This item is the archived peer-reviewed author-version of:

Quadrature methods for bayesian optimal design of experiments with nonnormal prior distributions

King's Research Portal

Link to publication record in King's Research Portal

Quadrature Methods for Bayesian Optimal Design of Experiments with Non-Normal Prior Distributions

Peter Goos

Faculty of Bioscience Engineering and Leuven Statistics Research Centre

KU Leuven

Faculty of Applied Economics and StatUa Center for Statistics

Universiteit Antwerpen

Kalliopi Mylona

Department of Statistics

Universidad Carlos III de Madrid

Department of Mathematics

King's College London

Southampton Statistical Sciences Research Institute

University of Southampton

December 12, 2016

Abstract

Many optimal experimental designs depend on one or more unknown model parameters. In such cases, it is common to use Bayesian optimal design procedures to seek designs that perform well over an entire prior distribution of the unknown model parameter(s). Generally, Bayesian optimal design procedures are viewed as computationally intensive. This is because they require numerical integration techniques in order to approximate the Bayesian optimality criterion at hand. The most common numerical integration technique involves pseudo Monte Carlo draws from the prior distribution(s). For a good approximation of the Bayesian optimality criterion, a large number of pseudo Monte Carlo draws is required. This results in long computation times. As an alternative to the pseudo Monte Carlo approach, we propose using computationally efficient Gaussian quadrature techniques. Since, for normal prior distributions, suitable quadrature techniques have already been used in the context of optimal experimental design, we focus on quadrature techniques for non-normal prior distributions. Such prior distributions are appropriate for variance components, correlation coefficients and any other parameters that are strictly positive or have

1

upper and lower bounds. In this article, we demonstrate the added value of the quadrature techniques we advocate by means of the Bayesian D-optimality criterion in the context of split-plot experiments, but we want to stress that the techniques can be applied to other optimality criteria and other types of experimental designs as well.

1 Introduction

In recent years, optimal design of experiments has gained substantial popularity due to the availability of fast algorithms for computing optimal experimental designs in user-friendly software. Generally, optimal experimental designs depend on one or more unknown parameters, the only exception being designs for completely randomized experiments and linear regression models (see, for instance, Atkinson et al. (2007) and Goos and Jones (2011)).

When the optimal experimental design depends on one or more unknown parameters, one option is to use the Bayesian optimal design approach, which involves a prior distribution on the unknown parameters, expressing the researcher's prior knowledge concerning the unknown parameter(s). A Bayesian optimal design is a design that maximizes the average performance across the entire prior distribution. The Bayesian optimal experimental design approach is superior to the locally optimal design approach, in which optimal designs are computed assuming a single educated guess for the unknown parameters. The added value of the Bayesian approach in optimal experimental design has been demonstrated convincingly by Woods et al. (2006) for generalized linear models in an industrial context and by Sándor and Wedel (2001) for binary choice models in a marketing context. In the context of the optimal design of biopharmaceutical experiments, Fedorov and Leonov (2014) mention the main advantage of Bayesian optimal designs, namely that they take into account prior uncertainty about the unknown parameters. However, they also indicate that this leads to computationally demanding optimization problems. The computational burden is also what prevents Diggle and Lophaven (2005) from using a Bayesian optimal design approach in geostatistical design. The Bayesian optimal design approach is not the only possible one: several authors have used a maximin optimal design approach, in which the worst case performance of an experimental design is optimized (see, for example, Dette (1997), Imhof (2001) and Dette et al. (2007)). The maximin approach requires the researcher only to specify sensible ranges for the unknown parameters on which the experimental design depends. Unlike the Bayesian approach, the maximin approach does not take into account any existing prior knowledge concerning the likelihood of various parameters values. A proper implementation of the maximin approach for realistic problems is oftentimes also

3

computationally expensive. In this article, we focus on the Bayesian approach, because it makes use of all available prior knowledge concerning the model parameters.

For the Bayesian approach to be feasible, computationally efficient implementations are required. For the case where the prior distribution for the model parameters is the multivariate normal, substantial work has been done to limit the computational time of Bayesian optimal designs through the use of Gaussian quadrature techniques (Bliemer et al.; 2009; Gotwalt et al.; 2009; Gotwalt; 2010; Yu et al.; 2010). In many cases, however, the multivariate normal is not an appropriate prior distribution. It is quite common to encounter model parameters which are necessarily positive (for instance, variances of random effects or ratios of variance components in mixed models), or which are bounded by -1 and 1 or by 0 and 1 (for instance, serial correlation coefficients or intra-block correlation coefficients). The optimal design of experiments in the presence of serial or spatial correlation has been studied Martin, Eccleston and Jones (1998), Martin, Jones and Eccleston (1998), Elliott et al. (1999), Garroi et al. (2009), Sethuraman and Raghavarao (2009), and Androulakis et al. (2013). In the presence of serial or spatial correlation, the designs depend on the magnitudes of correlation coefficients and on the parameters driving the spatial correlation. The optimal design of blocked experiments involving random block effects has been visited by Jones (1986) and Goos and Vandebroek (2001a). In that case, the optimal designs depend on the relative magnitude of the block effects' variance and the residual error variance. The optimal design of split-plot experiments, where designs depend on the relative magnitude of the whole-plot errors' variance and the residual error variance, has been investigated by Goos and Vandebroek (2001b, 2003, 2004), Jones and Goos (2007), Macharia and Goos (2010), Schoen et al. (2011), Mylona et al. (2014), Sambo et al. (2014) and Borrotti et al. (2016). Harari and Steinberg (2014) discuss the generation of optimal designs for Gaussian process models, and conclude by saying that it would be useful to develop a Bayesian approach that incorporates prior information on unknown correlation parameters. Atkinson et al. (2007) discuss the dependence of optimal experimental designs for estimating an exponential decay model and a compartmental model on one or more positive parameters. In all these cases, log-normal, gamma or beta prior distributions are required because the domain of these distributions matches that of variance compo-

nents and correlation coefficients. In this article, we provide an overview of the relatively unknown Gaussian quadrature approaches that are appropriate in these scenarios.

We view this paper as a tutorial on quadrature approaches that should be accessible to non-specialists in numerical analysis. For this reason, we provide much detail on each of the quadrature techniques to maximize understanding by non-specialists and to increase the likelihood that they will be adopted in future research on optimal experimental design. We also provide Matlab programs to generate the abscissas and weights required for the quadrature techniques in the supplementary materials. In the next section, we describe the design problem we utilize to demonstrate the usefulness of each of the quadrature techniques. Next, we discuss pseudo Monte Carlo sampling, which is the most widely used technique to evaluate Bayesian optimality criteria and we describe the basic principles of Gaussian quadrature. In Section 4, we explain how to deal with a log-normal prior distribution. In Section 5, we deal with a gamma prior distribution. In Sections 6 and 7, we shift our attention to beta and beta prime prior distributions, respectively. Finally, in Section 8, we discuss how to deal with multidimensional integrals in situations where the optimal experimental design depends on multiple model parameters.

2   Model and Design Criterion

For this tutorial paper on quadrature techniques, as a proof of concept, we selected one specific optimal experimental design problem to work with, namely the optimal design of split-plot experiments. A point-exchange design construction algorithm for this problem was presented in Goos and Vandebroek (2003), whereas a coordinate-exchange algorithm, similar to the one in Jones and Goos (2007, 2012), has been implemented in JMP and Design-Expert. Trinca and Gilmour (2001, 2015) present stratum-by-stratum algorithms to create split-plot and other multi-stratum designs. In this section, we discuss the model on which optimal split-plot experimental designs are based, as well as the D-optimality criterion which we selected to demonstrate the working of the quadrature techniques.

2.1 Model and estimation

The model for data from split-plot experiments is given by

$$\boldsymbol{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \tag{1}$$

where $\boldsymbol{Y}$ is an n-dimensional vector of responses, $\mathbf{X}$ is an $n \times p$ model matrix, $\mathbf{Z}$ is an $n \times b$ matrix of zeros and ones whose (i, j)th element is one when the ith observation is obtained in whole plot j and zero otherwise, $\boldsymbol{\gamma}$ is a $b \times 1$ vector of random effects describing the variation in the responses from one whole plot to another, and $\boldsymbol{\varepsilon}$ is an $n \times 1$ vector containing the random errors for each of the n measured responses. The elements of $\boldsymbol{\gamma}$ and $\boldsymbol{\varepsilon}$ are assumed to be mutually independently normally distributed random variables with zero mean and variances σ_γ^2 and σ_ε^2 respectively. The implied variance-covariance matrix for the response vector $\boldsymbol{Y}$ then is

$$\mathbf{V} = \sigma_\varepsilon^2 \mathbf{I}_n + \sigma_\gamma^2 \mathbf{Z}\mathbf{Z}' = \sigma_\varepsilon^2 (\mathbf{I}_n + \eta \mathbf{Z}\mathbf{Z}') = \sigma_\varepsilon^2 (\mathbf{I}_n + \frac{\rho}{1 - \rho} \mathbf{Z}\mathbf{Z}'),$$

where $\eta = \sigma_\gamma^2 / \sigma_\varepsilon^2$ and $\rho = \sigma_\gamma^2 / (\sigma_\gamma^2 + \sigma_\varepsilon^2)$. The larger the variance ratio η or the intra-whole-plot correlation coefficient ρ, the stronger observations within the same whole plot are correlated.

The best linear unbiased estimator for the parameter vector $\boldsymbol{\beta}$ is the generalized least squares (GLS) estimator

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\boldsymbol{Y}. \tag{2}$$

The corresponding information matrix equals

$$\mathbf{M} = \mathbf{X}'\mathbf{V}^{-1}\mathbf{X}. \tag{3}$$

2.2 Design criterion

When planning an experiment, the criterion used to select the experimental design should reflect the researcher's goals. Atkinson et al. (2007) list the most commonly used criteria. Some of these criteria are estimation-oriented, while others are prediction-oriented. The A-, D- and E-optimality criteria seek designs that maximize the precision of the parameter estimates, while the c- and L-optimality criteria seek designs that maximize the

precision of one or more linear combinations of estimates. Obviously, all these criteria are estimation-oriented. The V-optimality criterion and the G-optimality criterion seek designs that minimize the average variance of prediction and the maximum variance of prediction, respectively. In this paper, we use the D-optimality criterion to demonstrate Gaussian quadrature, because this is the best known optimality criterion.

2.2.1 D-optimality

A D-optimal design minimizes the volume of the confidence ellipsoid about the model parameters. Equivalently, a D-optimal design maximizes the determinant of the information matrix in (3). A D-optimal design thus maximizes

$$|\mathbf{M}| = |\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}|. \tag{4}$$

In general, D-optimal designs for split-plot experiments depend on the relative magnitude of σ_γ^2 and σ_ε^2, as measured by the variance ratio η and the correlation coefficient ρ, through $\mathbf{V}$. Goos (2002) argues that, when the number of whole plots and the whole-plot sizes are fixed, the dependence is minor. However, as shown in Tables 8.3 and 8.4 of the book, when the number of whole plots and the whole-plot sizes are not fixed, the D-optimal designs can be very different from one value of ρ or η to the other. This strong dependence of D-optimal split-plot designs on the unknown ρ or η value is a strong argument in favor of the Bayesian approach in the context of split-plot designs. Note that the absolute value of σ_γ^2 and σ_ε^2 has no impact on the optimal design. Therefore, without loss of generality, we use $\sigma_\varepsilon^2 = 1$ in this article, whenever reporting D-optimality criterion values.

2.2.2 Bayesian D-optimality

To cope with the prior uncertainty about η or ρ, we adopt the Bayesian approach introduced by Chaloner and Larntz (1989) in the context of single-factor logistic regression, and applied by Woods et al. (2006), Gotwalt et al. (2009), Bliemer et al. (2009) and Kessels et al. (2011), among others, in various multi-factor contexts.

The Bayesian D-optimality criterion for selecting split-plot designs is

$$D_B = \int_0^{+\infty} \Phi(\eta) \cdot \pi_1(\eta) \cdot d\eta \tag{5}$$

in situations where a prior distribution $\pi_1(\eta)$ is specified for the variance ratio η, and

$$D_B = \int_0^1 \Phi(\rho) \cdot \pi_2(\rho) \cdot d\rho \qquad (6)$$

in situations where a prior distribution $\pi_2(\rho)$ is specified for the intra-whole-plot correlation coefficient ρ, where

$$\Phi = \log|\mathbf{M}| = \log|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}|. \qquad (7)$$

To stress the dependence of the D-optimality criterion Φ on η or ρ, we denote it by $\Phi(\eta)$ or $\Phi(\rho)$ (depending on the exact parametrization) in the remainder of this article.

