KATHOLIEKE UNIVERSITEIT
LEUVEN

Faculty of Business and Economics

A comparison of different Bayesian design criteria to compute efficient conjoint choice experiments

J. Yu, P. Goos and M. Vandebroek

DEPARTMENT OF DECISION SCIENCES AND INFORMATION MANAGEMENT (KBI)

KBI 0817

A Comparison of Different Bayesian Design Criteria to Compute Efficient Conjoint Choice Experiments

Jie Yu

Faculty of Business and Economics, Katholieke Universiteit Leuven, Naamsestraat 69, B-3000 Leuven, Belgium.
Email: Jie.Yu@econ.kuleuven.be
Tel: +32 (0)16 32 69 62
Fax: +32 (0)16 32 66 24

Peter Goos

Faculty of Applied Economics, Universiteit Antwerpen, Prinsstraat 13, B-2000 Antwerpen, Belgium.
Email: Peter.Goos@ua.ac.be
Tel: +32 (0)3 220 40 59
Fax: +32 (0)3 220 48 17

Martina Vandebroek

Faculty of Business and Economics & Leuven Statistics Research Centre, Katholieke Universiteit Leuven, Naamsestraat 69, B-3000 Leuven, Belgium.
Email: Martina.Vandebroek@econ.kuleuven.be
Tel: +32 (0)16 32 69 75
Fax: +32 (0)16 32 66 24

A Comparison of Different Bayesian Design Criteria to Compute Efficient Conjoint Choice Experiments

Abstract

Bayesian design theory applied to nonlinear models is a promising route to cope with the problem of design dependence on the unknown parameters. The traditional Bayesian design criterion which is often used in the literature is derived from the second derivatives of the loglikelihood function. However, other design criteria are possible. Examples are design criteria based on the second derivative of the log posterior density, the expected posterior covariance matrix, or on the amount of information provided by the experiment. Not much is known in general about how well these criteria perform in constructing efficient designs and which criterion yields robust designs that are efficient for various parameter values. In this study, we apply these Bayesian design criteria to conjoint choice experimental designs and investigate how robust the resulting Bayesian optimal designs are with respect to other design criteria for which they were not optimized. We also examine the sensitivity of each design criterion to the prior distribution. Finally, we try to find out which design criterion is most appealing in a non-Bayesian framework where it is accepted that prior information must be used for design but should not be used in the analysis, and which one is most appealing in a Bayesian framework when the prior distribution is taken into account both for design and for analysis.

Keywords: Bayesian design criterion, posterior density, expected posterior covariance matrix, conjoint choice design, Laplace approximation, Fisher information

1 Introduction

Consumer behavioral models provide a quantitative way to assess the relative importance of one attribute against another. A very popular tool for modeling consumer behavior is discrete choice analysis. Data are collected through choice experiments which aim at evaluating consumers' preferences on a certain product or service. The quality of the outcome of such an experiment depends a lot on its design. The choice of the alternatives or profiles to be compared and the way of grouping different alternatives into choice sets is a central issue in the quality of the data collection. It has been shown that efficient choice designs indeed play an important role in improving the statistical inference about the quantities of interest (Huber and Zwerina 1996; Sándor and Wedel 2001, 2002). Optimal designs are commonly constructed by optimizing a criterion with respect to the design variables of interest.

A serious difficulty in constructing an efficient choice design for the probabilistic choice model is that it is nonlinear in the parameters and therefore requires knowledge of the values of the parameters (Atkinson and Donev 1992; Atkinson and Haines 1996; Dror and Steinberg 2006; Sándor and Wedel 2001). This implies that researchers need to assume values for the parameters before constructing the designs. In general, three approaches have been implemented to tackle this problem. The simplest one is to assume that the respondents have no preference for one alternative over another. This leads to zero prior parameter values for constructing designs (Anderson and Wiley 1992; Lazari and Anderson 1994). Huber and Zwerina (1996) introduced the nonzero prior strategy based on the belief that information is usually available prior to conducting the experiments. In their approach, designs were constructed based on the experimenter's best prior point estimate of the unknown parameters. This approach leads to locally optimal designs that are more efficient than those obtained by a zero prior if the assumed values are reasonably close to the unknown true values.

In recent years, the semi-Bayesian approach introduced in the marketing literature by Sándor and Wedel (2001) has been widely used for choice experiments (Bliemer et al. 2008; Kessels et al. 2006; Yu et al. 2008; Vermeulen et al. 2008). This approach takes into account all possible values of the parameters when constructing designs. A prior distribution is assumed for the parameters in the design stage, which is then incorporated into an appropriate design criterion. Sándor and Wedel (2001) showed the usefulness of this approach over the locally optimal approach in constructing experimental designs and concluded that taking into account the prior uncertainty in the design stage leads to designs that are robust against a poor initial guess. In recent years, constructing designs for nonlinear models in a Bayesian fashion has become the state of the art to cope with the problem of design dependence on the unknown parameters of the fitted models.

Both the local approach and the semi-Bayesian approach described above use the asymptotic covariance matrix of the maximum likelihood estimates (the inverse of the Fisher information matrix) as the design criterion to construct choice experiments. Therefore, prior information on the parameters is only taken into account in the design stage. This is different in a true Bayesian approach which use the design criterion developed in a Bayesian estimation context and it takes into account prior knowledge both for estimation and design procedures.

A true Bayesian approach looks at the covariance matrix of the posterior distribution in constructing choice designs. The corresponding design criteria can easily be formulated, but exact solutions are often intractable. This is because, in general, for the non-linear model, the posterior distribution of the parameters of interest cannot be found in closed form. To obtain practical solutions in the construction of choice designs, one often resorts to asymptotic Bayesian design criteria. Several asymptotic approximations to the posterior covariance matrix are given in Berger (1985). We will limit attention to two approximations. These include the second derivative of the loglikelihood function and the second derivative of the logposterior density.

However, in the literature on choice experiments, the validity of these asymptotic approximations to the posterior covariance matrix and the performance of the optimal designs constructed with the Bayesian criteria based on the expected posterior covariance matrix and those based on its approximations have not been compared. Little guidance is available on which type of Bayesian design criterion is best to use in practice when only a small number of observations are allowed in the experiment.

The approximation based on the second derivative of the loglikelihood function or the Fisher information matrix has been widely used as a design criterion by many authors for constructing choice-based experimental designs because of its computational simplicity (Huber and Zwerina 1996; Zwerina et al. 1996; Sándor and Wedel 2001, 2005; Kessels et al. 2006). However, in situations where the elements of the Fisher information matrix are small due to the sample size restriction, the asymptotic approximation based on the Fisher information matrix can be a poor approximation to the true posterior covariance matrix. In this situation, the optimal designs based on this asymptotic criterion might be very inefficient compared to the designs constructed with the true one.

Tsutakawa (1972) used a better approximation to the expected posterior variance based on the second derivative of the logposterior density for a design criterion in the computation of a Bayesian design for a one-parameter logistic regression model with known slope coefficient and unknown median lethal dose, LD50. A normal prior was used in his paper. Therefore, the approximation was simplified as the Fisher information plus the precision of the prior. Tsutakawa (1980) extended this to designs for the estimation of other percentiles than the median. The accuracy of the asymptotic approximation to the expected posterior variance in Tsutakawa (1972) has been studied by Sun et al. (1996). They showed a remarkable closeness of the asymptotic approximations to the exact ones in selecting dose levels.

In this paper, we examine the performance of design criteria based on asymptotic theory for constructing choice designs when multiple factors are allowed. In addition, we study design criteria such as the expected posterior covariance matrix and the amount of information provided by an experiment as given by the Shannon information. These approaches do not rely on the asymptotic approximations. Designs constructed with the former criterion aim at minimizing the expected posterior covariance matrix of the Bayesian estimator, while the latter criterion is based on the information theoretic approach and aims at maximizing the expected gain in Shannon information. Chaloner and Verdinelli (1995) have provided a general view of Bayesian experimental design criteria. In particular, they give an overview of a number of loss functions and the alphabetic Bayesian criteria that correspond to these loss functions. Their focus was on presenting various Bayesian design criteria based on the asymptotic theory rather than on evaluating how much is sacrificed when using asymptotic Bayesian criteria instead of the corresponding exact criteria in the construction of efficient designs.

Use of the exact expected posterior variance to select designs has been considered in the field of clinical trials. Typical examples are Han and Chaloner (2004), who compare eight candidate designs numerically by computing Monte Carlo estimates of the posterior variances, Stroud, Müller and Rosner (2001), who assume the same prior distribution for design and analysis, and use the expected posterior variances of some population pharmacokinetic quantities of interest as the design criterion, and Sun et al. (1996) and Sun and Tsutakawa (1997), who considered Bayesian design problems in choosing a set of dose levels. Note that all above authors considered the expected posterior variance instead of the covariance matrix of the parameters as a design criterion. This reduces the design criterion to only one parameter dimension. In our study, we use the posterior covariance matrix of the parameters in the design criterion.

