X}}^2}{\sigma_{\boldsymbol{X}}^2 + N\sigma_\theta^2}\theta_0 + \frac{\sigma_\theta^2}{\frac{\sigma_{\boldsymbol{X}}^2}{N} + \sigma_\theta^2}\bar{\boldsymbol{X}}\right) - c,$$

which is a linear function of a normal distribution. Therefore, the distribution of the preposterior mean is normal with mean and variance equal to the mean and variance of the preposterior mean;

$$\mathrm{E}_{\theta|\boldsymbol{X}}(\mathrm{INB}^\theta) \sim N\left(k\theta_0 - c, k^2 \frac{\sigma_\theta^4}{\frac{\sigma_{\boldsymbol{X}}^2}{N} + \sigma_\theta^2}\right).$$

Firstly, note that our approximation to the distribution of the preposterior mean is $a\,\mathrm{INB}^\theta + b$, which is a linear combination of a normal distribution. Consequently, our approximation is also a normal distribution with the same mean and variance as the preposterior mean. Therefore, our approximation is exactly equal to the true distribution of the preposterior mean in this setting.

Secondly, note that as $N \to \infty$, it is clear to see that the variance of the preposterior mean INB^θ tends to $k^2\sigma_\theta^2$, meaning that the distribution of the preposterior mean tends to the prior for the incremental net benefit. Another way to think about this property is that for large sample sizes is that the sample mean $\bar{\boldsymbol{X}}$ collapses to the underlying mean of X_i, which in this setting is θ where θ itself is subject to uncertainty. Clearly, therefore, for an infinite sample

size, i.e. when the sample mean is exactly equal to θ, the distribution of the preposterior mean is exactly equal to the prior, as the preposterior mean is exactly equal to $\bar{X} \sim N(\theta, \sigma_\theta^2)$.

In addition to this, observe that, as N increases, the size of the denominator decreases, meaning that the distribution of the preposterior mean gets *more* variable as the sample size increases. This confirms our intuition that as more information is contained in the data, i.e. the sample size increases, the posterior mean can be "pulled" further from the prior mean and therefore more weight in the distribution of the preposterior mean is given to values further from the prior and the variance of the preposterior mean increases.

Finally, it is trivial to see how much the distribution of the preposterior mean is dependent on our prior beliefs. Not only is it centred on the prior mean but the prior variance has a larger impact on the variance of the preposterior mean than the sample variance. The distributional assumptions for the prior also impact the distribution of the preposterior mean. This underlines why this moment matching methodology can be used so successfully as the hypothetical data collection exercise gives us *no* additional information than the information contained in the prior, remembering that once the data collection exercise has been realised then this *will* give us additional information.