Some researchers might consider specifying a joint prior distribution for the two variance components σ_γ^2 and σ_ε^2. In this paper, we only discuss the special case of two independent prior distributions $\pi_a(\sigma_\gamma^2)$ and $\pi_b(\sigma_\varepsilon^2)$ for σ_γ^2 and σ_ε^2, in which case the Bayesian D-optimality criterion becomes

$$D_B = \int_0^{+\infty} \int_0^{+\infty} \Phi(\sigma_\gamma^2, \sigma_\varepsilon^2) \cdot \pi_a(\sigma_\gamma^2) \cdot \pi_b(\sigma_\varepsilon^2) \cdot d\sigma_\gamma^2 \cdot d\sigma_\varepsilon^2. \qquad (8)$$

3   Evaluating the Bayesian D-optimality Criteria

3.1   Pseudo Monte Carlo sampling

A technical problem with the Bayesian D-optimality criterion in (5), (6) and (8) is that there is no analytical solution for the integral over the prior distributions. The Bayesian criterion therefore has to be approximated numerically. This can be done by taking N pseudo Monte Carlo draws from the prior distributions, computing the value of Φ for each of these N draws, and averaging the resulting values. Representing the ith pseudo random draw from prior distribution $\pi_1(\eta)$ or $\pi_2(\rho)$ by d_i, the Bayesian D-optimality criteria in (5) and (6) are thus approximated by the (unweighted) average

$$D_B \approx \frac{1}{N} \sum_{i=1}^N \Phi(d_i). \qquad (9)$$

When the Bayesian D-optimality criterion in (8) is used, pseudo random draws d_{ai} and d_{bi} are needed from both the prior distributions $\pi_a(\sigma_\gamma^2)$ and $\pi_b(\sigma_\varepsilon^2)$, and the criterion is

approximated by

$$D_B \approx \frac{1}{N_a N_b} \sum_{i=1}^{N_a} \sum_{j=1}^{N_b} \Phi(d_{ai}, d_{bj}), \tag{10}$$

where N_a and N_b are the numbers of draws from the two prior distributions.

For this approach to work well, large numbers of pseudo Monte Carlo draws are needed. For each of the draws, the design optimality criterion in (7) has to be evaluated. This is computationally expensive, especially because point- and coordinate-exchange algorithms for constructing optimal experimental designs involve many evaluations of the optimality criterion. Hence, Bayesian optimal experimental design algorithms based on pseudo Monte Carlo samples are prohibitively slow. There exist, however, several alternatives to pseudo Monte Carlo sampling, an important one of which is Gaussian quadrature.

3.2   Gaussian quadrature

There are two important differences between pseudo Monte Carlo sampling and Gaussian quadrature techniques for numerical integration. The first difference is that Gaussian quadrature uses systematic, deterministic draws d_i from the prior distribution (which is called weight function in the specialized literature). In this paper, we call the number of systematic draws R. The second important difference is that, in Gaussian quadrature, not every draw receives the same weight. Instead of approximating the integral of interest using an unweighted average of function evaluations, this is done using a weighted average. With Gaussian quadrature, the Bayesian D-optimality criteria in (5) and (6) are approximated by a weighted sum

$$D_B \approx \sum_{i=1}^{R} w_i \Phi(d_i), \tag{11}$$

where w_i represents the weight of the ith systematic draw d_i. The Bayesian D-optimality criterion in (8) is approximated by

$$D_B \approx \sum_{i=1}^{R_a} \sum_{j=1}^{R_b} w_{ai} w_{bj} \Phi(d_{ai}, d_{bj}), \tag{12}$$

where w_{ai} and w_{bj} are the weights of the systematic draws d_{ai} and d_{bj} from the prior distributions of σ_γ^2 and σ_ε^2, and R_a and R_b are the numbers of draws from the two prior distributions.

9

A key feature of Gaussian quadrature is that, by using R systematic draws, any polynomial of degree $2R - 1$ is integrated exactly. Since any k-times differentiable function can be approximated by means of a polynomial of degree k (due to Taylor's theorem), this suggests it is possible to integrate any function well, even with just a relatively small number of systematic draws. Expressions for the errors made when using Gaussian quadrature to approximate integrals can be found in standard textbooks on numerical analysis (see, for instance, Chapter 3 in Kythe and Schäferkotter (2005) and Theorem 3.6.24 in Stoer and Bulirsch (2002)). The hope is that these errors vanish when the number of systematic draws is sufficiently large. It is important to check that this is indeed the case, in any particular kind of application.

In this paper, we show that Gaussian quadrature is a very useful technique in the context of optimal experimental design, where optimality criteria are differentiable functions of one or more unknown parameters and need to be integrated numerically over a prior distribution to identify Bayesian optimal designs. To investigate whether the errors of the various quadrature approaches drop to zero as R is increased, we compare the results of Gaussian quadrature to the value obtained by utilizing one million pseudo Monte Carlo draws. Since there is no analytical solution to the integral in the formulation of the Bayesian optimality criterion, the true value of the Bayesian D-optimality criterion is unknown. We therefore assume that the value we obtain from one million pseudo Monte Carlo draws represents the true value of the Bayesian D-optimality criterion.

One challenge in Gaussian quadrature is to produce the systematic draws and the corresponding weights to obtain a high-quality approximation of the Bayesian D-optimality criterion. The way in which the systematic draws are determined depends on the prior distribution(s) used. In the numerical integration jargon, a different quadrature rule is required for each weight function. In case the prior distribution is normal, the Gauss-Hermite quadrature rule can be used. The systematic draws and their weights are then obtained from a special type of orthogonal polynomials, called Hermite polynomials. When the prior distribution is log-normal, Gauss-Hermite quadrature can also be employed, but Wilck (2001) recommended Gauss-Stieltjes-Wigert quadrature instead, based on Stieltjes-Wigert polynomials, another type of orthogonal polynomials. Likewise, Gauss-Jacobi quadrature,

which is suited for beta prior distributions, is based on Jacobi polynomials, and Gauss-Laguerre quadrature, which is suited for gamma prior distributions, is based on Laguerre polynomials. Finally, Gauss-Legendre quadrature is based on Legendre polynomials and can be used for uniform prior distributions.

Generally, the systematic draws d_i are obtained from the roots of the orthogonal polynomials using a simple transformation. The roots of the polynomial are referred to as abscissas. We denote the abscissas by a_i in this paper.

In this article, we demonstrate the usefulness of Gaussian quadrature for log-normal, gamma, beta and beta prime prior distributions. Because we view this paper as a tutorial, we provide some of the mathematical background as well and illustrate how the abscissas a_i are converted into systematic draws d_i from each of the prior distributions considered. We do not discuss normal prior distributions in this paper, because the use of Gaussian quadrature for normal prior distributions was discussed in detail in Bliemer et al. (2009), Yu et al. (2010) and Gotwalt et al. (2009). Instead, we discuss several lesser-known quadrature techniques for useful prior distributions other than the normal.

In the next two sections, we discuss quadrature techniques when a log-normal or gamma prior distribution is assumed for the variance ratio η. Next, in Section 6, we discuss the appropriate quadrature technique when a beta prior distribution is assumed for the correlation coefficient ρ. In Section 7, we discuss the use of independent prior distributions for σ_γ^2 and σ_ε^2, and explain how to cope with a beta prime prior distribution. Finally, in Section 8, we sketch various possible approaches to handle more general scenarios involving multidimensional integrals.

4   Log-normal prior distribution for η

Assuming the variance ratio η can only take positive values, one possible prior distribution $\pi_1(\eta)$ for η is the log-normal distribution

$$\pi_1(\eta) = \frac{1}{\eta\sigma\sqrt{2\pi}} \cdot e^{-\{\log(\eta)-\mu\}^2/(2\sigma^2)},$$

where $\eta \geq 0$. The Bayesian design criterion in (5) then becomes

$$D_B = \int_0^{+\infty} \Phi(\eta) \cdot \frac{1}{\eta\sigma\sqrt{2\pi}} \cdot e^{-\{\log(\eta)-\mu\}^2/(2\sigma^2)} \cdot d\eta. \tag{13}$$

To evaluate this integral numerically, we can use Gauss-Hermite quadrature, which has been utilized by Bliemer et al. (2009) and Yu et al. (2010) in the context of the optimal design of choice experiments. An alternative, more direct and computationally cheaper way, suggested by Wilck (2001), involves Gaussian quadrature based on Stieltjes-Wigert polynomials.

4.1 Gauss-Hermite quadrature

Gauss-Hermite quadrature approximates integrals of the form

$$\int_{-\infty}^{\infty} f(x) \cdot e^{-x^2} \cdot dx \tag{14}$$

by the weighted sum

$$\sum_{i=1}^{R} w_i^{\mathrm{GH}} \cdot f(a_i^{\mathrm{GH}}),$$

where the R abscissas a_i^{GH} and the corresponding weights w_i^{GH} are obtained from the Hermite polynomials (see, for instance, Rabinowitz (2001)).

In order to use Gauss-Hermite quadrature when a log-normal prior distribution is utilized for the variance ratio η, η has to be substituted by e^y in (13), so that

$$D_B = \int_{-\infty}^{+\infty} \Phi(e^y) \cdot \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-(y-\mu)^2/(2\sigma^2)} \cdot dy. \tag{15}$$

Replacing y by $\mu + z\sigma\sqrt{2}$ then yields

$$D_B = \frac{1}{\sqrt{\pi}} \cdot \int_{-\infty}^{+\infty} \Phi(e^{\mu+z\sigma\sqrt{2}}) \cdot e^{-z^2} \cdot dz. \tag{16}$$

The integral in this expression has the same form as that in (14). It can therefore be approximated using Gauss-Hermite quadrature:

$$\begin{aligned} D_B &\approx \frac{1}{\sqrt{\pi}} \cdot \sum_{i=1}^{R} w_i^{\mathrm{GH}} \cdot \Phi(e^{\mu+a_i^{\mathrm{GH}}\sigma\sqrt{2}}), \\ &\approx \frac{1}{\sqrt{\pi}} \cdot \sum_{i=1}^{R} w_i^{\mathrm{GH}} \cdot \Phi(d_i^{\mathrm{GH}}), \\ &\approx \sum_{i=1}^{R} \widetilde{w}_i^{\mathrm{GH}} \cdot \Phi(d_i^{\mathrm{GH}}), \end{aligned} \tag{17}$$

12

where

$$d_i^{\mathrm{GH}} = e^{\mu + a_i^{\mathrm{GH}} \sigma \sqrt{2}}$$

is the ith systematic draw from the log-normal prior distribution with parameters μ and σ, $\widetilde{w}_i^{\mathrm{GH}} = w_i^{\mathrm{GH}}/\sqrt{\pi}$ is the corresponding normalized weight, and the R abscissas a_i^{GH} and (non-normalized) weights w_i^{GH} are obtained from the Hermite polynomial of order R. We call the weights $\widetilde{w}_i^{\mathrm{GH}}$ normalized because they sum to one for each value of R. The weights w_i^{GH} produced from the Hermite polynomial sum to $\sqrt{\pi}$.

Table 1 shows the abscissas a_i^{GH} and weights w_i^{GH} and $\widetilde{w}_i^{\mathrm{GH}}$ for R up to 5, as well as the corresponding systematic draws d_i^{GH} for two choices of the log-normal prior distribution's parameters μ and σ. The table shows that the abscissas are symmetric about zero, and illustrates the conversion of the abscissas into systematic draws from the exact prior distribution utilized. When $\mu = 0$ and $\sigma = 0.5$, the prior distribution attaches much importance to small variance ratios η. Table 1 shows that, in that case, Gauss-Hermite quadrature results in systematic draws that do not go beyond 4.1724. When $\mu = 3$ and $\sigma = 1$, the prior distribution attaches substantial importance to big values of η. The Gauss-Hermite approach then results in much larger systematic draws.

Figure 1 compares Gauss-Hermite quadrature involving $R = 16$ abscissas with pseudo Monte Carlo sampling with $N = 1000$, $N = 10000$, $N = 10^5$ and $N = 10^6$ for a five-factor split-plot response surface design involving 21 whole plots of two runs, using a log-normal prior distribution with $\mu = 0$ and $\sigma = 0.75$. The design we evaluate in the figure using the Bayesian D-optimality criterion is the design proposed in Trinca and Gilmour (2001) for a protein extraction experiment. We assume that a second-order response surface model is used, involving an intercept, five main effects, ten two-factor interactions and five quadratic effects.