The exact expected posterior covariance matrix has been rarely used up to now for planning experiments in the area of conjoint choice study due to computational problems. these prohibited to check how much worse the optimal designs constructed with the asymptotic criteria are compared to those based on the exact criterion, to investigate the robustness of different Bayesian criteria to the prior specification, and to examine how close the asymptotic approximation to the posterior covariance matrix is. In this paper, we provide answers to those questions. Our focus is on choice situations where only small sample sizes are feasible. Note that in situations where large samples are allowed, there is little meaning to check the performance of the asymptotic design criteria since they will all converge to the exact one.

So far, we have focused on the asymptotic approximation of the expected posterior covariance matrix. Another criterion that we consider is based on the information theoretic approach which is related to the concept of the Shannon information (Shannon 1948). Shannon introduced the notion of entropy to measure the uncertainty associated with a random variable. Based on this concept, Lindley (1956) introduced a Bayesian information approach to experimental designs which maximizes the gain in knowledge about the parameters. Knowledge is then measured by the amount of information provided by the experiment. The design which maximizes the expected Kullback-Leibler distance between the posterior and the prior distributions or, equivalently, maximizes the information gain in moving from the prior distribution to the posterior distribution was preferred. When the prior distribution does not depend on the design, this criterion is equivalent to maximizing the expected Shannon information of the posterior distribution. Sebastiani and Wynn (1997) reviewed the information theoretic approach to Bayesian experimental designs and examined computational issues related to constructing designs using the expected Shannon information of the posterior distribution. An application of the approximation to the Shannon information to experimental designs for non-linear models is given in Sebastiani and Settimi (1997, 1998). Merlé and Mentré constructed designs with one or two measurements for a pharmacokinetic and a pharmacodynamic model using the Shannon information and the expected posterior covariance matrix. They showed that these two criteria generally lead to the same designs except for the E_{max} model and a multiplicative measurement error.

In this paper, six Bayesian design criteria are introduced and used to optimize experiments for estimating the parameters of the conditional logit model. These designs are compared to investigate how robust they are with respect to other design criteria for which they are not optimized and to examine the sensitivity of each Bayesian design with respect to the prior distribution. We study the closeness of the asymptotic approximations to the posterior covariance matrix and explore how good the asymptotic Bayesian design criteria are for constructing efficient choice designs compared to the criteria which do not rely on the asymptotic theory. We also find out which design criterion is most appealing in a non-Bayesian framework where it is accepted that prior information must be used for design but not for the analysis, and which one is most appealing in a Bayesian framework when the prior distribution is taken into account for design and for analysis.

The results of this study are useful in the context of the sequential design construction, in which, for a given respondent, the parameters are estimated each time a choice set is rated. The design of the next choice set then depends on the parameter estimates based on all previous answers. Obviously, the number of observations used to compute the design for each individual respondent is small.

Another example where small samples are likely is in prototype experiments. Assume that a manufacturer would like to innovate the design of shaving machines, and that prototype shaving machines are produced. Each respondent is given one set of shaving machines to try for a certain period. The respondent is then asked to give his preference and to indicate the smoothness of the skin after using each type of shaving machines. This type of experiment in which prototypes are used is also common in other fields.

In the next section, we sketch the conditional logit model and introduce different Bayesian design criteria. In Section 3, we investigate the robustness of each Bayesian design with respect to design criteria for which they were not optimized. In Section 4, we describe the details of the simulation study and discuss the results, and Section 5 contains a summary of the main findings.

2 Conditional Logit Model and Design Efficiency Criteria

2.1 Conditional Logit model

The conditional logit probability that alternative k is chosen from choice set s is given by

$$p_{ks}(\boldsymbol{\beta}) = \frac{\exp(\mathbf{x'}_{ks}\boldsymbol{\beta})}{\sum_{i=1}^{K} \exp(\mathbf{x'}_{is}\boldsymbol{\beta})}, k = 1, ..., K, \tag{1}$$

with K the number of profiles in each choice set, $\mathbf{x}_{ks}$ a p-dimensional vector characterizing the attributes of profile k in choice set s, and $\boldsymbol{\beta}$ a p-dimensional coefficient vector containing the effects of the different attribute levels on the utility. For reasons of notational simplicity, we will denote $p_{ks}(\boldsymbol{\beta})$ by p_{ks}.

Suppose that there are S choice sets, then the likelihood function is given by

$$L(\mathbf{y}|\boldsymbol{\beta}) = \prod_{s=1}^{S} \prod_{k=1}^{K} p_{ks}^{y_{ks}}, \tag{2}$$

where y_{ks} denotes the number of times that respondents choose alternative k in choice set s. The maximum likelihood estimator $\hat{\boldsymbol{\beta}}$ for the parameter vector $\boldsymbol{\beta}$ is the vector of values that maximizes the likelihood function. As any prior information is ignored in the analysis, this approach is a non-Bayesian estimation approach.

In the Bayesian framework, an estimate for the parameter vector $\boldsymbol{\beta}$ is obtained from the posterior density of $\boldsymbol{\beta}$ given the data $\mathbf{y}$. Suppose that $\pi_0^I(\boldsymbol{\beta})$ is the prior distribution of the parameters. The posterior density is then given by

$$q(\boldsymbol{\beta}|\mathbf{y}) = \frac{L(\mathbf{y}|\boldsymbol{\beta})\ \pi_0^I(\boldsymbol{\beta})}{\int L(\mathbf{y}|\boldsymbol{\beta})\ \pi_0^I(\boldsymbol{\beta})\ d\boldsymbol{\beta}} = \frac{L(\mathbf{y}|\boldsymbol{\beta})\ \pi_0^I(\boldsymbol{\beta})}{p_Y(\mathbf{y})}, \tag{3}$$

where $p_Y(\mathbf{y})$ is the marginal distribution of Y. The posterior mode, which is the parameter vector that maximizes the posterior density, is used as an estimation for $\boldsymbol{\beta}$.

2.2 Design Efficiency Criteria

In the following section, we introduce six design criteria. The first four are based on an asymptotic approximation to the expected posterior covariance matrix. The last two criteria (which do not rely on asymptotic theory) are based on the expected posterior covariance matrix and on the Shannon information.

2.2.1 Fisher Information Matrix (FIM)

So far, the most widely used design criterion for constructing choice-based conjoint experiments is based on the asymptotic covariance matrix of the maximum likelihood estimator. According to the Rao-Cramer inequlity, the inverse of the Fisher information matrix is the asymptotic covariance matrix of the best asymptotically normal estimators. Thus designs which yield maximal Fisher information are associated with minimal asymptotic covariance. For the conditional logit model, the Fisher information matrix has the following expression:

$$\begin{aligned} \boldsymbol{\mathcal{I}}_{FIM}(\boldsymbol{\beta}, \mathbf{X}) &= -E_{\mathbf{Y}}\left[\frac{\partial^2 \log[L(\mathbf{y}|\boldsymbol{\beta})]}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\right], \\ &= N\sum_{s=1}^{S}\mathbf{X}_s'(\mathbf{P}_s - \mathbf{p}_s\mathbf{p}_s')\mathbf{X}_s, \end{aligned} \tag{4}$$

where N is the number of respondents, $\mathbf{X}_s$ is the design matrix for choice set s, $\mathbf{P}_s = \text{diag}\,[p_{1s}, p_{2s}, ..., p_{Ks}]$ and $\mathbf{p}_s = [p_{1s}, p_{2s}, ..., p_{Ks}]'$.

To take into account the uncertainty of the parameter values, the Fisher information matrix is used in a Bayesian framework. Let $\pi_0^D(\boldsymbol{\beta})$ denote the prior distribution used for design. Then, the D-optimal Bayesian designs are constructed by minimizing either

$$\phi_{FIM}^A = \int \det\left[\boldsymbol{\mathcal{I}}_{FIM}(\boldsymbol{\beta}|\mathbf{X})\right]^{-1/p}\ \pi_0^D(\boldsymbol{\beta})\ d\boldsymbol{\beta} \tag{5}$$

or

$$\phi_{FIM}^B = \det\left[\int \boldsymbol{\mathcal{I}}_{FIM}(\boldsymbol{\beta}|\mathbf{X})\ \pi_0^D(\boldsymbol{\beta})\ d\boldsymbol{\beta}\right]^{-1/p}. \tag{6}$$

We call ϕ_{FIM}^A the FIM type A criterion and ϕ_{FIM}^B the FIM type B criterion. Criterion ϕ_{FIM}^A minimizes the expected determinant of the maximum likelihood covariance matrix over the design prior, and has been considered by many authors in constructing choice experiments (Bliemer et al. 2008; Sándor and Wedel 2001, 2005; Kessels et al. 2006; Yu et al. 2008; Vermeulen et al. 2008). Criterion ϕ^B_{FIM} which maximizes the determinant of the expected Fisher information has not been used frequently. Atkinson and Donev (1992) constructed one parameter designs for a truncated model using both ϕ^A_{FIM} and ϕ^B_{FIM}. They found a striking difference between both designs. Note that ϕ^A_{FIM} and ϕ^B_{FIM} are essentially semi-Bayesian design criteria as they are derived in a non-Bayesian maximum likelihood estimation context and only the design prior $\pi^D_0(\boldsymbol{\beta})$ was taken into account in these criteria.

2.2.2 Generalized Fisher Information Matrix ($GFIM$)

In Bayesian estimation, one uses prior information to compute the posterior distribution of the model parameters. We call this prior distribution the inference prior, $\pi^I_0(\boldsymbol{\beta})$. In most cases, the inference prior, $\pi^I_0(\boldsymbol{\beta})$, is identical to the design prior,$\pi^D_0(\boldsymbol{\beta})$. The posterior distribution can be approximated by a normal distribution with the posterior mode and the inverse of the Generalized Fisher information matrix as the mean and the covariance matrix. The Generalized Fisher information matrix ($GFIM$) is computed as minus the Hessian matrix of the log posterior density and given by

$$\begin{aligned}
\boldsymbol{\mathcal{I}}_{GFIM}(\boldsymbol{\beta}, \mathbf{X}) &= -E_{\mathbf{Y}}\left[\frac{\partial^2 \log\left[L(\mathbf{y}|\boldsymbol{\beta})\ \pi^I_0(\boldsymbol{\beta})\right]}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\right], \\
&= E_{\mathbf{Y}}\left[-\frac{\partial^2 \log L(\mathbf{y}|\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'} - \frac{\partial^2 \pi^I_0(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\right], \\
&= \boldsymbol{\mathcal{I}}_{FIM}(\boldsymbol{\beta}, \mathbf{X}) - E_{\mathbf{Y}}\left[\frac{\partial^2 \pi^I_0(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\right],
\end{aligned} \tag{7}$$

where $\boldsymbol{\mathcal{I}}_{FIM}(\boldsymbol{\beta}, \mathbf{X})$ is given in (4). When the inference prior follows a multivariate normal distribution with covariance $\boldsymbol{\Sigma}_{I0}$, the second term in the last expression in (7) is simplified to

$$E_Y\left[\frac{\partial^2 \pi^I_0(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\right] = -\boldsymbol{\Sigma}^{-1}_{I0}, \tag{8}$$

and

$$\boldsymbol{\mathcal{I}}_{GFIM}(\boldsymbol{\beta}, \mathbf{X}) = \boldsymbol{\mathcal{I}}_{FIM}(\boldsymbol{\beta}, \mathbf{X}) + \boldsymbol{\Sigma}^{-1}_{I0}. \tag{9}$$

In this paper, we focus on a multivariate normal prior as it is the standard prior used by many authors (Han and Chaloner 2004; Sándor and Wedel 2001, 2005; Tsutakawa 1972, 1980; Vermeulen et al. 2008; Yu et al. 2008;). The two asymptotic Bayesian design criteria based on (9) are

$$\phi^A_{GFIM} = \int \det\left[\mathcal{I}_{FIM}(\boldsymbol{\beta}|\mathbf{X}) + \boldsymbol{\Sigma}^{-1}_{I0}\right]^{-1/p} \ \pi^D_0(\boldsymbol{\beta}) \ d\boldsymbol{\beta} \tag{10}$$

and

$$\phi^B_{GFIM} = \det\left[\int \left(\mathcal{I}_{FIM}(\boldsymbol{\beta}|\mathbf{X}) + \boldsymbol{\Sigma}^{-1}_{I0}\right) \ \pi^D_0(\boldsymbol{\beta}) \ d\boldsymbol{\beta}\right]^{-1/p}. \tag{11}$$

We call ϕ^A_{GFIM} the $GFIM$ type A optimality criterion and ϕ^B_{GFIM} the $GFIM$ type B optimality criterion. The corresponding Bayesian D-optimal design is the one that minimizes either (10) or (11). From the expression of the $GFIM$ criteria, we notice that the FIM criteria introduced in (5) and (6) are limiting cases of the $GFIM$ design criteria when non informative inference prior distribution is considered, or equivalently, when $\boldsymbol{\Sigma}^{-1}_{I0}$ in (10) and (11) is close to the zero matrix.

2.2.3 Expected Posterior Covariance Matrix ($EPCV$)

The posterior covariance matrix which measures the accuracy of the Bayesian estimator after the experiment has been conducted is given by

$$\begin{aligned} Var(\boldsymbol{\beta}|\mathbf{y}) &= \int \left[\boldsymbol{\beta} - \bar{\boldsymbol{\beta}}(\mathbf{y})\right] \left[\boldsymbol{\beta} - \bar{\boldsymbol{\beta}}(\mathbf{y})\right]' q(\boldsymbol{\beta}|\mathbf{y}) \ d\boldsymbol{\beta} \\ &= \frac{1}{p^I_Y(\mathbf{y})} \int \left[\boldsymbol{\beta} - \bar{\boldsymbol{\beta}}(\mathbf{y})\right] \left[\boldsymbol{\beta} - \bar{\boldsymbol{\beta}}(\mathbf{y})\right]' L(\mathbf{y}|\boldsymbol{\beta}) \ \pi^I_0(\boldsymbol{\beta}) \ d\boldsymbol{\beta}, \end{aligned}$$

where $p^I_Y(\mathbf{y})$ is the marginal distribution of $\mathbf{Y}$:

$$p^I_Y(\mathbf{y}) = \int L(\mathbf{y}|\boldsymbol{\beta}) \ \pi^I_0(\boldsymbol{\beta}) \ d\boldsymbol{\beta}. \tag{12}$$

The posterior mean $\bar{\boldsymbol{\beta}}(\mathbf{y})$ is given by

$$\bar{\boldsymbol{\beta}}(\mathbf{y}) = \int \boldsymbol{\beta} \ q(\boldsymbol{\beta}|\mathbf{y}) \ d\boldsymbol{\beta}. \tag{13}$$

Suppose that a possibly different prior distribution $\pi^D_0(\boldsymbol{\beta})$ is used for constructing the design. The goal is to find a design that minimizes the determinant of the expected posterior covariance matrix $E^D_Y\left[Var(\boldsymbol{\beta}|\mathbf{y})\right]$, where the expectation is with respect to the marginal distribution of $\mathbf{y}$ when the prior distribution is the design prior $\pi^D_0(\boldsymbol{\beta})$. Note that the expected posterior covariance matrix is the posterior covariance matrix one expects before observing the data. Let $f_I(\mathbf{y},\boldsymbol{\beta}) = L(\mathbf{y}|\boldsymbol{\beta})\pi_0^I(\boldsymbol{\beta})$ and $f_D(\mathbf{y},\boldsymbol{\beta}) = L(\mathbf{y}|\boldsymbol{\beta})\pi_0^D(\boldsymbol{\beta})$. Then,

$$\begin{aligned} E_Y^D\left[Var(\boldsymbol{\beta}|\mathbf{y})\right] &= \int Var(\boldsymbol{\beta}|\mathbf{y})\ p_Y^D(\mathbf{y})\ d\mathbf{y} \\ &= \int \left[\frac{1}{p_Y^I(\mathbf{y})} \int \left[\boldsymbol{\beta} - \bar{\boldsymbol{\beta}}(\mathbf{y})\right] \left[\boldsymbol{\beta} - \bar{\boldsymbol{\beta}}(\mathbf{y})\right]'\ f_I(\mathbf{y},\boldsymbol{\beta})\ d\boldsymbol{\beta}\right] \times \left[\int f_D(\mathbf{y},\boldsymbol{\beta})\ d\boldsymbol{\beta}\right] d\mathbf{y} \quad (14) \\ &= \int\int \left[\boldsymbol{\beta} - \bar{\boldsymbol{\beta}}(\mathbf{y})\right] \left[\boldsymbol{\beta} - \bar{\boldsymbol{\beta}}(\mathbf{y})\right]'\ f_I(\mathbf{y},\boldsymbol{\beta})\ d\boldsymbol{\beta} \times \frac{\int f_D(\mathbf{y},\boldsymbol{\beta})d\boldsymbol{\beta}}{\int f_I(\mathbf{y},\boldsymbol{\beta})d\boldsymbol{\beta}} d\mathbf{y}. \end{aligned}$$