For each number of pseudo Monte Carlo draws, N, we approximated the Bayesian D-optimality criterion ten times and made box plots of the resulting approximations. We also approximated the Bayesian D-optimality criterion by means of $R = 16$ systematic draws produced by the Gauss-Hermite quadrature technique. The resulting approximate value of the optimality criterion was 49.1833, which is depicted in Figure 1 by means of a horizontal reference line. Pseudo Monte Carlo sampling utilizing $N = 1000$ draws results

13
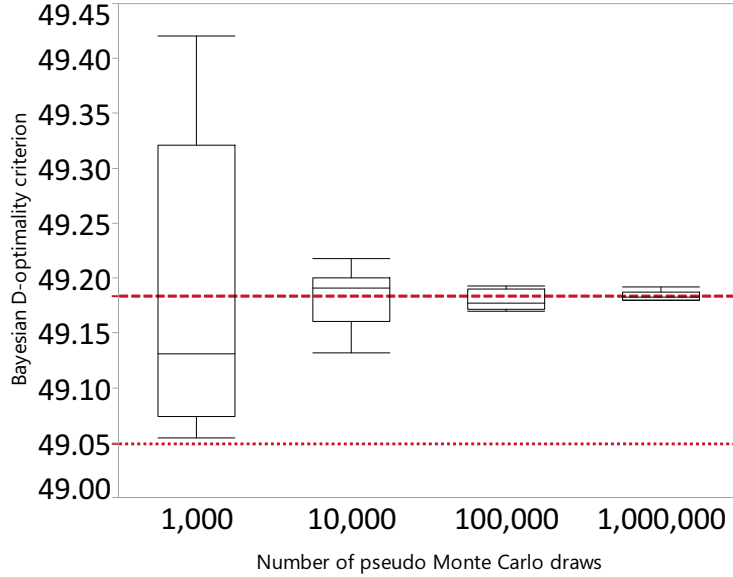
Figure 1: Box plots comparing the pseudo Monte Carlo approach with the Gauss-Hermite and Gauss-Stieltjes-Wigert approaches for a log-normal prior distribution with $\mu = 0$ and $\sigma = 0.75$. The dashed reference line indicates the result from the Gauss-Hermite approach using $R = 16$, while the dotted line indicates the result from the Gauss-Stieltjes-Wigert approach.

in substantial variability in the approximations of the optimality criterion. To reduce that variability substantially requires many more pseudo random draws.

Since there is no analytical solution to the integral in the formulation of the Bayesian optimality criterion, the true value of the Bayesian D-optimality criterion is unknown. We therefore assumed that the value we obtain from one million pseudo Monte Carlo draws represents the true value of the Bayesian criterion. That value equals 49.1839. We can see that using 1,000 pseudo Monte Carlo draws results in poor approximations of this value: the approximations based on 1,000 draws fluctuate between 49.0542 and 49.420. For 10,000 pseudo Monte Carlo draws, the approximation is substantially better, with a minimum value of 49.1319 and a maximum value of 49.2175. Using one million random draws results in values between 49.1697 and 49.1931, so that this large a number of draws guarantees a good approximation of the Bayesian D-optimality criterion.

However, the dashed reference line in Figure 1 shows that Gauss-Hermite quadrature,

14

with just $R = 16$ systematic draws and hence only 16 evaluations of $\Phi(\eta)$, produces an excellent approximation of the true Bayesian D-optimality criterion value. It even turns out that using only $R = 5$ systematic draws would be as good as using $R = 16$. This is shown in Table 2, where we evaluated the Bayesian D-optimality criterion with R values ranging from 1 to 16 when $\mu = 0$ and $\sigma = 0.75$. Starting from $R = 5$, the approximation of the Bayesian D-optimality criterion does not change any more, suggesting that as few as five systematic draws suffices here.

As a result, the use of Gauss-Hermite quadrature allows an accurate evaluation of the Bayesian D-optimality criterion at a computational cost that is a factor of 200,000 smaller than pseudo Monte Carlo sampling.

4.2   Gauss-Stieltjes-Wigert quadrature

An alternative way to approximate the Bayesian D-optimality criterion in (13) is to use Gauss-Stieltjes-Wigert quadrature. The most well-known form of this type of quadrature is suitable for integrals of the form

$$\int_0^\infty f(x) \cdot \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-(\log x)^2/(2\sigma^2)} \cdot dx,$$

but this approach can be generalized to cope with integrals of the form

$$\int_0^\infty f(x) \cdot \frac{1}{x\sigma\sqrt{2\pi}} \cdot e^{-(\log x)^2/(2\sigma^2)} \cdot dx. \tag{18}$$

In order to apply this type of quadrature to our problem, the parameter μ in (13) can be replaced by $\log(\lambda)$, where $\lambda > 0$, so that

$$
\begin{aligned}
D_B &= \int_0^{+\infty} \Phi(\eta) \cdot \frac{1}{\eta\sigma\sqrt{2\pi}} \cdot e^{-\{\log(\eta)-\log(\lambda)\}^2/(2\sigma^2)} \cdot d\eta, \\
&= \int_0^{+\infty} \Phi(\eta) \cdot \frac{1}{\eta\sigma\sqrt{2\pi}} \cdot e^{-\{\log(\eta/\lambda)\}^2/(2\sigma^2)} \cdot d\eta.
\end{aligned}
\tag{19}
$$

Substituting y for η/λ then yields

$$
\begin{aligned}
D_B &= \int_0^{+\infty} \Phi(\lambda y) \cdot \frac{1}{y\sigma\sqrt{2\pi}} \cdot e^{-(\log y)^2/(2\sigma^2)} \cdot dy, \\
&= \int_0^{+\infty} \Phi(e^\mu y) \cdot \frac{1}{y\sigma\sqrt{2\pi}} \cdot e^{-(\log y)^2/(2\sigma^2)} \cdot dy,
\end{aligned}
\tag{20}
$$

which has the same form as (18). As a result, the Bayesian D-optimality criterion in (13) can be approximated as

$$
\begin{aligned}
D_B &\approx \sum_{i=1}^{R} w_i^{\mathrm{GSW}} \cdot \Phi(e^\mu a_i^{\mathrm{GSW}}), \\
&\approx \sum_{i=1}^{R} w_i^{\mathrm{GSW}} \cdot \Phi(d_i^{\mathrm{GSW}}),
\end{aligned}
\tag{21}
$$

where

$$
d_i^{\mathrm{GSW}} = e^\mu a_i^{\mathrm{GSW}}
$$

is the ith systematic draw from the log-normal prior distribution with parameters μ and σ, and a_i^{GSW} and w_i^{GSW} represent the abscissas and weights obtained from the Stieltjes-Wigert orthogonal polynomials.

Compared to Hermite polynomials and to the Jacobi and Laguerre polynomials, which we will encounter later in this paper, the Stieltjes-Wigert polynomials are relatively unknown. The most convenient way to determine the abscissas and their weights is by computing the eigenvalues and the eigenvectors of the tridiagonal, symmetric Jacobi matrix

$$
\mathbf{J}_R = \begin{bmatrix}
a_0 & b_1 & 0 & 0 & \dots & 0 \\
b_1 & a_1 & b_2 & 0 & \dots & 0 \\
0 & b_2 & a_2 & b_3 & \dots & 0 \\
0 & 0 & \ddots & \ddots & \ddots & \vdots \\
\vdots & \vdots & \ddots & b_{R-2} & a_{R-2} & b_{R-1} \\
0 & 0 & \dots & 0 & b_{R-1} & a_{R-1}
\end{bmatrix},
$$

where

$$
a_0 = e^{\sigma^2/2},
$$

$$
a_i = ((a_0^2 + 1)a_0^{2i} - 1)a_0^{2i-1},
$$

and

$$
b_i = \sqrt{(a_0^{2i} - 1)a_0^{6i-4}},
$$

for $i = 1, \dots, R-1$. The ith abscissa a_i^{GSW} is the ith eigenvalue of $\mathbf{J}_R$, and the corresponding weight w_i^{GSW} is the square of the first component of the ith normalized eigenvector of $\mathbf{J}_R$.

This approach to computing the abscissas and weights for Gaussian quadrature is outlined in Golub and Welsch (1969), Gautschi (2004) and Mastroianni and Milovanovic (2008).

A major difference between the abscissas a_i^{GSW} for Gauss-Stieltjes-Wigert quadrature and the abscissas a_i^{GH} for Gauss-Hermite quadrature is that, through the values a_i and b_i in the Jacobi matrix $\mathbf{J}_R$, the former depend on the parameter σ of the log-normal prior distribution, whereas the latter are independent of the parameters of the prior distribution. The weights w_i^{GSW} produced by Gauss-Stieltjes-Wigert quadrature are automatically normalized.

Table 3 shows the abscissas a_i^{GSW}, systematic draws d_i^{GSW} and weights w_i^{GSW} for R up to 5, when $\mu = 0$ and $\sigma = 0.5$ and when $\mu = 3$ and $\sigma = 1$. Comparing the systematic draws produced by the Gauss-Stieltjes-Wigert approach to those produced by the Gauss-Hermite approach (see Table 1), we can see that the systematic draws produced by the former approach are much larger than those produced by the latter. Also, some of the draws have weights that are so small that they become irrelevant. Such extremely small weights do not occur with Gauss-Hermite quadrature for $R \leq 5$. It is conceivable that the large systematic draws and the small weights for the Gauss-Stieltjes-Wigert approach will lead to numerical problems in some instances.

The second column of Table 2 shows the results of evaluating the Bayesian D-optimality criterion for the split-plot design of Trinca and Gilmour (2001) by means of Gauss-Stieltjes-Wigert quadrature with R values ranging from 1 to 16. The value produced by Gauss-Stieltjes-Wigert quadrature creeps up very slowly to stabilize at 49.0488 when $R = 16$. Gauss-Stieltjes-Wigert quadrature thus leads to a different value when $R = 16$ than Gauss-Hermite quadrature (which produced 49.1833). The value 49.0488 is also depicted in Figure 1 by means of the dotted line. The picture shows that Gauss-Stieltjes-Wigert quadrature underestimates the true value of the Bayesian D-optimality criterion. We observed the same kind of underestimation by the Gauss-Stieltjes-Wigert approach for other designs. We believe that it is due to the large values for the modified abscissas d_i^{GSW}. These values require evaluating the determinant of the information matrices for very large variance ratios η, in which case the whole-plot-to-whole-plot variation goes to infinity and the information matrix for a split-plot design becomes singular. This problem does not occur with

Gauss-Hermite quadrature, since the d_i^{GH} values are much smaller than the d_i^{GSW} values. The underestimation by the Gauss-Stieltjes-Wigert approach disappears when σ is lowered to 0.5, but such a σ value may not be in line with the prior beliefs about the variance ratio η.

4.3   Comparison with pseudo Monte Carlo sampling

Poor approximations of the Bayesian D-optimality criterion are, of course, not desirable. However, what matters most is that the approximation is good enough to rank sets of alternative designs correctly. Kessels et al. (2009), for instance, state that, in the computation of Bayesian optimal designs, the approximation of the Bayesian optimality criterion should only be accurate enough to judge whether one design is better than another.

To investigate this issue, we generated ten high-quality alternative designs for the protein extraction experiment. To this end, we used one try of the coordinate-exchange algorithm by Jones and Goos (2007), as embedded in the JMP statistics software package, ten times. We used the default settings of the algorithm and specified a full second-order model. For each of the ten resulting designs, we evaluated the performance in terms of the Bayesian D-optimality criterion, assuming a log-normal prior distribution with $\mu = 0$ and $\sigma = 0.75$. To evaluate the performance, we used pseudo Monte Carlo sampling with 1,000,000 draws, and Gauss-Hermite and Gauss-Stieltjes-Wigert quadrature with R values of 1, 2, 4 and 8.

For each of the ten designs we created, Figure 2 compares the Bayesian D-optimality criterion evaluated using Gauss-Stieltjes-Wigert quadrature with $R = 1$, $R = 2$, $R = 4$ and $R = 8$ abscissas and using a sample of one million pseudo Monte Carlo draws. The figure first of all confirms that the Gauss-Stieltjes-Wigert approach systematically underestimates the true value of the Bayesian criterion. Figure 2a shows that Gauss-Stieltjes-Wigert quadrature using $R = 1$ leads to a quite different ranking of the ten alternative designs than pseudo Monte Carlo sampling with 1,000,000 draws. In the ranking according to Gauss-Stieltjes-Wigert quadrature with $R = 1$, design 8 is better than designs 3 and 7, and design 9 is better than design 1, while it is the other way around when considering the 1,000,000 pseudo Monte Carlo samples. Using a larger number of systematic samples remediates this

Figure 2: The Bayesian D-optimality criterion for ten designs evaluated using Gauss-Stieltjes-Wigert quadrature (vertical axis) and 1,000,000 pseudo Monte Carlo draws.

problem. Figures 2b, 2c and 2d shows that the ranking obtained using $R = 2$, $R = 4$ and $R = 8$ is the same as for 1,000,000 pseudo Monte Carlo samples. However, as Figure 2b shows, this is only barely the case for $R = 2$. Therefore, we recommend going beyond two systematic draws.