If the design prior is identical to the inference prior, then $f_I(\mathbf{y},\boldsymbol{\beta}) = f_D(\mathbf{y},\boldsymbol{\beta})$ and the expected posterior covariance matrix can be simplified to:

$$E_Y^D\left[Var(\boldsymbol{\beta}|\mathbf{y})\right] = \int\int \left[\boldsymbol{\beta} - \bar{\boldsymbol{\beta}}(\mathbf{y})\right] \left[\boldsymbol{\beta} - \bar{\boldsymbol{\beta}}(\mathbf{y})\right]'\ f_I(\mathbf{y},\boldsymbol{\beta})\ d\boldsymbol{\beta}\ d\mathbf{y}. \quad (15)$$

In this paper, we assume $\pi_0^I(\boldsymbol{\beta})$=$\pi_0^D(\boldsymbol{\beta})$ and denote this prior by $\pi_0(\boldsymbol{\beta})$. This assumption is realistic because the design and analysis are often conducted by the same person. The expected posterior covariance matrix given in expression (15) can then be written as

$$\phi_{EPCV} = E_Y^D\left[Var(\boldsymbol{\beta}|\mathbf{y})\right] = \int\int \left[\boldsymbol{\beta} - \bar{\boldsymbol{\beta}}(\mathbf{y})\right] \left[\boldsymbol{\beta} - \bar{\boldsymbol{\beta}}(\mathbf{y})\right]'\ L(\mathbf{y}|\boldsymbol{\beta})\ \pi_0(\boldsymbol{\beta})\ d\boldsymbol{\beta}\ d\mathbf{y}. \quad (16)$$

To simplify the computation, we approximate the likelihood by a normal distribution. Rossi et al. (2005) stated that for the conditional logit likelihood, the normal approximation is excellent. The design which minimizes the determinant of the expected posterior covariance matrix in (16) is desirable.

2.2.4 Shannon information

In a Bayesian information theoretic approach, the optimal design is chosen by maximizing the expected gain in Shannon information or, equivalently, maximizing the amount of information provided by the experiment. This gain can be assessed by comparing the information in the prior and in the posterior distribution.

The amount of information associated with a prior distribution $\pi_0^D(\boldsymbol{\beta})$ is defined as

$$g_0 = E_{\boldsymbol{\beta}}\left\{\log[\pi_0^D(\boldsymbol{\beta})]\right\} = \int \log[\pi_0^D(\boldsymbol{\beta})]\ \pi_0^D(\boldsymbol{\beta})\ d\boldsymbol{\beta}. \quad (17)$$

After the experiment has been performed, the prior distribution is updated to the posterior distribution, $q(\boldsymbol{\beta}|\mathbf{y})$, which has an information content equal to

$$\begin{aligned} g_1(\mathbf{y}) &= \int \log[q(\boldsymbol{\beta}|\mathbf{y})]\ q(\boldsymbol{\beta}|\mathbf{y})\ d\boldsymbol{\beta} \\ &= \frac{1}{p^I_{\mathbf{Y}}(\mathbf{y})} \int \log[q(\boldsymbol{\beta}|\mathbf{y})]\ L(\mathbf{y}|\boldsymbol{\beta})\ \pi^I_0(\boldsymbol{\beta})\ d\boldsymbol{\beta}, \\ &= \frac{1}{p^I_{\mathbf{Y}}(\mathbf{y})} \int \log[q(\boldsymbol{\beta}|\mathbf{y})]\ f_I(\mathbf{y},\boldsymbol{\beta})\ d\boldsymbol{\beta}. \end{aligned} \tag{18}$$

The amount of information provided by the experiment with design matrix X when the design prior is $\pi^D_0(\boldsymbol{\beta})$, the inference prior is $\pi^I_0(\boldsymbol{\beta})$ and the response is $\mathbf{y}$, is

$$g(\mathbf{X}, \pi^D_0(\boldsymbol{\beta}), \pi^I_0(\boldsymbol{\beta}), \mathbf{y}) = g_1(\mathbf{y}) - g_0. \tag{19}$$

Notice that $g(\mathbf{X}, \pi^D_0(\boldsymbol{\beta}), \pi^I_0(\boldsymbol{\beta}), \mathbf{y})$ depends on the responses $\mathbf{y}$. Since they are not yet available when the experiment is set up, the design criterion is based on the expected Shannon information provided by the experiment. It is computed by taking the expectation over the marginal distribution of the data when the prior is $\pi^D_0(\boldsymbol{\beta})$. Denoting $p^D_Y(\mathbf{y}) = \int L(\mathbf{y}|\boldsymbol{\beta})\pi^D_0(\boldsymbol{\beta})d\boldsymbol{\beta}$, the expected Shannon information provided by the experiment is defined as

$$g(\mathbf{X}, \pi^D_0(\boldsymbol{\beta}), \pi^I_0(\boldsymbol{\beta})) = E^D_{\mathbf{Y}}[g_1(\mathbf{y}) - g_0] = \int g_1(\mathbf{y})\ p^D_Y(\mathbf{y})\ d\mathbf{y} - g_0. \tag{20}$$

A design that maximizes the expected Shannon information provided by the experiment is desirable. As g_0 in (20) does not depend on the design, maximizing (20) is equivalent to maximizing $\int g_1(\mathbf{y})p^D_Y(\mathbf{y})d\mathbf{y}$, which is the expected Shannon information of the posterior distribution. Similar to the expected posterior covariance matrix, the design criterion based on the Shannon information is given by

$$\begin{aligned} \int g_1(\mathbf{y})p^D_Y(\mathbf{y})d\mathbf{y} &= \int \left[\frac{1}{p^I_{\mathbf{Y}}(\mathbf{y})} \int \log[q(\boldsymbol{\beta}|\mathbf{y})]\ f_I(\mathbf{y},\boldsymbol{\beta})\ d\boldsymbol{\beta}\right] \times \left[\int f_D(\mathbf{y},\boldsymbol{\beta})\ d\boldsymbol{\beta}\right] d\mathbf{y} \\ &= \int\int \log[q(\boldsymbol{\beta}|\mathbf{y})]\ f_I(\mathbf{y},\boldsymbol{\beta})\ d\boldsymbol{\beta} \times \frac{\int f_D(\mathbf{y},\boldsymbol{\beta})\ d\boldsymbol{\beta}}{\int f_I(\mathbf{y},\boldsymbol{\beta})\ d\boldsymbol{\beta}} d\mathbf{y}. \end{aligned} \tag{21}$$

If $\pi^D_0(\boldsymbol{\beta}) = \pi^I_0(\boldsymbol{\beta}) = \pi_0(\boldsymbol{\beta})$, then the above expression reduces to:

$$\begin{aligned} \phi_{Shannon} = \int g_1(\mathbf{y})\ p^D_Y(\mathbf{y})\ d\mathbf{y} &= \int\int \log[q(\boldsymbol{\beta}|\mathbf{y})]\ f_I(\mathbf{y},\boldsymbol{\beta})\ d\boldsymbol{\beta}\ d\mathbf{y} \\ &= \int\int \log\left(\frac{L(\mathbf{y}|\boldsymbol{\beta})\pi_0(\boldsymbol{\beta})}{\int L(\mathbf{y}|\boldsymbol{\beta})\pi_0(\boldsymbol{\beta})d\boldsymbol{\beta}}\right)\ L(\mathbf{y}|\boldsymbol{\beta})\ \pi_0(\boldsymbol{\beta})\ d\boldsymbol{\beta}\ d\mathbf{y}. \end{aligned} \tag{22}$$

2.2.5 Computational Issues

To compute optimal designs using these criteria, several highly dimensional integrals have to be computed precisely. For the integral $\int L(\mathbf{y}|\boldsymbol{\beta})\pi_0(\boldsymbol{\beta})d\boldsymbol{\beta}$ in (22) and for the posterior mean $\bar{\boldsymbol{\beta}}$ in (13), we used the Laplace approximation (Bradley and Thomas 1996). To make the construction of the designs based on the expected posterior covariance matrix and the Shannon information feasible, we used systematic draws instead of Monte Carlo draws to approximate the integrals in expression (16) and (22). The outer integral in both expression computed using the randomized spherical-radial theory was which has been used by Gotwalt et al. (2007) and Monahan and Genz (1997). The inner integral in (16) and (22) was computed using the extensible shifted lattice points transformed by Baker's transformation based on the work of Sándor and András (2004) and Hickernell et al. (2000). Furthermore, we used the well-known coordinate-exchange algorithm to search for the best design (Meyer and Nachtsheim 1995; Kessels et al. 2008; Yu et al. 2008a). The accuracy of the computations was checked and is reported in detail in the technical report (Yu et al. 2008b)

3 Relative Design Efficiency

Let $D_C(\mathbf{X})$ denote the value of design criterion C for design $\mathbf{X}$ where C is one of the six design criteria discussed in this paper. The relative design efficiency of any pair of designs $\mathbf{X}_1$ and $\mathbf{X}_2$ in terms of criterion C is computed as $D_C(\mathbf{X}_1)/D_C(\mathbf{X}_2)$. It measures how efficient design $\mathbf{X}_2$ is relative to design $\mathbf{X}_1$ when evaluated by criterion C. Values larger than one are obtained if design $\mathbf{X}_2$ is more efficient than $\mathbf{X}_1$ according to criterion C.

In this section, we examine how well the different Bayesian designs perform with respect to the design criteria for which they were not optimized. More specifically, suppose that design $\mathbf{X}^*$ is the optimal design constructed with design criterion ϕ^A_{FIM}. The relative efficiency of any other design $\mathbf{X}$ relative to design $\mathbf{X}^*$, $\frac{D^A_{FIM}(\mathbf{X}^*)}{D^A_{FIM}(\mathbf{X})}$, enables us to explore how good design $\mathbf{X}$ is compared to the optimal design $\mathbf{X}^*$ when the goal of the experiment is to minimize the ϕ^A_{FIM} criterion.

To investigate a wide variety of situations while keeping the computations manageable, we consider a design problem with specification $3^2/2/6$, that is, 6 choice sets with 2 alternatives per choice set and 2 attributes, each at 3 levels. The prior is specified as $\boldsymbol{\beta} \sim N(\boldsymbol{\mu}_0, \mathbf{I}_4)$, where $\boldsymbol{\mu}_0$=(-1 0 -1 0) and $\mathbf{I}_4$ is the 4-dimensional identity matrix.

The criterion values were computed for the parameter values assumed when constructing the designs. That is, the comparison was conducted under the assumption that the prior was correctly specified. This allowed us to investigate which design is the least sensitive to the design criterion. Table 1 shows that in general, the efficiencies of the design based on the expected posterior covariance matrix ($EPCV$) are quite high for all other criteria.

This implies that the $EPCV$ criterion enabled us to find a design which performs well for different purposes of the choice experiments. For example, the $EPCV$ design is efficient not only for a Bayesian analysis, but also for a non-Bayesian analysis where the inference prior is ignored in the analysis and the asymptotic covariance matrix of the maximum likelihood estimates (the inverse of the Fisher information matrix) is minimized. This is also the case for the design based on the Shannon information.

Comparing the ϕ^A_{FIM} and ϕ^B_{FIM} designs to the ϕ^A_{GFIM} and ϕ^B_{GFIM} designs learns that the $GFIM$ criteria in general, lead to better efficiencies than the corresponding FIM criteria. Comparing the ϕ^A_{FIM} design with the ϕ^B_{FIM} design, we found that the FIM type A criterion led to a design which is more robust to the other design criteria than the FIM type B. Similarly, the $GFIM$ type A criterion led to a design which is more robust than the $GFIM$ type B criterion. The Bayesian design generated by ϕ^B_{FIM} is the least robust to other design criteria.

Table 1: Evaluation of the Bayesian optimal designs in terms of other design criteria

		Evaluation based on					
		ϕ^A_{FIM}	ϕ^B_{FIM}	ϕ^A_{GFIM}	ϕ^B_{GFIM}	ϕ_{EPCV}	$\phi_{Shannon}$
Designs based on	ϕ^A_{FIM}	100.00%	99.95%	99.95%	99.99%	87.3%	90.69%
	ϕ^B_{FIM}	58.44%	100.00%	86.37%	99.97%	66.53%	74.25%
	ϕ^A_{GFIM}	99.88%	96.91%	100.00%	97.55%	93.36%	93.58%
	ϕ^B_{GFIM}	88.1%	99.96%	96.62%	100.00%	87.73%	94.2%
	ϕ_{EPCV}	98.20%	99.56%	98.54%	96.08%	100.00%	97.92%
	$\phi_{Shannon}$	95.68%	96.15%	99.41%	97.60%	96.07%	100.00%

4 Simulation Study

The simulation study consists of two parts. In the first part, we check whether the inverse of the FIM and $GFIM$ are good approximations of the true posterior covariance matrix. In the second part, we investigate how much we sacrifice in terms of design efficiency by applying the simpler asymptotic Bayesian design criteria instead of the computationally more demanding ϕ_{EPCV}.

4.1 Validity of the Asymptotic Approximation

In this section, we examine the closeness of the inverse of the FIM and $GFIM$ to the posterior covariance matrix under various scenarios. Since we are working on the multi-dimensional parameters, it is difficult to compare different matrices directly. Therefore, we use a widely accepted scalar measure and compute the determinants of the inverse of the

FIM, $GFIM$ and the posterior covariance matrices, to investigate how close these matrices are to each other.

We drew $r = 1, ..., 512$ true parameter vectors from $N(\boldsymbol{\mu}_0, \mathbf{I}_4)$, $\boldsymbol{\mu}_0$=(-1 0 -1 0). For a given design, a given true parameter vector $\boldsymbol{\beta}^r$ and a given number of respondents, we simulated 1000 sets of responses. For each set, the posterior covariance matrix was computed using the Markov Chain Monte Carlo (MCMC) approach. We then averaged their determinants over the 1000 data sets in order to get a reliable result. This result was compared to the determinants obtained from the asymptotic approximations. More specifically, for each true parameter vector $\boldsymbol{\beta}^r$, we compute

$$\left|[\boldsymbol{\mathcal{I}}_{FIM}(\boldsymbol{\beta}, \mathbf{X})]^{-1}\right|, \tag{23}$$

and

$$\left|\left[\boldsymbol{\mathcal{I}}(\boldsymbol{\beta}, \mathbf{X}) + \frac{1}{(\sigma \mathbf{I}_4)}\right]^{-1}\right|. \tag{24}$$

We then computed the percentage difference between the determinants from the posterior covariance matrix and these approximations. Finally, we averaged all these percentage differences over the 512 draws $\boldsymbol{\beta}^r$. Since the performance of the approximations to the posterior covariance matrix depends on the sample size, we took 7 different numbers of respondents between 10 and 70. In addition, as the quality of the $GFIM$ approximation also depends on σ, we took 15 values of σ between 1 and 20.

We summarize the results in Figure 1. The Y-axis represents the percentage difference between the posterior covariance matrix and its approximations. The X-axis shows the number of respondents. The results obtained from FIM and $GFIM$ with $\sigma = 1, 5, 10$ are shown. All curves decrease with the number of respondents, which was to be expected. In addition, it is clear from the plot that the inverse of the FIM leads to very inefficient approximations to the posterior covariance matrix when the sample size is small. The $GFIM$ with relatively small σ provides a significantly better approximation compared to FIM. The FIM curve lies higher than above all other curves as the asymptotic approximation based on the FIM is the upper bound for all the approximations by $GFIM$ when σ goes to infinity. For any given number of respondents, the inverse of the $GFIM$ approximation is closer to the posterior covariance matrix than the inverse of the FIM approximation. However, the advantage of using the $GFIM$ over the FIM decreases as the sample size increases.

In Figure 1, we present the results for only a few cases among the $7 \times 15 = 85$ different combinations of σ and the number of respondents n. To get a more detailed picture, we plot all the combinations of these two parameters in Figure 2. This plot enables us to visualize the comparison between the $GFIM$ and FIM approximations under all scenarios we have studied.

For each true parameter, we first compute the percentage difference between (23) and (24) from the determinant of the true posterior covariance matrix, respectively. We then took the ratio of the deviation obtained from the $GFIM$ over that obtained from the FIM, and averaged this ratio over the 512 true parameters β^r. The smaller the value on the plot, or equivalently, the lighter the color is, the larger the advantage of using the $GFIM$ approach. A ratio with value close to 1 indicates that the inverses of the $GFIM$ and the FIM lead to almost the same error in approximating the posterior covariance matrix.

Figure 2 clearly demonstrates the domains where the inverse of the $GFIM$ is most appealing and where it only has little advantage over the inverse of the FIM in approximating the posterior covariance matrix. These domains correspond to the pink, blue and green area, and to the deep orange area, respectively. In addition, the color of the plot changes from the right to the left and from the front to the back which demonstrates how the relative performance of the $GFIM$ and FIM changes with the value of σ and the sample size. It is clear that the relative performance of the two designs strongly depends on σ. For reasonably small values of σ, the inverse of the $GFIM$ is a much better approximation to the posterior covariance matrix. That the $GFIM$ converges to the FIM as the value of σ increases is now clearly visualized in Figure 2.