Figure 3 compares the Bayesian D-optimality criterion evaluated using Gauss-Hermite quadrature with $R = 1$, $R = 2$, $R = 4$ and $R = 8$ abscissas and using a sample of 1,000,000 pseudo Monte Carlo draws. The patterns in the figure are not very different from those in Figure 2. Again, using $R = 1$ leads to a different ranking of the ten designs than 1,000,000 pseudo Monte Carlo samples, while using $R = 2$, $R = 4$ and $R = 8$ preserves the ranking. The main difference between Gauss-Hermite and Gauss-Stieltjes-Wigert quadrature is that the former method neither overestimates nor underestimates the Bayesian D-optimality

19

(a) $R = 1$

(b) $R = 2$

(c) $R = 4$

(d) $R = 8$
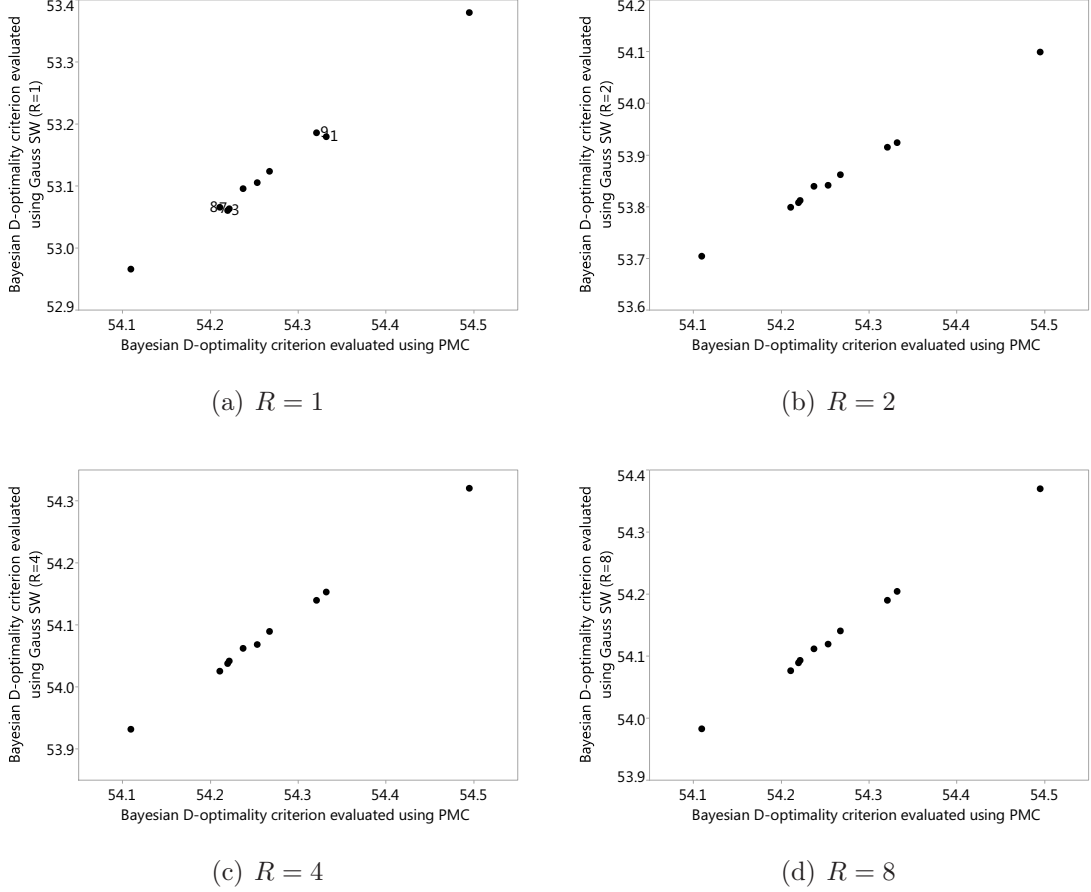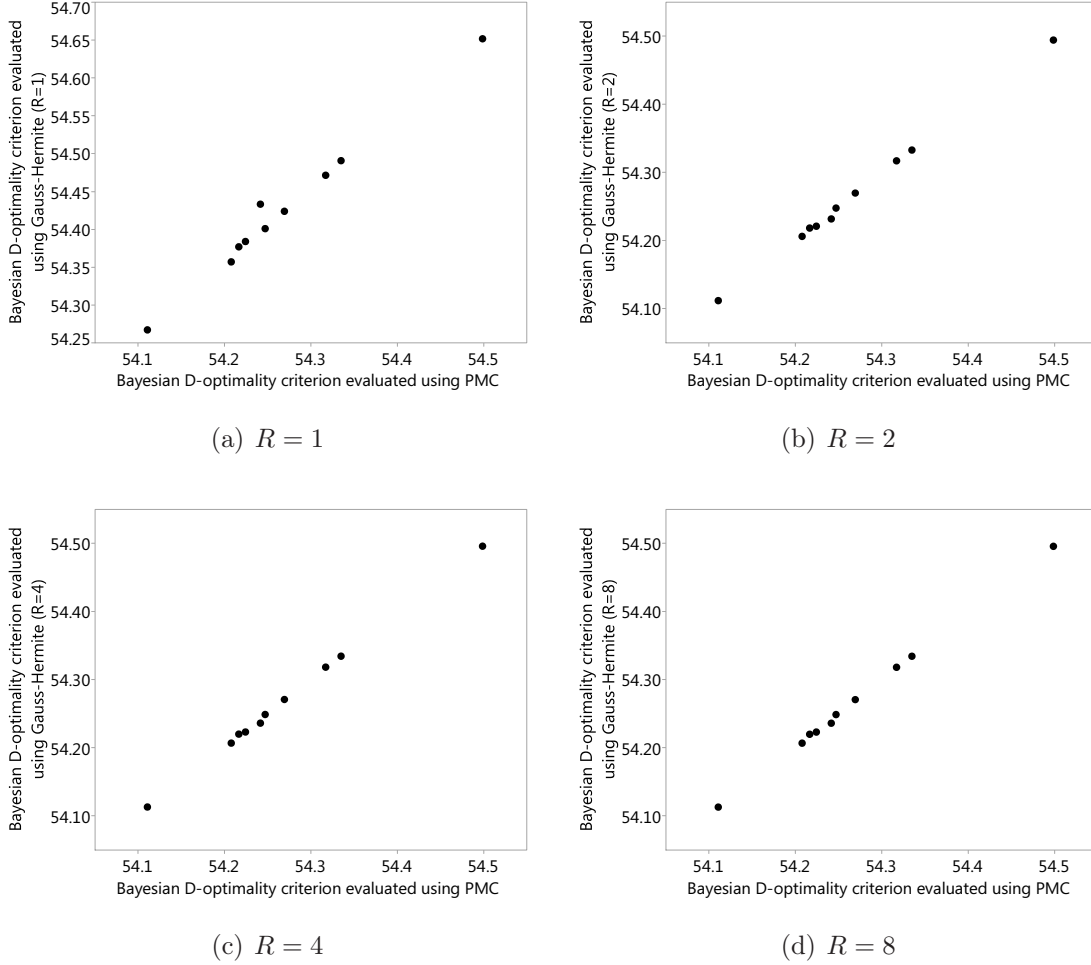
Figure 3: The Bayesian D-optimality criterion for ten designs evaluated using Gauss-Hermite quadrature (vertical axis) and 1,000,000 pseudo Monte Carlo draws.

criterion as soon as $R > 1$.

The fact that, for the ten designs we selected, we obtain the same ranking with $R = 2$ as with 1,000,000 pseudo Monte Carlo draws does not imply that we actually recommend using $R = 2$. However, it clearly shows that it is possible to rank designs at the expense of a much smaller computational effort than with large numbers of pseudo Monte Carlo draws.

Given the computational ease with which Gauss-Hermite and Gauss-Stieltjes-Wigert result in a correct ranking of the designs, one may think that any method involving a few draws from the prior distribution results in a correct ranking. For this reason, we studied the rankings obtained using 16 pseudo Monte Carlo draws. First, we evaluated the Bayesian

(a) $N = 16$ (b) $N = 1,000$
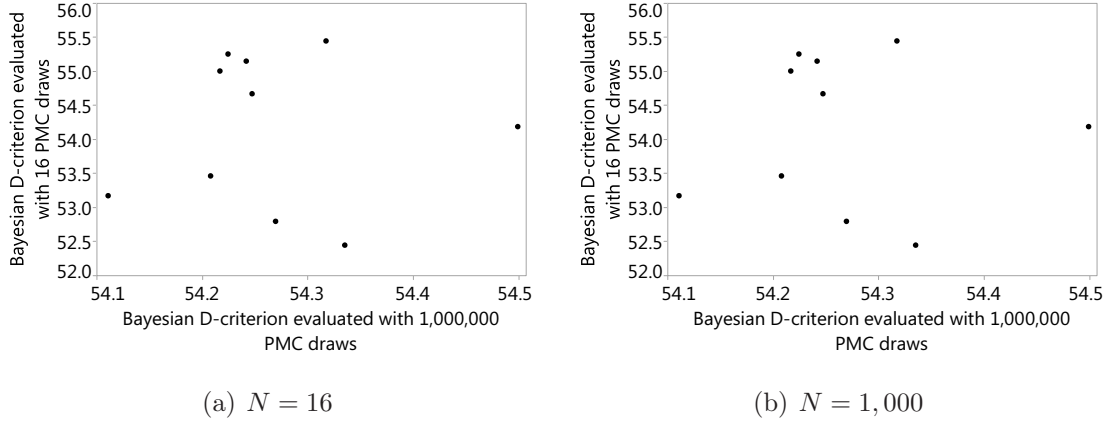
Figure 4: The Bayesian D-optimality criterion for ten designs evaluated using a limited number of pseudo Monte Carlo draws versus 1,000,000 pseudo Monte Carlo draws.

D-optimality criterion value of each of the 10 alternative designs with a different set of 16 pseudo Monte Carlo draws. This approach leads to an entirely incorrect ranking of the 10 designs, as can be seen from Figure 4a. For instance, the design which is best according to 1,000,000 pseudo Monte Carlo draws is only ranked 6th when 16 draws are used. Using the same approach with 1,000 instead of 16 pseudo Monte Carlo draws still does not result in a good ranking, as shown by Figure 4b. The design which is best according to 1,000,000 pseudo Monte Carlo draws is only ranked 3rd when 1,000 draws are used. With 1,000 draws, this implementation of the pseudo Monte Carlo approach is capable of identifying the top-3 designs correctly, but not in the correct order.

The results with 16 and 1,000 pseudo Monte Carlo draws are substantially better if the same set of 16 or 1,000 draws is used to evaluate the ten alternative designs. However, even then, the rankings obtained are generally incorrect. We tested the procedure described here four times. For one set of 16 draws, the resulting ranking only had a correlation of 0.8788 with the ranking based on 1,000,000 draws. For another set of 16 draws, the correlation was 1. The other two correlations were 0.9636 and 0.9879. The worst rank correlation for a sample of 1,000 pseudo Monte Carlo draws was 0.9997, indicating that using one sample of 1,000 draws nearly guarantees a perfect ranking.

As a summary, it should be clear that pseudo Monte Carlo sampling is not a good idea. Gauss-Hermite and Gauss-Stieltjes-Wigert sampling lead to a correct ranking of the ten

competing designs even with two, four or eight systematic draws. For pseudo Monte Carlo sampling, even 1,000 draws does not suffice. Given that generating systematic draws is instantaneous, there is no reasonable argument to opt for pseudo Monte Carlo sampling.

5   Gamma prior distribution for η

Another possible prior distribution $\pi_1(\eta)$ for the variance ratio η is the gamma distribution

$$\pi_1(\eta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot \eta^{\alpha-1} \cdot e^{-\beta\eta},$$

where $\eta \geq 0$, α is a strictly positive shape parameter, and β is a strictly positive rate parameter. The Bayesian design criterion in (5) then becomes

$$
\begin{aligned}
D_B &= \int_0^{+\infty} \Phi(\eta) \cdot \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot \eta^{\alpha-1} \cdot e^{-\beta\eta} \cdot d\eta, \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot \int_0^{+\infty} \Phi(\eta) \cdot \eta^{\alpha-1} \cdot e^{-\beta\eta} \cdot d\eta.
\end{aligned}
\tag{22}
$$

To evaluate this integral numerically, we can use generalized Gauss-Laguerre quadrature which handles integrals of the form

$$\int_0^{+\infty} f(x) \cdot x^\kappa \cdot e^{-\lambda x} \cdot dx. \tag{23}$$

The integrand in (22) has the same form as that in (23), which implies that the Bayesian optimality criterion can now be approximated as

$$
\begin{aligned}
D_B &\approx \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot \sum_{i=1}^R w_i^{\mathrm{GL}} \cdot \Phi\left(a_i^{\mathrm{GL}}\right), \\
&\approx \sum_{i=1}^R \widetilde{w}_i^{\mathrm{GL}} \cdot \Phi\left(d_i^{\mathrm{GL}}\right),
\end{aligned}
\tag{24}
$$

where $d_i^{\mathrm{GL}} = a_i^{\mathrm{GL}}$ is the ith systematic draw from the gamma prior distribution,

$$\widetilde{w}_i^{\mathrm{GL}} = \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot w_i^{\mathrm{GL}}$$

is the normalized weight, and the abscissas a_i^{GL} and weights w_i^{GL} are obtained from the generalized Laguerre polynomial (see, for instance, Kythe and Schäferkotter (2005)).

Table 4 shows the abscissas and weights for $R = 4$ and $R = 8$ for three sets of parameter values for the prior distribution: (i) $\alpha = \beta = 1$ (in which case the prior mean is one), (ii) $\alpha = 1$ and $\beta = 2$ (in which case the prior mean is 0.5), and (iii) $\alpha = 2$ and $\beta = 1$ (in which case the prior mean is 2). Note that the weights $\widetilde{w}_i^{\text{GL}}$ and the systematic draws d_i^{GL} attach a larger importance to small values of the variance ratio η when the prior mean is small.

Figure 5 compares generalized Gauss-Laguerre quadrature with pseudo Monte Carlo sampling with $N = 1,000$, $N = 10,000$, $N = 10^5$ and $N = 10^6$ for the split-plot design of Trinca and Gilmour (2001) involving 21 whole plots of two runs, using a gamma prior distribution with $\alpha = \beta = 1$. For each number of pseudo Monte Carlo draws, N, we approximated the Bayesian D-optimality criterion ten times and made box plots of the resulting approximations. We also approximated the Bayesian D-optimality criterion by means of $R = 16$ systematic draws produced by the generalized Gauss-Laguerre quadrature technique. The resulting approximate value of the optimality criterion was 51.0578, which is depicted by means of a horizontal reference line. Again, pseudo Monte Carlo sampling utilizing $N = 1,000$ draws results in substantial variability in the approximations of the optimality criterion. To reduce that variability substantially requires many more pseudo random draws.

In Table 5, we show the values of the Bayesian D-optimality criterion for the design of Trinca and Gilmour (2001) when evaluated by the generalized Gauss-Laguerre quadrature approach using 1 to 16 abscissas. The results in the table show that using 12–16 systematic draws leads to a close match to the values produced by one million pseudo Monte Carlo draws, shown in Figure 5.

Finally, for the gamma prior distribution, we created pictures similar to those in Figures 2 and 3. The patterns we observed for generalized Gauss-Laguerre quadrature and the gamma distribution are very similar to those for the log-normal distribution. The picture for $R = 8$, however, deserves a special mention, as its points very nearly lie on top of the 45 degree line. This is shown in Figure 6. This is different from Figures 2d and 3d, in which the points visibly deviate from the line. This means that generalized Gauss-Laguerre quadrature in combination with a gamma prior distribution allows a very precise approximation of the Bayesian D-optimality criterion, for each of the ten designs we used in our

23
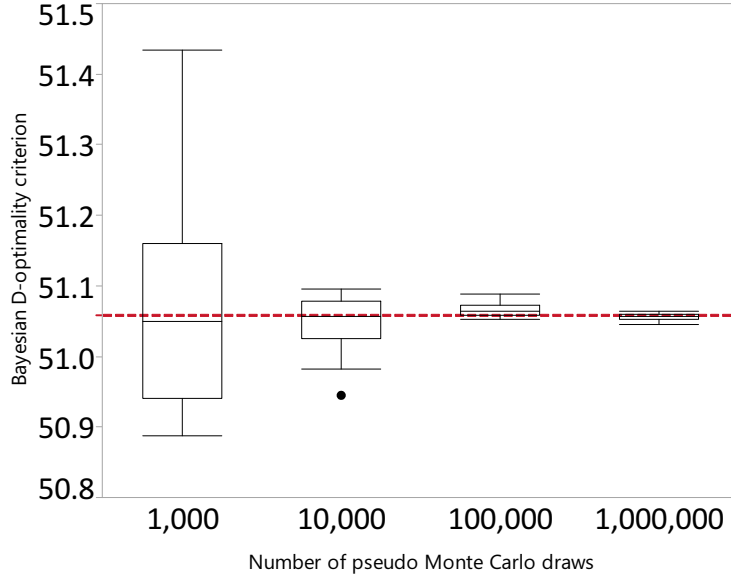
Figure 5: Box plots comparing the pseudo Monte Carlo approach with the generalized Gauss-Laguerre approach for a gamma prior distribution with $\alpha = \beta = 1$ and the split-plot design of Trinca and Gilmour (2001). The dashed reference line indicates the result from the generalized Gauss-Laguerre approach with $R = 16$.

comparison.

6 Beta prior distribution for ρ

When a prior distribution is specified for the intra-whole-plot correlation coefficient ρ, a natural choice is to use a beta distribution. This is because the domain of the beta distribution is the $[0, 1]$ interval, which is also the domain of ρ. We denote the beta prior distribution for ρ by

$$\pi(\rho) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \rho^{\alpha-1} \cdot (1 - \rho)^{\beta-1},$$

where $0 \leq \rho \leq 1$, $\alpha > 0$ and $\beta > 0$. The Bayesian design criterion in (6) then becomes

$$D_B = \int_0^1 \Phi(\rho) \cdot \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \rho^{\alpha-1} \cdot (1 - \rho)^{\beta-1} \cdot d\rho. \tag{25}$$
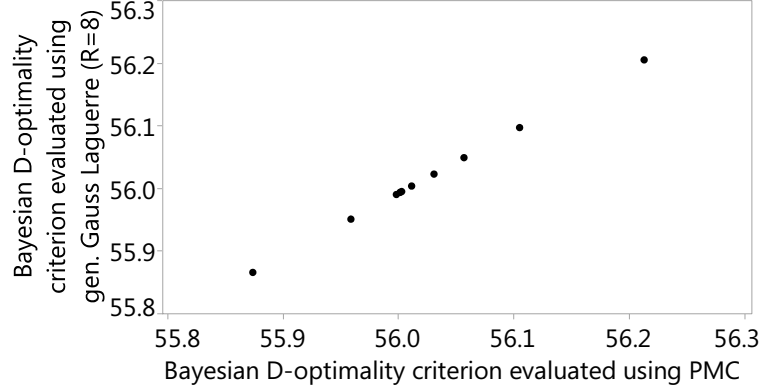
Figure 6: The Bayesian D-optimality criterion for ten designs evaluated using generalized Gauss-Laguerre quadrature with $R = 8$ (vertical axis) and 1,000,000 pseudo Monte Carlo draws.

This integral also has to be evaluated numerically. The appropriate quadrature technique for that purpose is Gauss-Jacobi quadrature, which handles integrals of the form

$$\int_{-1}^{1} f(x) \cdot (1+x)^{\kappa} \cdot (1-x)^{\lambda} \cdot dx. \tag{26}$$

The use of Gauss-Jacobi quadrature requires substituting ρ by $(1+y)/2$ in (25), which leads to the following alternative expression for the Bayesian design criterion:

$$
\begin{aligned}
D_B &= \frac{1}{2} \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \int_{-1}^{1} \Phi\Big(\frac{1+y}{2}\Big) \cdot \Big(\frac{1+y}{2}\Big)^{\alpha-1} \cdot \Big(1 - \frac{1+y}{2}\Big)^{\beta-1} \cdot dy, \\
&= \frac{1}{2^{\alpha+\beta-1}} \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \int_{-1}^{1} \Phi\Big(\frac{1+y}{2}\Big) \cdot (1+y)^{\alpha-1} \cdot (1-y)^{\beta-1} \cdot dy.
\end{aligned}
\tag{27}
$$

The integral in this expression is of the form given in (26). This implies that the Bayesian D-optimality criterion in (6) can be approximated as

$$
\begin{aligned}
D_B &\approx \frac{1}{2^{\alpha+\beta-1}} \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \sum_{i=1}^{R} w_i^{\mathrm{GJ}} \cdot \Phi\Big(\frac{1+a_i^{\mathrm{GJ}}}{2}\Big), \\
&\approx \sum_{i=1}^{R} \widetilde{w}_i^{\mathrm{GJ}} \cdot \Phi(d_i^{\mathrm{GJ}}),
\end{aligned}
\tag{28}
$$

where $d_i^{\mathrm{GJ}} = (1 + a_i)^{\mathrm{GJ}}/2$ is the ith systematic draw and

$$\widetilde{w}_i^{\mathrm{GJ}} = \frac{1}{2^{\alpha+\beta-1}} \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot w_i^{\mathrm{GJ}}$$

25

is the corresponding normalized weight. The R abscissas a_i^{GJ} and the weights w_i^{GJ} are obtained from the Jacobi polynomial (see, for instance, Rabinowitz (2001)).

Table 6 shows the abscissas a_i^{GJ}, the systematic draws d_i^{GJ} and weights w_i^{GJ} and $\widetilde{w}_i^{\text{GJ}}$ for $R = 12$ when $\alpha = \beta = 1$ (in which case the beta distribution reduces to the uniform distribution on the [0,1] interval and the prior mean is 1/2) or when $\alpha = 1$ and $\beta = 2$ (in which case the prior mean is 1/3). Because of the symmetry of the beta distribution around 1/2 whenever $\alpha = \beta$, the systematic draws and the weights will also exhibit symmetry. This can be seen in the left panel of Table 6, where $\alpha = \beta = 1$.

In Table 7, we show the values of the Bayesian D-optimality criterion obtained when evaluating the split-plot design of Trinca and Gilmour (2001) using Gauss-Jacobi quadrature for a beta prior distribution with $\alpha = \beta = 1$ and R values ranging from 1 to 16. Figure 7 compares Gauss-Jacobi quadrature using $R = 16$ with pseudo Monte Carlo sampling with $N = 1,000$, $N = 10,000$, $N = 10^5$ and $N = 10^6$ in that case. The figure shows, once more, that many pseudo Monte Carlo draws are required to achieve the same accuracy as Gauss-Jacobi quadrature. Finally, for the beta prior distribution, we created pictures similar to those in Figures 2 and 3. The patterns we observed for Gauss-Jacobi quadrature and the beta distribution are very similar to those for the log-normal distribution and the gamma distribution. The pictures for $R \geq 4$ have points that very nearly lie on top of the 45 degree line, as in Figure 6.

7 Independent prior distributions for σ_γ^2 and σ_ε^2

So far, we have focused on a log-normal or a gamma prior distribution for η and a beta prior distribution for ρ. Some researchers might prefer specifying a bivariate prior distribution or two independent prior distributions for σ_γ^2 and σ_ε^2. In the latter case, the Bayesian D-optimality criterion can be written as in (8).

7.1 Log-normal prior distributions

When independent log-normal prior distributions for σ_γ^2 and σ_ε^2 are considered, it is useful to realize that the ratio of two independent log-normally distributed random variables is also
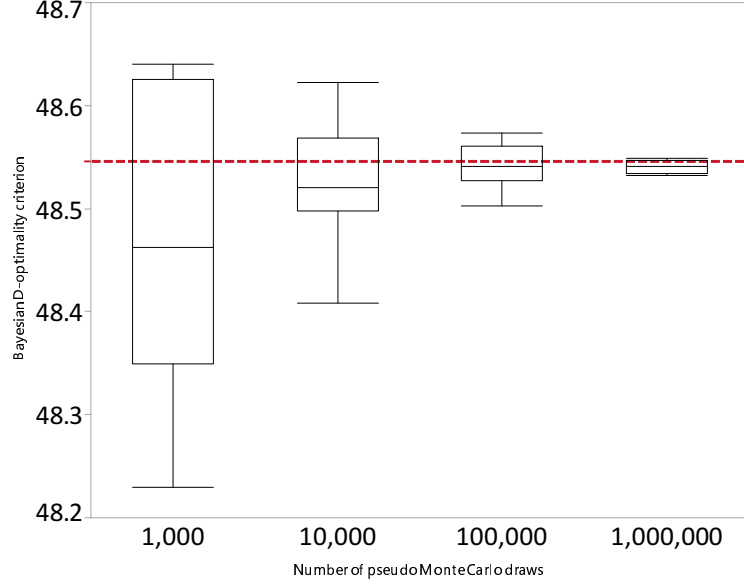
Figure 7: Box plots comparing the pseudo Monte Carlo approach with the Gauss-Jacobi approach for a beta prior distribution with $\alpha = \beta = 1$ and the split-plot design of Trinca and Gilmour (2001). The dashed reference line indicates the result from the Gauss-Jacobi approach with $R = 16$.

log-normally distributed. Assuming σ_γ^2 would be log-normally distributed with parameters μ_γ and σ_γ and that σ_ε^2 would be log-normally distributed with parameters μ_ε and σ_ε, the variance ratio $\eta = \sigma_\gamma^2/\sigma_\varepsilon^2$ would have a log-nogmal distribution with parameters $\mu_\gamma - \mu_\varepsilon$ and $\sqrt{\sigma_\gamma^2 + \sigma_\varepsilon^2}$. Using a single log-normal prior distribution for η is a computationally more efficient approach than using two independent log-normal distributions for σ_γ^2 and for σ_ε^2, since using the Bayesian optimality criterion then only requires evaluating a one-dimensional instead of a two-dimensional integral. In Section 4, we discussed how to handle a single log-normal prior distribution for the variance ratio η.