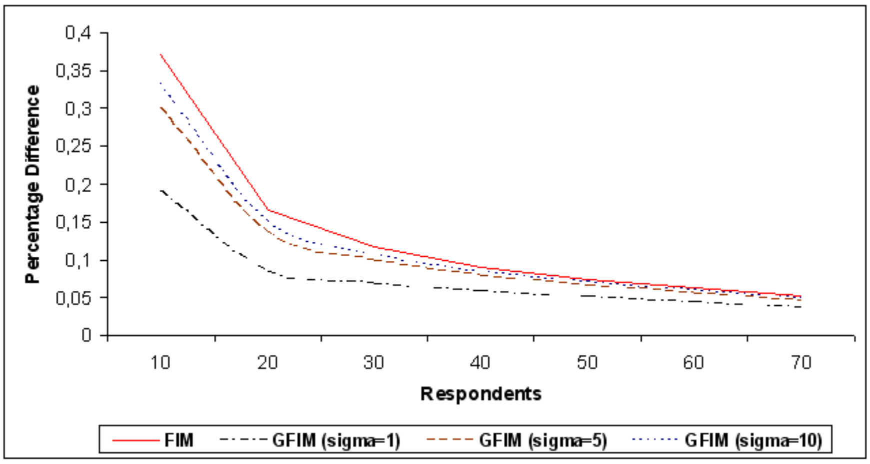


Figure 1: Percentage difference between the posterior covariance matrix and its approximations

4.2 Comparing Design Sensitivity to Prior Specification

In this section, we compare the six optimal designs constructed with the design criteria introduced in Section 2 under a wide variety of parameter spaces. We used 10 respondents in the study. The goal is to examine how well these designs perform when the sample size is

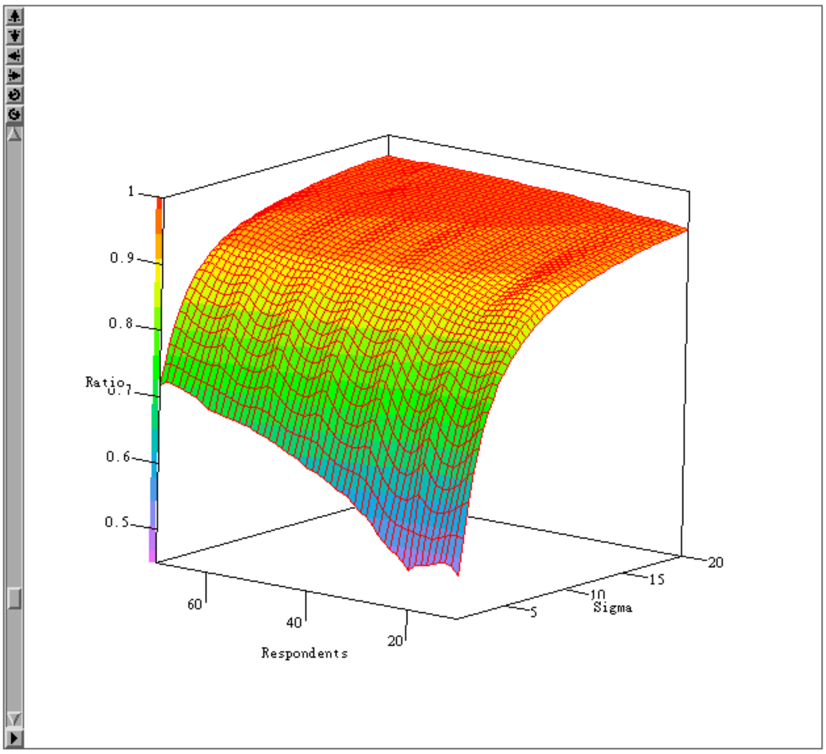


Figure 2: Ratio of the deviation of the $GFIM$ from the posterior covariance matrix over the deviation of the FIM

small, and to study the robustness of each design when prior information about the parameters is incorrect. The comparison is done in both a Bayesian and a non-Bayesian framework.

Since the prior specification does have an impact on the performance of the resulting designs, it is interesting to investigate whether the relative performance of different designs is sensitive to the choice of the prior mean as well as to the prior covariance matrix. We evaluated the performance of the six designs in 465 different parameter spaces. For each parameter space, the true parameters were drawn from a different multivariate normal distribution defined by a different combination of the mean $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$. The mean $\boldsymbol{\mu}$ was specified as $\boldsymbol{\mu} = \boldsymbol{\mu}_0 + \lambda[\mathbf{1}_4]$, where λ reflects the deviation of the true mean $\boldsymbol{\mu}$ from the assumed mean $\boldsymbol{\mu}_0$ and took values between [-1.5, 1.5]. The covariance matrix $\boldsymbol{\Sigma}$ was specified as $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}_4$, where σ took values between [0.2, 3].

These parameter spaces allow us to study the impact of well defined prior distributions and poorly defined prior distributions when constructing the designs. Under well specified prior distributions, the true parameters do not deviate much from the ones assumed in the design construction. With poorly defined prior distributions, the true parameters can lie in the tail of the design prior or they are not even covered by the design prior. The larger the value of σ or the larger the absolute value of the λ, the more likely the prior is poorly defined. We hope to identify designs that are robust so that if the design prior is misspecified to a reasonable extent, it is still possible to obtain efficient parameter estimates.

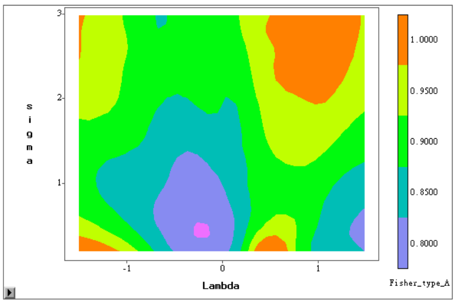
We first investigate how well each design performs in a Bayesian framework when the prior distribution is taken into account for design and for analysis. To study this, we compare the efficiencies of the six designs in each parameter space. We drew 90 true parameter vectors $\boldsymbol{\beta}^r$ from each parameter space. For each $\boldsymbol{\beta}^r$ and for each Bayesian design $\mathbf{X}$, we simulated data 128 times and computed the corresponding posterior covariances. The expected posterior covariance matrix for each $\boldsymbol{\beta}^r$ was computed by averaging 128 computed posterior covariances. We then took the determinant of the expected posterior covariance matrix for each $\boldsymbol{\beta}^r$. We consider the Bayesian design constructed with the ϕ_{EPCV} criterion as the benchmark design. From the Bayesian perspective, the ϕ_{EPCV} design is a natural one to serve as the benchmark. For each draw of the true parameter $\boldsymbol{\beta}^r$, the relative efficiency of design $\mathbf{X}$ over the benchmark design was computed by taking the ratio of the determinant obtained from the benchmark design over that from design $\mathbf{X}$. For each parameter space, we averaged the ratios over all parameter draws. The results for the different parameter spaces and different designs are presented in the contour plot in Figure 3.

We also conducted the comparison in a non-Bayesian framework where it is accepted that prior information is used for design but should not be used in the analysis. The benchmark design for this comparison was constructed by the ϕ^A_{FIM} criterion as it is the most commonly used Bayesian design criterion for the classic non-Bayesian analysis. Here each design is evaluated by the ϕ^A_{FIM} criterion. The efficiency of each design relative to the benchmark design was computed and is shown in Figure 4 for all parameter spaces.

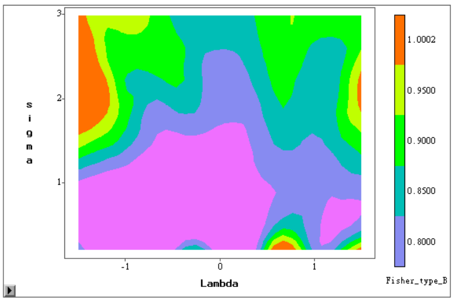
4.2.1 Comparisons in a Bayesian Framework

In this section, we examine how good each design is for making efficient Bayesian analyses. The five contour plots in Figure 3 enable us to explore the sensitivity of each design to the specification of the prior distribution. Each plot presents the relative efficiency of the design over the benchmark design under various parameter spaces. The value of λ, which reflects how much the mean of the multivariate normal distribution from which the true parameters are drawn deviates from the one assumed in generating the design, is displayed on the horizontal axis. The vertical axis presents the values of σ defined in Section 4.2.