7.2 Gamma prior distributions

When independent gamma prior distributions for σ_γ^2 and for σ_ε^2 are considered, it is useful to know that using independent gamma prior distributions for σ_γ^2 and σ_ε^2 with the same rate parameter and shape parameters α_γ and α_ε, respectively, is equivalent to using a beta distribution with parameters α_γ and α_ε for the correlation coefficient $\rho = \sigma_\gamma^2/(\sigma_\gamma^2 + \sigma_\varepsilon^2)$.

27

Using a single beta prior distribution for ρ is a computationally more efficient alternative than using two independent gamma distributions for σ_γ^2 and for σ_ε^2, since using the Bayesian optimality criterion then only requires evaluation of a one-dimensional instead of a two-dimensional integral. In Section 6, we explained how the Bayesian D-optimality criterion can be evaluated at a low computational cost when a beta prior distribution is utilized for the correlation coefficient ρ.

Using independent gamma prior distributions with the same rate parameter and shape parameters α_γ and α_ε, respectively, for σ_γ^2 and for σ_ε^2 is also equivalent to taking a beta prime prior distribution with parameters α_γ and α_ε for the variance ratio η. A beta prime distribution, also known as a beta distribution of the second kind or an inverted beta distribution, with parameters α_γ and α_ε is given by

$$\pi_1(\eta) = \frac{\Gamma(\alpha_\gamma + \alpha_\varepsilon)}{\Gamma(\alpha_\gamma)\Gamma(\alpha_\varepsilon)} \cdot \eta^{\alpha_\gamma - 1} \cdot (1 + \eta)^{-\alpha_\gamma - \alpha_\varepsilon}$$

for $\eta > 0$, where the two shape parameters α_γ and α_ε are both strictly positive.

When a beta prime prior distribution is used for η, the Bayesian D-optimality criterion in (5) becomes

$$D_B = \int_0^{+\infty} \Phi(\eta) \cdot \frac{\Gamma(\alpha_\gamma + \alpha_\varepsilon)}{\Gamma(\alpha_\gamma)\Gamma(\alpha_\varepsilon)} \cdot \eta^{\alpha_\gamma - 1} \cdot (1 + \eta)^{-\alpha_\gamma - \alpha_\varepsilon} \cdot d\eta. \tag{29}$$

To evaluate this integral numerically, we can again use Gauss-Jacobi quadrature which handles integrals of the form (26). This first requires substituting η by $y/(1 - y)$ in (29), which leads to the following alternative expression for the Bayesian D-optimality criterion:

$$D_B = \frac{\Gamma(\alpha_\gamma + \alpha_\varepsilon)}{\Gamma(\alpha_\gamma)\Gamma(\alpha_\varepsilon)} \cdot \int_0^1 \Phi\left(\frac{y}{1 - y}\right) \cdot y^{\alpha_\gamma - 1} \cdot (1 - y)^{\alpha_\varepsilon - 1} \cdot dy. \tag{30}$$

Now, replacing y by $(1 + x)/2$ results in

$$D_B = \frac{\Gamma(\alpha_\gamma + \alpha_\varepsilon)}{\Gamma(\alpha_\gamma)\Gamma(\alpha_\varepsilon)} \cdot \int_{-1}^1 \Phi\left(\frac{(1 + x)/2}{1 - (1 + x)/2}\right) \cdot \left(\frac{1 + x}{2}\right)^{\alpha_\gamma - 1} \cdot \left(1 - \frac{1 + x}{2}\right)^{\alpha_\varepsilon - 1} \cdot dx.$$
$$= \frac{1}{2^{\alpha_\gamma + \alpha_\varepsilon - 1}} \cdot \frac{\Gamma(\alpha_\gamma + \alpha_\varepsilon)}{\Gamma(\alpha_\gamma)\Gamma(\alpha_\varepsilon)} \cdot \int_{-1}^1 \Phi\left(\frac{1 + x}{1 - x}\right) \cdot (1 + x)^{\alpha_\gamma - 1} \cdot (1 - x)^{\alpha_\varepsilon - 1} \cdot dx. \tag{31}$$

The integral in this expression has the same form as that in (26). The Bayesian D-optimality

criterion can therefore be approximated using Gauss-Jacobi quadrature, as

$$
\begin{aligned}
D_B &\approx \frac{1}{2^{\alpha_\gamma + \alpha_\varepsilon - 1}} \cdot \frac{\Gamma(\alpha_\gamma + \alpha_\varepsilon)}{\Gamma(\alpha_\gamma)\Gamma(\alpha_\varepsilon)} \cdot \sum_{i=1}^{R} w_i^{\mathrm{GJ}} \cdot \Phi\left(\frac{1 + a_i^{\mathrm{GJ}}}{1 - a_i^{\mathrm{GJ}}}\right), \\
&\approx \sum_{i=1}^{R} \widetilde{w}_i^{\mathrm{GJ}} \cdot \Phi(d_i^{\mathrm{GJ}}),
\end{aligned}
\tag{32}
$$

where $d_i^{\mathrm{GJ}} = (1 + a_i^{\mathrm{GJ}})/(1 - a_i^{\mathrm{GJ}})$ is the ith systematic draw from the beta prime prior distribution,

$$
\widetilde{w}_i^{\mathrm{GJ}} = \frac{1}{2^{\alpha_\gamma + \alpha_\varepsilon - 1}} \cdot \frac{\Gamma(\alpha_\gamma + \alpha_\varepsilon)}{\Gamma(\alpha_\gamma)\Gamma(\alpha_\varepsilon)} \cdot w_i^{\mathrm{GJ}}
$$

is the corresponding normalized weight, and the R abscissas a_i^{GJ} and weights w_i^{GJ} are again obtained from the Jacobi polynomial.

Table 8 shows the abscissas a_i^{GJ}, the systematic draws d_i^{GJ} and weights w_i^{GJ} and $\widetilde{w}_i^{\mathrm{GJ}}$ for $R = 12$ when the beta prime prior distribution has parameters $\alpha_\gamma = \alpha_\varepsilon = 1$ and when it has parameters $\alpha_\gamma = 1$ and $\alpha_\varepsilon = 2$. Comparing Table 8 for the beta prime distribution with Table 6 for the beta distribution, we can see that the abscissas a_i^{GJ} give rise to completely different draws d_i^{GJ} in both cases, but to the same normalized weights $\widetilde{w}_i^{\mathrm{GJ}}$.

The values of the Bayesian D-optimality criterion for the split-plot design of Trinca and Gilmour (2001), when evaluated using R values ranging from 1 to 16 in case a beta prime prior distribution with $\alpha_\gamma = \alpha_\varepsilon = 1$ is used for η, are exactly the same as those in Table 7 for a beta distribution with $\alpha = \beta = 1$ for ρ. Figure 8 compares Gauss-Jacobi quadrature using $R = 16$ with pseudo Monte Carlo sampling with $N = 1,000$, $N = 10,000$, $N = 10^5$ and $N = 10^6$ for the beta prime prior distribution. The figure is in line with earlier figures for other distributions, showing the superiority of the quadrature technique.

8   Handling multidimensional integrals

In this tutorial, we discussed scenarios in which the design optimality criterion either involves a one-dimensional integration over a prior distribution for one parameter, or a two-dimensional integral that can be reduced to a one-dimensional integral due to the independence and compatibility of the two prior distributions involved. For practical applications in which the optimal experimental design depends on two or more unknown parameters,
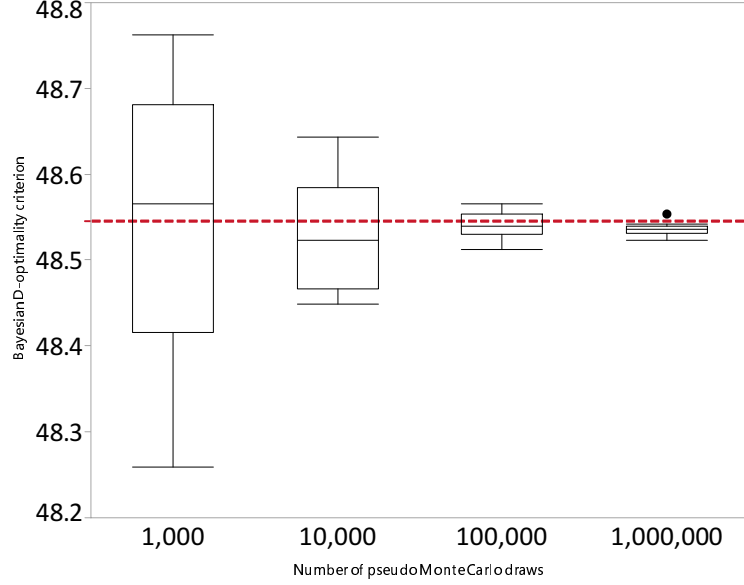
Figure 8: Box plots comparing the pseudo Monte Carlo approach with the Gauss-Jacobi approach for a beta prime prior distribution with $\alpha_\gamma = \alpha_\varepsilon = 1$. The dashed reference line indicates the result from the Gauss-Jacobi approach with $R = 16$.

however, it will be impossible to reduce the multidimensional integral to a one-dimensional integral. In that case, we can use multidimensional quadrature, which is sometimes referred to as cubature.

If we denote the k unknown parameters an optimal experimental design depends on by $\theta_1, \theta_2, \ldots, \theta_k$, and use independent prior distributions $\pi_1(\theta_1), \pi_2(\theta_2), \ldots, \pi_k(\theta_k)$ for them, then the Bayesian D-optimality criterion in (8) can be generalized to

$$D_B = \int_{D_1} \int_{D_2} \cdots \int_{D_k} \Phi(\theta_1, \theta_2, \ldots, \theta_k) \cdot \left(\prod_{i=1}^{k} \pi_i(\theta_i) \right) \cdot d\theta_1 \cdot d\theta_2 \cdots d\theta_k, \qquad (33)$$

where $D_1, D_2, \ldots, D_k$ denote the domains for the k parameters. This integral can be evaluated numerically by generalizing (12) to

$$D_B \approx \sum_{i_1=1}^{R} \sum_{i_2=1}^{R} \cdots \sum_{i_k=1}^{R} \left(\prod_{l=1}^{k} w_l^{i_l} \right) \Phi(\theta_1^{i_1}, \theta_2^{i_2}, \ldots, \theta_k^{i_k}), \qquad (34)$$

where $\theta_l^{i_l}$ is the i_lth systematic draw from the prior distribution $\pi_l(\theta_l)$ for the lth parameter θ_l and $w_l^{i_l}$ is the corresponding quadrature weight. In mathematical terms, this approach

30

views the multidimensional integral as a sequence of nested univariate integrals and combines univariate quadrature techniques in a tensor product fashion. It is generally taken to be the multivariate quadrature approach. This approach's drawback is that it uses a total of R^k evaluations of the optimality criterion $\Phi(\theta_1, \theta_2, \ldots, \theta_k)$. So, in this approach, the number of criterion evaluations as well as the computing time increase exponentially with the number of unknown parameters. This problem is sometimes called the curse of dimensionality.

It should be pointed out, however, that, for many problems, the dimension k of the integral in (33) is rather low. One reason for this is that researchers generally use covariance structures which are parsimonious in the number of parameters. As a matter of fact, the most commonly used serial and spatial correlation patterns are described by just a few parameters. Similarly, the most complex multi-stratum experimental designs in the literature only involve a few strata and, hence, only a few variance components. For instance, compared to the split-plot designs considered in this article, split-split-plot, strip-plot and staggered-level designs only involve one additional variance component. Finally, the most commonly used non-linear regression models involve at most 2–5 parameters. Given that it generally suffices to use small R values, the total number of optimality criterion evaluations required, R^k, will still be rather small compared to the number needed when using pseudo Monte Carlo sampling.

One possible remedy for the curse of dimensionality in case the number of parameters is too large to use the multidimensional quadrature approach in (34) is to use so-called sparse grids. Heiss and Winschel (2008) explain that the approach in (34) is exact for products of Rth-order univariate polynomials. Therefore, it is exact for multivariate polynomials containing specific terms of order R^k, since the highest-order terms obtained when multiplying k polynomials of order R have order R^k. The main idea of sparse grids is that the interest is in a polynomial of order R and that, therefore, no computational effort should be spent on terms of a higher order. Heiss and Winschel (2008) nicely visualize the principle behind sparse grids, provide the required mathematical expressions and explain that the computing time increase considerably slower than exponentially. For example, for a problem with $k = 5$ unknown parameters, the number of optimality criterion evaluations required for

achieving a polynomial exactness of 5 will drop from $R^k = 3^5 = 243$ to 61 when using the Smolyak construction of sparse grids. Some technical background concerning the Smolyak construction can be found in Petras (2003). At present, the sparse grid technology is available for uniform prior distributions and for normal prior distributions for the k parameters, due to Genz and Keister (1996), Petras (2003) and Heiss and Winschel (2008). Matlab and Stata code for generating the grids can be found at http://www.sparse-grids.de/.