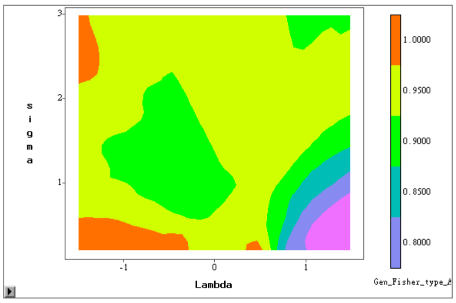
The relative efficiencies in each plot in Figure 3 are classified into six categories. Each color on the plot represents a category with a specific range of efficiencies. For example, the pink area shows the relative efficiencies with values below 0.8, while the orange area presents those with values above 1. For a specific plot, the orange area thus shows those conditions



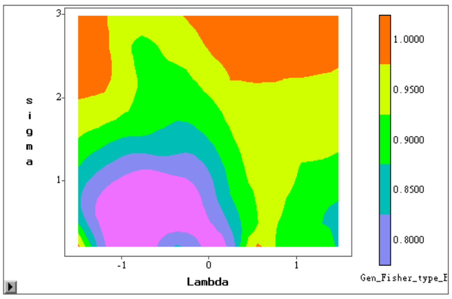
(a) Fisher information type A



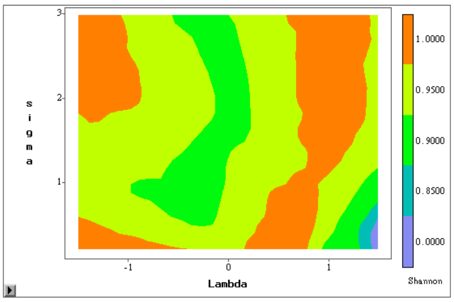
(b) Fisher information type B



(c) Generalized Fisher information type A



(d) Generalized Fisher information type B



(e) Shannon Information

Figure 3: Efficiencies relative to the $EPCV$ optimal design in a Bayesian framework

under which the alternative design is more efficient than the benchmark. The larger the green, blue, purple and pink areas, the less efficient the associated design compared to the benchmark design.

The large non-orange area in each plot indicates that the ϕ_{EPCV}-optimal design is generally more efficient than the alternatives under various scenarios. This implies that the $EPCV$ criterion leads to designs which are robust to the prior specification. Most of the area in plot (c) is colored by yellow and green which indicates that the asymptotic $GFIM$ type A criterion does lead to a reasonably efficient Bayesian conjoint choice design. Plot (e) tells us that the design criterion which aims at maximizing the expected gain in Shannon information also performs well for constructing designs for efficient Bayesian analysis. Comparing Plot (d) to Plot (c) learns us that the ϕ^B_{GFIM} criterion leads to a design which is less stable than that constructed with the ϕ^A_{GFIM} criterion.

The lower parts of Plot (a) and (b) show us that the FIM criteria are rather inefficient in constructing optimal Bayesian designs for efficient Bayesian inference compared to the $EPCV$ criterion. This area includes those parameter spaces where the true parameters are not far from the one assumed in the design construction or equivalently, when the prior information is correctly specified or misspecified to a reasonable extent. Under these situations, the loss in efficiency due to the use of the popular FIM criteria instead of the $EPCV$ criterion for constructing choice experiments is large. This implies that the popular asymptotic Bayesian design criteria based on the Fisher information matrix are not adequate for generating choice experiments in a fully Bayesian setting. Note that the sacrifice is even larger for ϕ^B_{FIM} than for ϕ^A_{FIM}.

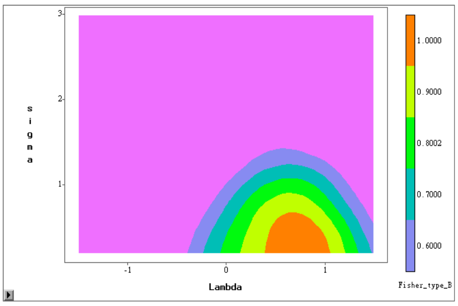
4.2.2 Comparisons in a non-Bayesian framework

In the previous section, we showed the good performance of design criteria such as the expected posterior covariance matrix, the $GFIM$ and the Shannon information in constructing choice experiments for efficient Bayesian inference. In this section, we are interested in examining how these criteria perform in a non-Bayesian framework. Recall that the benchmark design used here was constructed with the ϕ^A_{FIM} criterion because it is the most commonly used criterion for constructing choice experiments in the non-Bayesian setting. As in Figure 3, the relative efficiencies in Figure 4 are classified into six categories. The colors on the plot have the same interpretation as in Figure 3 but correspond to larger ranges of relative efficiencies.

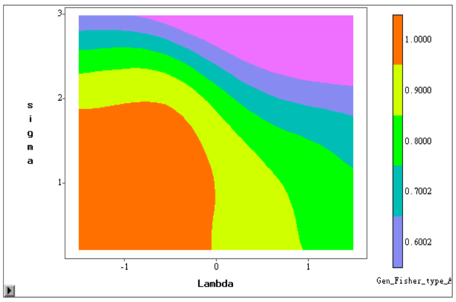
Plot (d) reveals that, in general, the design constructed with the $EPCV$ criterion is also efficient for non-Bayesian analysis. This particularly holds for those conditions the results of which are displayed in the lower part of Plot (d). The upper part of Plot (d) contains results for those parameter spaces with a lot of true parameters that are not covered by the design prior. In those situations, the $EPCV$ design becomes less efficient for non-Bayesian estimation compared to the benchmark design.

The large orange and yellow areas in Plot (b) correspond to the relative efficiencies which are higher than 90%. This implies that the ϕ^A_{GFIM} design performs well for the non-Bayesian analysis in those scenarios. Compared to the ϕ^A_{GFIM} design, the ϕ^B_{GFIM} design is less efficient in a non-Bayesian framework.

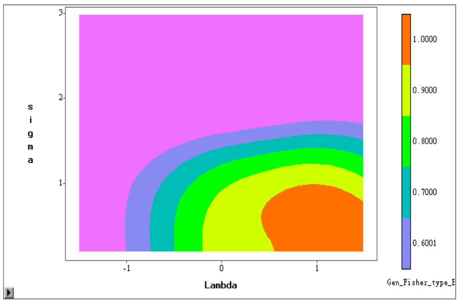
Compared to the design based on the expected posterior covariance matrix, the design



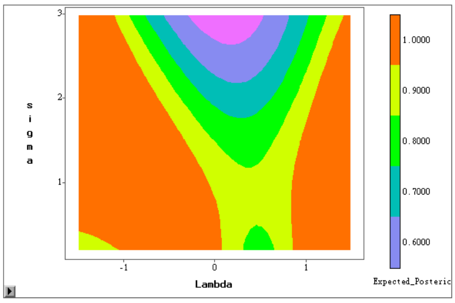
(a) Fisher information type B



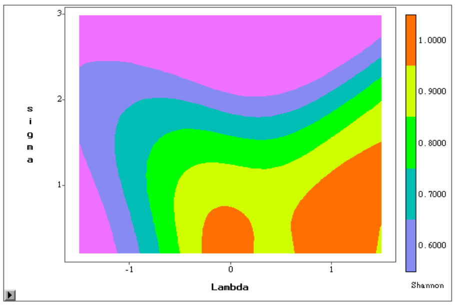
(b) Generalized Fisher information type A



(c) Generalized Fisher information type B



(d) Expected Posterior Covariance Matrix



(e) Shannon information

Figure 4: Design criteria comparison in terms of Relative efficiency in a non-Bayesian framework

based on the Shannon information in Plot (e) is less robust in a non-Bayesian context. However, it still provides good performance when the design prior is well specified or misspecified to a certain extent. Plot (a) shows the inefficiency of ϕ^{B}_{FIM} compared to ϕ^{A}_{FIM}.

4.3 Computational Complexity

In the previous section, it was shown that the design criterion based on the expected posterior covariance matrix is most appealing in most cases. However, computing the exact posterior distribution at each iteration of the design construction algorithm is not a trivial task and requires substantial computations. In many cases, the small marginal gains using the exact criterion were obtained at the expense of an excessive computational effort. In this section, we compare each of the six Bayesian design criteria in terms of computational convenience.

It turns out that the four asymptotic Bayesian criteria (ϕ^A_{FIM}, ϕ^B_{FIM}, ϕ^A_{GFIM} and ϕ^B_{GFIM}) are remarkably fast. For the design problem that we considered in this paper, the computation times for one try of the coordinate exchange algorithm to search for the best design based on each of these four asymptotic criteria were less than 0.5 seconds. In contrast, the computation times for computing designs based on the expected posterior covariance matrix and the Shannon information are much higher. For only one try of the algorithm, these two criteria require at least 21600 times more time than required by the asymptotic Bayesian criteria. This is because there might be millions of integrals related to the posterior distribution that need to be evaluated when searching for a best design by means of the expected posterior covariance matrix and the Shannon information.

The computation time also increases dramatically with the design size. For example, for a design problem with specification $3^3/2/12$, one needs more than 16 hours for a single try of the coordinate exchange algorithm. Given the good performance of the ϕ^A_{GFIM} criterion and its computational attractiveness, using the ϕ^A_{GFIM} criterion to construct Bayesian designs when the sample size is small is a sensible thing to do in practice.