9 Discussion

In this paper, we have described several quadrature techniques that can be used for the evaluation of the Bayesian D-optimality criterion when constructing optimal designs that depend on one or more unknown model parameters. To illustrate the techniques, we focused on split-plot experiments which depend on one (the variance ratio η or the correlation coefficient ρ) or two (the two variance components) parameters. We wish to emphasize that the techniques are more broadly applicable, however, in scenarios where the usual normal prior distribution makes no sense.

The small computational cost of the quadrature techniques lowers the hurdle for using Bayesian D-optimal experimental design more routinely, instead of the locally optimal design approach, where the prior uncertainty about the unknown model parameter(s) is entirely ignored. This tutorial paper aims to make the quadrature approach accessible to researchers who are not familiar with state-of-the-art numerical integration techniques. By making our Matlab programs for generating abscissas and weights for various Gaussian quadrature techniques available in the supplementary materials, we hope that many researchers make the step from locally optimal designs to Bayesian optimal designs, or from a computationally intensive Bayesian approach involving large pseudo Monte Carlo samples to more efficient Bayesian approaches involving a limited number of systematic draws from the prior distribution(s) chosen. It should be pointed out that our Matlab programs are based on the parametrizations of the prior distributions used in this paper, and that different software packages generally use different implementations of the quadrature rules.

Throughout the paper, we focus on precise approximations of the Bayesian optimality criterion. However, in the search for optimal experimental designs, it is perhaps the ranking

of designs that matters most, rather than a precise approximation. For our Bayesian D-optimality criterion, using four systematic draws always led to a correct ranking of ten high quality designs for a particular split-plot scenario. As Figures 2, 3 and 4 illustrate, using one systematic draw and using 1,000 or fewer pseudo Monte Carlo draws are definitely not good ideas when it comes to ranking competing designs. Also, for the split-plot design problem we studied in this paper, the optimal design construction algorithm does not yield substantially better designs when more than four systematic draws are used. We do not recommend Gauss-Stieltjes-Wigert quadrature as we have observed that it leads to an underestimation of the Bayesian D-optimality criterion when $R \leq 16$, even though it is, in theory, more appropriate than Gauss-Hermite quadrature for the log-normal distribution and seems to perform well in terms of ranking designs.

We limited our attention to univariate (log-normal, gamma, beta and beta prime) prior distributions for a variance ratio and a correlation coefficient, and to two independent log-normal or gamma distributions for a variance. In some scenarios, for instance those involving designs with more than one blocking factor, split-split-plot, strip-plot or staggered-level designs (see Goos and Donev (2006), Jones and Goos (2009), Arnouts et al. (2010, 2013) and Arnouts and Goos (2012, 2015)), prior distributions are required for more than one variance ratio or for more than one correlation coefficient. Similarly, the optimal design for the compartmental model in Atkinson, Donev and Tobias (2007) depends on two parameters. It will usually not be realistic to assume that the prior information concerning several variances, several correlation coefficients or several other parameters are independent. It is therefore useful to study multivariate prior distributions for variances and correlation coefficients, as well as suitable quadrature techniques for these. This research will also prove useful for optimal experimental design problems where the serial or spatial correlation pattern is described by more than one parameter.

In the paper, we did not pay explicit attention to uniform prior distributions on a general interval $[a, b]$ or to inverse gamma prior distributions. To handle uniform prior distributions on the $[-1, 1]$ interval, we rely on Gauss-Legendre quadrature, which is a special case of Gauss-Jacobi quadrature. Also, note that the interval $[-1, 1]$ can be replaced with any other interval through a simple change of variable.

Obviously, the univariate and multivariate normal prior distributions are of great importance too. We did not pay attention to normal prior distributions because the Gauss-Hermite quadrature technique for a multivariate normal distribution was thoroughly described by Bliemer et al. (2009) and Yu et al. (2010), who also demonstrate the usefulness of the spherical-radial integration approach in Gotwalt et al. (2009) (see also Gotwalt (2010)) as an alternative to Gauss-Hermite quadrature. To the best of our knowledge, however, neither the plain multidimensional Gauss-Hermite quadrature not the spherical-radial integration approach of Gotwalt et al. (2009) has been compared to the sparse grids for Gaussian prior distributions, introduced by Genz and Keister (1996). Such a comparison would be an interesting topic for future research.

Finally, some readers may have read this paper with the hope that we would recommend specific quadrature techniques for specific kinds of models and specific design optimality criteria. All quadrature techniques we studied, except for Gauss-Stieltjes-Wigert quadrature, performed well for small numbers of systematic draws, for the problem we studied in this paper. In our view, which quadrature technique should be used should depend only on the way in which the prior information concerning the unknown parameters is specified. So, the choice of quadrature technique should only depend on the nature of the prior distribution. We also believe that it is a good practice to always compare the quadrature technique considered to a large pseudo Monte Carlo sample for a few feasible experimental designs, whenever studying a new kind of model and a new kind of design optimality criterion. Also, quadrature techniques for Bayesian optimal experimental design are useful for any design construction algorithm, since any algorithm for constructing designs will have to evaluate the Bayesian optimality criterion at hand on many occasions. In this paper, the designs we evaluated had been generated using a coordinate-exchange algorithm, but we could have used a point-exchange algorithm, a variable neighborhood search algorithm or any other metaheuristic instead.

Acknowledgment

would also like to thank the two referees for their constructive comments and suggestions.

References

Androulakis, E., Angelopoulos, P. and Koukouvinos, C. (2013). A comparison of three-level orthogonal arrays in the presence of different correlation structures in observations, *Communications in Statistics: Simulation and Computation* **42**: 552–569.

Arnouts, H. and Goos, P. (2012). Staggered-level designs for experiments with more than one hard-to-change factor, *Technometrics* **54**: 355–366.

Arnouts, H. and Goos, P. (2015). Staggered-level designs for response surface modeling, *Journal of Quality Technology* **47**: 156–175.

Arnouts, H., Goos, P. and Jones, B. (2010). Design and analysis of industrial strip-plot experiments, *Quality and Reliability Engineering International* **26**: 127–136.

Arnouts, H., Goos, P. and Jones, B. (2013). Three-stage strip-plot designs, *Journal of Quality Technology* **45**: 1–17.

Atkinson, A. C., Donev, A. N. and Tobias, R. D. (2007). *Optimum Experimental Designs, with SAS*, Oxford: Oxford University Press.

Bliemer, M. C. J., Rose, J. M. and Hess, S. (2009). Approximation of Bayesian efficiency in experimental choice designs, *Journal of Choice Modeling* **1**: 98–127.

Borrotti, M., Sambo, F., Mylona, K. and Gilmour, S. G. (2016). A multi-objective coordinate-exchange two-phase local search algorithm for multi-stratum experiments, *Statistics and Computing* **26**: to appear.

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

Dette, H. (1997). Designing experiments with respect to 'standardized' optimality criteria, *Journal of the Royal Statistical Society, Series B* **59**: 97–110.

Dette, H., Haines, L. M. and Imhof, L. A. (2007). Maximin and Bayesian optimal designs for regression models, *Statistica Sinica* **17**: 463–480.

Diggle, P. and Lophaven, S. (2005). Bayesian geostatistical design, *Scandinavian Journal of Statistics* **33**: 53–65.

Elliott, L. J., Eccleston, J. and Martin, R. J. (1999). An algorithm for the design of factorial experiments when the data are correlated, *Statistics and Computing* **9**: 195–201.

Fedorov, V. V. and Leonov, S. L. (2014). *Optimal Design for Nonlinear Response Models*, Chapman and Hall/CRC.

Garroi, J.-J., Goos, P. and Sörensen, K. (2009). A variable-neighbourhood search algorithm for finding optimal run orders in the presence of serial correlation, *Journal of Statistical Planning and Inference* **139**: 30–44.

Gautschi, W. (2004). *Orthogonal Polynomials: Computation and Approximation*, New York: Oxford University Press.

Genz, A. and Keister, B. D. (1996). Fully symmetric interpolatory rules for multiple integrals over infinite regions with Gaussian weight, *Journal of Computational and Applied Mathematics* **71**: 299–309.

Golub, G. H. and Welsch, J. H. (1969). Calculation of Gauss quadrature rules, *Mathematics of Computation* **23**: 221–230.

Goos, P. (2002). *The Optimal Design of Blocked and Split-plot Experiments*, New York: Springer.

Goos, P. and Donev, A. N. (2006). Blocking response surface designs, *Computational Statistics and Data Analysis* **51**: 1075–1088.

Goos, P. and Jones, B. (2011). *Design of Experiments: A Case Study Approach*, New York: Wiley.

Goos, P. and Vandebroek, M. (2001a). D-optimal response surface designs in the presence of random block effects, *Computational Statistics and Data Analysis* **37**: 433–453.

Goos, P. and Vandebroek, M. (2001b). Optimal split-plot designs, *Journal of Quality Technology* **33**: 436–450.

Goos, P. and Vandebroek, M. (2003). D-optimal split-plot designs with given numbers and sizes of whole plots, *Technometrics* **45**: 235–245.

Goos, P. and Vandebroek, M. (2004). Outperforming completely randomized designs, *Journal of Quality Technology* **36**: 12–26.

Gotwalt, C. M. (2010). Addendum to "Fast computation of designs robust to parameter uncertainty for nonlinear settings", *Technometrics* **52**: 137–137.

Gotwalt, C. M., Jones, B. A. and Steinberg, D. M. (2009). Fast computation of designs robust to parameter uncertainty for nonlinear settings, *Technometrics* **51**: 88–95.

Harari, O. and Steinberg, D. M. (2014). Optimal designs for Gaussian process models via spectral decomposition, *Journal of Statistical Planning and Inference* **154**: 87–101.

Heiss, F. and Winschel, V. (2008). Likelihood approximation by numerical integration on sparse grids, *Journal of Econometrics* **144**: 62–80.

Imhof, L. A. (2001). Maximin designs for exponential growth models and heteroscedastic polynomial models, *Annals of Statistics* **17**: 561–576.

Jones, B. (1986). Searching for optimal block designs when block effects are random, *Biometrical Journal* **28**: 909–918.

Jones, B. and Goos, P. (2007). A candidate-set-free algorithm for generating D-optimal split-plot designs, *Journal of the Royal Statistical Society. Series C* **56**: 347–364.

Jones, B. and Goos, P. (2009). D-optimal design of split-split-plot experiments, *Biometrika* **96**: 67–82.

Jones, B. and Goos, P. (2012). I-optimal versus D-optimal split-plot response-surface designs, *Journal of Quality Technology* **44**: 85–101.

Kessels, R., Jones, B., Goos, P. and Vandebroek, M. (2009). An efficient algorithm for constructing Bayesian optimal choice designs, *Journal of Business and Economic Statistics* **27**: 279–291.

Kessels, R., Jones, B., Goos, P. and Vandebroek, M. (2011). The usefulness of Bayesian optimal designs for discrete choice experiments, *Applied Stochastic Models in Business and Industry* **27**: 173–188.

Kythe, P. K. and Schäferkotter, M. R. (2005). *Handbook of Computational Methods for Integration*, Boca Raton: Chapman and Hall/CRC.

Macharia, H. and Goos, P. (2010). D-optimal and D-efficient equivalent-estimation second-order split-plot designs, *Journal of Quality Technology* **42**: 358–372.

Martin, R. J., Eccleston, J. A. and Jones, G. (1998). Some results on multi-level factorial designs with dependent observations, *Journal of Statistical Planning and Inference* **73**: 91–111.

Martin, R. J., Jones, G. and Eccleston, J. A. (1998). Some results on two-level factorial designs with dependent observations, *Journal of Statistical Planning and Inference* **66**: 363–384.

Mastroianni, G. and Milovanovic, G. V. (2008). *Interpolation Processes: Basic Theory and Applications*, Springer: Verlag Berlin Heidelberg.

Mylona, K., Goos, P. and Jones, B. (2014). Optimal design of blocked and split-plot experiments for fixed effects and variance component estimation, *Technometrics* **56**: 132–144.

Petras, K. (2003). Smolyak cubature of given polynomial degree with few nodes for increasing dimension, *Numerische Mathematik* **93**: 729–753.

Rabinowitz, P. (2001). *A First Course in Numerical Analysis*, New York: Dover Publications.

Sambo, F., Borrotti, M. and Mylona, K. (2014). A coordinate-exchange two-phase local search algorithm for the D- and I-optimal design of split-plot experiments, *Computational Statistics and Data Analysis* **71**: 1193–1207.

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

Schoen, E. D., Jones, B. and Goos, P. (2011). A split-plot experiment with factor-dependent whole-plot sizes, *Journal of Quality Technology* **43**: 66–79.

Sethuraman, V. and Raghavarao, D. (2009). Balanced 2^n factorial designs when observations are spatially correlated, *Journal of Biopharmaceutical Statistics* **19**: 332–344.

Stoer, J. and Bulirsch, R. (2002). *Introduction to Numerical Analysis*, Springer.

Trinca, L. A. and Gilmour, S. G. (2001). Multi-stratum response surface designs, *Technometrics* **43**: 25–33.

Trinca, L. A. and Gilmour, S. G. (2015). Improved split-plot and multi-stratum designs, *Technometrics* **57**: 145–154.

Wilck, M. (2001). A general approximation method for solving integrals containing a lognormal weighting function, *Journal of Aerosol Science* **32**: 1111–1116.

Woods, D. C., Lewis, S. M., Eccleston, J. A. and Russell, K. G. (2006). Designs for generalized linear models with several variables and model uncertainty, *Technometrics* **48**: 284–292.

Yu, J., Goos, P. and Vandebroek, M. (2010). Comparing different sampling schemes for approximating the integrals involved in the efficient design of stated choice experiments, *Transportation Research Part B: Methodological* **44**: 1268–1289.

Table 1: Abscissas a_i^{GH} and weights w_i^{GH} and $\widetilde{w}_i^{\mathrm{GH}}$ for Gauss-Hermite quadrature, as well as systematic draws d_i^{GH} constructed from them assuming that $\mu = 0$ and $\sigma = 0.5$ and that $\mu = 3$ and $\sigma = 1$.

R	i	a_i^{GH}	w_i^{GH}	$\widetilde{w}_i^{\mathrm{GH}}$	d_i^{GH} $(\mu = 0, \sigma = 0.5)$	d_i^{GH} $(\mu = 3, \sigma = 1)$
1	1	0.0000	1.7725	1.0000	1.0000	20.0855
2	1	-0.7071	0.8862	0.5000	0.6065	7.3891
	2	0.7071	0.8862	0.5000	1.6487	54.5982
3	1	-1.2247	0.2954	0.1667	0.4206	3.5536
	2	0.0000	1.1816	0.6667	1.0000	20.0855
	3	1.2247	0.2954	0.1667	2.3774	113.5281
4	1	-1.6507	0.0813	0.0459	0.3112	1.9456
	2	-0.5246	0.8049	0.4541	0.6901	9.5643
	3	0.5246	0.8049	0.4541	1.4492	42.1807
	4	1.6507	0.0813	0.0459	3.2130	207.3513
5	1	-2.0202	0.0200	0.0113	0.2397	1.1538
	2	-0.9586	0.3936	0.2221	0.5077	5.1778
	3	0.0000	0.9453	0.5333	1.0000	20.0855
	4	0.9586	0.3936	0.2221	1.9696	77.9156
	5	2.0202	0.0200	0.0113	4.1724	349.6631

Table 2: The effect of different numbers of systematic draws, R, on the evaluation of the Bayesian D-optimality criterion for the five-factor split-plot design of Trinca and Gilmour (2001), using a log-normal prior distribution with $\mu = 0$ and $\sigma = 0.75$, by means of Gauss-Hermite and Gauss-Stieltjes-Wigert quadrature.

R	Gauss-Hermite	Gauss-Stieltjes-Wigert
1	49.3401	47.9558
2	49.1822	48.7448
3	49.1827	48.9246
4	49.1835	48.9892
5	49.1833	49.0178
6	49.1833	49.0321
7	49.1833	49.0396
8	49.1833	49.0436
12	49.1833	49.0483
16	49.1833	49.0488

Table 3: Abscissas a_i^{GSW}, systematic draws d_i^{GSW} and weights w_i^{GSW} for Gauss-Stieltjes-Wigert quadrature with $\mu = 0$ and $\sigma = 0.5$ and with $\mu = 3$ and $\sigma = 1$ for various values of R.

		$\mu = 0$, $\sigma = 0.5$			$\mu = 3$, $\sigma = 1$		
R	i	a_i^{GSW}	d_i^{GSW}	w_i^{GSW}	a_i^{GSW}	d_i^{GSW}	w_i^{GSW}
1	1	1.1331	1.1331	1.0000	1.6487	33.1155	1.0000
2	1	0.8591	0.8591	0.8293	1.3080	26.2714	0.9757
	2	2.4641	2.4641	0.1707	15.3562	308.4377	0.0243
3	1	0.7359	0.7359	0.6652	1.2173	24.4511	0.9608
	2	1.8682	1.8682	0.3285	12.1825	244.6919	0.0392
	3	4.7432	4.7432	0.0063	121.9152	2448.7330	9.57E-06
4	1	0.6657	0.6657	0.5507	1.1873	23.8469	0.9547
	2	1.5954	1.5954	0.4248	11.3306	227.5804	0.0453
	3	3.6069	3.6069	0.0245	96.7854	1943.9876	2.48E-05
	4	8.6446	8.6446	6.98E-05	923.6629	18552.2649	7.27E-11
5	1	0.6212	0.6212	0.4729	1.1766	23.6325	0.9524
	2	1.4392	1.4392	0.4789	11.0470	221.8856	0.0475
	3	3.0802	3.0802	0.0477	90.0171	1808.0424	3.33E-05
	4	6.5924	6.5924	5.08E-04	733.5078	14732.8987	3.00E-10
	5	15.2744	15.2744	2.61E-07	6886.9134	138327.3526	1.06E-17

Table 4: Abscissas and systematic draws $a_i^{\mathrm{GL}} = d_i^{\mathrm{GL}}$ and weights w_i^{GL} and $\widetilde{w}_i^{\mathrm{GL}}$ for generalized Gauss-Laguerre quadrature for a gamma prior distribution with (i) $\alpha = \beta = 1$, (ii) $\alpha = 1$ and $\beta = 2$, and (iii) $\alpha = 2$ and $\beta = 1$, when $R = 4$ and $R = 8$.

		$\alpha = \beta = 1$			$\alpha = 1, \beta = 2$			$\alpha = 2, \beta = 1$		
R	i	$a_i^{\mathrm{GL}} = d_i^{\mathrm{GL}}$	w_i^{GL}	$\widetilde{w}_i^{\mathrm{GL}}$	$a_i^{\mathrm{GL}} = d_i^{\mathrm{GL}}$	w_i^{GL}	$\widetilde{w}_i^{\mathrm{GL}}$	$a_i^{\mathrm{GL}} = d_i^{\mathrm{GL}}$	w_i^{GL}	$\widetilde{w}_i^{\mathrm{GL}}$
4	1	0.3225	0.6032	0.6032	0.1613	0.3016	0.6032	0.7433	0.4469	0.4469
	2	1.7458	0.3574	0.3574	0.8729	0.1787	0.3574	2.5716	0.4776	0.4776
	3	4.5366	0.0389	0.0389	2.2683	0.0194	0.0389	5.7312	0.0742	0.0742
	4	9.3951	0.0005	0.0005	4.6975	0.0003	0.0005	10.9539	0.0013	0.0013
8	1	0.1703	0.3692	0.3692	0.0851	0.1846	0.3692	0.4094	0.1876	0.1876
	2	0.9037	0.4188	0.4188	0.4519	0.2094	0.4188	1.3850	0.4390	0.4390
	3	2.2511	0.1758	0.1758	1.1255	0.0879	0.1758	2.9563	0.2900	0.2900
	4	4.2667	0.0333	0.0333	2.1334	0.0167	0.0333	5.1819	0.0751	0.0751
	5	7.0459	0.0028	0.0028	3.5230	0.0014	0.0028	8.1617	0.0079	0.0079
	6	10.7585	0.0001	0.0001	5.3793	4.54E-05	9.08E-05	12.0701	3.09E-04	3.09E-04
	7	15.7407	8.49E-07	8.49E-07	7.8703	4.24E-07	8.49E-07	17.2497	3.35E-06	3.35E-06
	8	22.8631	1.05E-09	1.05E-09	11.4316	5.24E-10	1.05E-09	24.5860	4.72E-09	4.72E-09

Table 5: Values obtained for the Bayesian D-optimality criterion for a gamma prior distribution with $\alpha = \beta = 1$ when using the generalized Gauss-Laguerre quadrature approach with one to sixteen abscissas for the five-factor split-plot design of Trinca and Gilmour (2001).

R	D_B	R	D_B
1	49.3401	6	51.0387
2	50.5815	7	51.0474
3	50.8806	8	51.0519
4	50.9808	12	51.0571
5	51.0208	16	51.0578

Table 6: Abscissas a_i^{GJ}, systematic draws d_i^{GJ} and weights w_i^{GJ} and $\widetilde{w}_i^{\mathrm{GJ}}$ obtained using Gauss-Jacobi quadrature with $R = 12$ when a beta prior distribution is used for ρ, with parameters $\alpha = \beta = 1$ or $\alpha = 1$ and $\beta = 2$.

i	$\alpha = \beta = 1$				$\alpha = 1,\ \beta = 2$			
	a_i^{GJ}	d_i^{GJ}	w_i^{GJ}	$\widetilde{w}_i^{\mathrm{GJ}}$	a_i^{GJ}	d_i^{GJ}	w_i^{GJ}	$\widetilde{w}_i^{\mathrm{GJ}}$
1	-0.9816	0.0092	0.0472	0.0236	-0.9829	0.0085	0.0867	0.0433
2	-0.9041	0.0479	0.1069	0.0535	-0.9111	0.0444	0.1897	0.0949
3	-0.7699	0.1150	0.1601	0.0800	-0.7863	0.1069	0.2664	0.1332
4	-0.5873	0.2063	0.2032	0.1016	-0.6157	0.1922	0.3076	0.1538
5	-0.3678	0.3161	0.2335	0.1167	-0.4092	0.2954	0.3108	0.1554
6	-0.1252	0.4374	0.2491	0.1246	-0.1789	0.4105	0.2804	0.1402
7	0.1252	0.5626	0.2491	0.1246	0.0619	0.5310	0.2265	0.1132
8	0.3678	0.6839	0.2335	0.1167	0.2992	0.6496	0.1618	0.0809
9	0.5873	0.7937	0.2032	0.1016	0.5192	0.7596	0.0995	0.0498
10	0.7699	0.8850	0.1601	0.0800	0.7091	0.8546	0.0497	0.0249
11	0.9041	0.9521	0.1069	0.0535	0.8579	0.9289	0.0178	0.0089
12	0.9816	0.9908	0.0472	0.0236	0.9569	0.9784	0.0031	0.0016

Table 7: Values obtained for the Bayesian D-optimality criterion for the split-plot design of Trinca and Gilmour (2001) using the Gauss-Jacobi approach with one to sixteen abscissas for a beta prior distribution for ρ with $\alpha = \beta = 1$ and using the Gauss-Jacobi approach for the beta prime distribution for η with $\alpha_\gamma = \alpha_\varepsilon = 1$.

R	D_B	R	D_B
1	49.3401	6	48.5833
2	48.8551	7	48.5721
3	48.6960	8	48.5646
4	48.6328	12	48.5504
5	48.6013	16	48.5453

Table 8: Abscissas a_i^{GJ}, systematic draws d_i^{GJ} and weights w_i^{GJ} and $\widetilde{w}_i^{\mathrm{GJ}}$ obtained using Gauss-Jacobi quadrature with $R = 12$ when a beta prime prior distribution is used for η, with parameters $\alpha_\gamma = \alpha_\varepsilon = 1$ or $\alpha_\gamma = 1$ and $\alpha_\varepsilon = 2$.

i	$\alpha_\gamma = \alpha_\varepsilon = 1$				$\alpha_\gamma = 1$ and $\alpha_\varepsilon = 2$			
	a_i^{GJ}	d_i^{GJ}	w_i^{GJ}	$\widetilde{w}_i^{\mathrm{GJ}}$	a_i^{GJ}	d_i^{GJ}	w_i^{GJ}	$\widetilde{w}_i^{\mathrm{GJ}}$
1	-0.9816	0.0093	0.0472	0.0236	-0.9829	0.0086	0.0867	0.0433
2	-0.9041	0.0504	0.1069	0.0535	-0.9111	0.0465	0.1897	0.0949
3	-0.7699	0.1300	0.1601	0.0800	-0.7863	0.1196	0.2664	0.1332
4	-0.5873	0.2600	0.2032	0.1016	-0.6157	0.2379	0.3076	0.1538
5	-0.3678	0.4622	0.2335	0.1167	-0.4092	0.4192	0.3108	0.1554
6	-0.1252	0.7774	0.2491	0.1246	-0.1789	0.6965	0.2804	0.1402
7	0.1252	1.2863	0.2491	0.1246	0.0619	1.1320	0.2265	0.1132
8	0.3678	2.1637	0.2335	0.1167	0.2992	1.8539	0.1618	0.0809
9	0.5873	3.8463	0.2032	0.1016	0.5192	3.1597	0.0995	0.0498
10	0.7699	7.6920	0.1601	0.0800	0.7091	5.8753	0.0497	0.0249
11	0.9041	19.8588	0.1069	0.0535	0.8579	13.0730	0.0178	0.0089
12	0.9816	107.4636	0.0472	0.0236	0.9596	45.3778	0.0031	0.0016