5 Summary

This paper reviews six Bayesian criteria ranging from those based on the Fisher information matrix and the Generalized Fisher information matrix to those based on the expected posterior covariance matrix and the Shannon information to compute designs for conjoint choice experiments. First we investigated how robust the resulting Bayesian optimal designs are with respect to criteria for which they are not optimized. Then, we examined how close the inverse of the Fisher information matrix and the inverse of the Generalized Fisher information matrix are to the true posterior covariance matrix. Finally, we studied the quality of each Bayesian optimal choice designs. Especially, we investigated how much we sacrifice when using the simpler asymptotic design criteria instead of the computationally more involved expected posterior covariance matrix criterion, and we check the sensitivity of each design to the misspecification of the prior distribution.

Our study reveals that the efficiency of the $EPCV$ design is quite high when evaluated by other design criteria for which it was not optimized. This is also the case for the design based on the Shannon information. In contrast, the $GFIM$ designs are less robust with respect to the other design criteria. However, they still perform better than the FIM designs.

The simulation study leads to the conclusion that the inverse of the FIM might be a poor approximation to the posterior covariance matrix when the sample size is small. The results also show that the approximation based on the inverse of the $GFIM$ is superior to that based on the inverse of the FIM.

In addition, we show that the $EPCV$ design is quite robust to the misspecification of the prior but hard to compute. It is efficient not only in a Bayesian framework but also in a non-Bayesian framework under various degrees of misspecifying the prior distribution. An interesting finding is that although the Bayesian information theoretic approach does not explicitly aim at minimizing the posterior covariance of the parameter estimates when constructing a design, it does lead to a good design which is efficient for Bayesian analysis and non-Bayesian analysis under a number of scenarios, especially when the prior information about the possible values of the true parameters is properly defined. In contrast, it was shown that the widely used design criteria based on the inverse of the Fisher information matrix lead to inefficient designs for Bayesian estimation when the sample size is small.

The $GFIM$ type A design performs well in this study. It is shown that in a Bayesian framework, using $GFIM$ type A instead of the expected posterior covariance matrix for constructing choice experimental designs does not lead to a dramatic efficiency loss under most conditions. In a non-Bayesian framework, using the $GFIM$ type A criterion is also quite efficient for a large number of parameter spaces. We conclude that the asymptotic Bayesian design criterion $GFIM$ type A is a reasonable alternative to the Bayesian design criterion based on the expected posterior covariance matrix, and it is much cheaper to compute.

References

Anderson, D. A., J. B. Wiley. 1992. Efficient choice designs for estimating availability cross-effects models. *Marketing Letters* 3 357-370.

Atkinson, A. C., A. N. Donev. 1992. Optimum experimental designs. Oxford U. K. Clarendon Press.

Atkinson, A. C. and Haines, L. M. (1996). Designs for non-linear and generalized linear models. In Handbook of Statistics, Volume 13, S. Ghosh and C. R. Rao (eds), 437- 475. New York: Elsevier Science.

Berger, J. O. 1985. Statistical decision theory and Bayesian analysis. *Springer Series in Statistics* 224.

Bliemer, M. C. J., J. M. Rose, S. Hess. 2008. Approximation of Bayesian efficiency in experimental choice designs. To appear in *Journal of Choice Modeling*

Bradley, P. C., A. L. Thomas. 1996. Bayes and empirical bayes methods for data analysis. Chapman and Hall.

Deák, I. (1990), Random number generation and simulation. Budapest: Akademiniai Kiado.

Dror, H. A., D. M. Steinberg 2006. Robust experimental design for multivariate generalized liner models. *American Statistical Association and the American Society for Quality* 48 NO. 4.

Gotwalt, C. M., B. A. Jones, D. M. Steinberg. 2007. Fast computation of designs roust to parameter uncertainty for nonlinear settings. Working paper.

Han, C., K. Chaloner. 2004. Bayesian experimental design for nonlinear mixed-effects models with application to HIV dynamics. *Biometrics* 60 25-33.

Hickernell, F. J., H. S. Hong, P. L'Ecuyer, C. Lemieux. 2000. Extensible lattice sequences for quasi-Monte Carlo quadrature. *SIAM Journal on Scientific Computing* 22 1117-1138.

Huber, J., K. Zwerina. 1996. The importance of utility balance in efficient choice designs. *Journal of Marketing Research* 307-317.

Kessels, R., P. Goos, M. Vandebroek. 2006. A comparison of criteria to design efficient choice experiments. *Journal of Marketing Research* 43 409-419.

Kessels, R., B. Jones, P. Goos, M. Vandebroek. 2008. An efficient algorithm for constructing Bayesian optimal choice designs. To appear in Journal of Business and Industrial Statistics.

Lazari, A. G., D. A. Anderson. 1994. Designs of discrete choice experiments for estimating both attibute and availability cross effects. *Journal of Marketing Research* 31 375-383.

Lindley, D. V. 1956. On the measure of information provided by an experiment. *The Annals of Mathematical Statistics* 27 986-1005

Merlé, Y., F. Mentré. 1995. Bayesian design criteria: computation, comparison, and application to a pharmacokinetic and a pharmacodynamic model. *Journal of Pharmacokinetics and Biopharmaceutics* 23 101-125

Meyer, R. K., C. J. Nachtsheim. 1995. The coordinate-exchange algorithm for constructing exact optimal experimental designs. *Technometrics* 37 60-69.

Monahan, J., A. Genz. 1997. Spherical-Radial Integration Rules for Bayesian Computation. *Journal of the American Statistical Association* 92 664-674.

Rossi, P. E., G. M. Allenby, R. McCulloch. 2005. Bayesian statistics and marketing. *Wiley series in probability and statistics.*

Sándor, Z., M. Wedel. 2001. Designing conjoint choice experiments using managers' prior beliefs. *Journal of Marketing Research* 38 (November) 430-444.

Sńdor, Z., M. Wedel. 2005. Differentiated Bayesian conjoint choice designs. *Journal of Marketing Research* 55 210-218.

Sándor, Z., M. Wedel. 2002. Profile construction in experimental choice designs for mixed logit models. *Marketing Science* 21 (Fall) 455-475.

Sándor, Z., P. P. András. 2004. Alternative sampling methods for estimating multivariate normal probabilities. *Journal of Econometrics* 120 207-234.

Sebastiani, P., G. Parmigiani. 1997. A note on d-optimal designs for a logistic regression models. *Journal of Statistical Planning and Inference* 59 359-368.

Sebastiani, P., G. Parmigiani. 1998. First-order optimal deisgns for non-linear models. *Journal of Statistical Planning and Inference* 74 177-192.

Sebastiani, P., H. P. Wynn. 1997. Bayesian experimental design and Shannon information. 1997. Proceedings of the Section on Bayesian Statistical Science. 176–181.

Shannon, C. E. 1948. A mathematical theory of communication. *Bell System Technical Journal* 27 379-423, 623-656.

Stroud, J. R., P. Müller, G. L. Rosner. 2001. Optimal sampling times in population pharmacokinetic studies. *Applied Statistics* 50 345-359.

Sun, D., R. K. Tsutakawa, W. S. Lu. 1996. Bayesian design of experiment for quantal responses: what is promised versus what is delivered. *Journal of Statistical Planning and Inference* 52 289-306.

Sun, D., R. K. Tsutakawa. 1997. Bayesian design for dose-response curves wit penalized risk. *Biometrics* 53 1262-1273.

Tsutakawa, R. K. 1972. Design of an experiment for bioassay. *Journal of the American Statistical Association* 67 584-590

Tsutakawa, R. K. 1980. Selection of dose levels for estimating a percentage point of a logistic quantal response curve. *Applied Statistics* 29 25-33.

Vermeulen, B., P. Goos, M. Vandebroek. 2008. Models and optimal designs for conjoint choice experiments including a no-choice option. *International Journal of Research in Marketing* 25 94-103.

Woodworth, G. G. 1976. t for two, or preposterior analysis for two decision makers: interval estimates for the mean. *The American Statistician* 30 168-171.

Yu, J., P. Goos, M. Vandebroek. 2008a. Efficient conjoint choice designs in the presence of respondent heterogeneity. To appear in *Marketing Science*.

Yu, J., P. Goos, M. Vandebroek. 2008b. Comparing different approaches for approximating the integrals involved in the Bayesian optimal design of choice experiments. Technical Report.

Zwerina, K., J. Huber, W. F. Kuhfeld. 1996. A general method for constructing efficient choice designs. Working paper, Fuqua School of Business, Duke University, Durham, NC 27708.