UNIVERSITY OF SOUTHAMPTON

FACULTY OF SOCIAL, HUMAN, AND MATHEMATICAL SCIENCES

Mathematical Sciences

Frequentist Decision-theoretic Optimal Design and its Applications to Generalised Linear Models

by

Meshayil M. Alsolmi

Thesis submitted for the degree of Doctor of Philosophy

March 16, 2020

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF SOCIAL, HUMAN, AND MATHEMATICAL SCIENCES
Mathematical Sciences

Doctor of Philosophy

FREQUENTIST DECISION-THEORETIC OPTIMAL DESIGN AND ITS APPLICATIONS TO GENERALISED LINEAR MODELS

by Meshayil M. Alsolmi

Alphabetical optimal designs are found by minimising a scalar function of the inverse Fisher information matrix, and represent the de-facto standard in optimal design of experiments. For example, the well-known D-optimal design is found by minimising the log determinant of the inverse Fisher information matrix. In this thesis, an alternative decision-theoretic basis for frequentist design is proposed and investigated whereby designs are found by minimising the risk function defined as the expectation of an appropriate loss function that represents the aim of an experiment (e.g, parameter estimation or model discrimination). The expectation is taken with respect to the distribution of responses that observed from the experiment given the design. Generally, the loss function compares a parameter estimator or result of a model selection procedure to the true values of the parameters or the model, respectively, given the observed data obtained from the design. The conceptual advantages of the decision-theoretic framework over alphabetical optimal designs are that it is suitable for small sample sizes and can be applied for bespoke experimental aims. This research aims to find frequentist decision-theoretic optimal designs for the experimental aim of parameter estimation and model discrimination for generalised linear models.

Finding decision-theoretic designs is accomplished by minimising the risk function over a design space $\mathcal{X}$. The risk function that we minimise depends on unknown parameters. Local and pseudo-Bayesian approaches are proposed for defining an objective function in cases where the risk depends on the parameters. However, finding such designs is complicated due to the considerable computational challenges of minimising an analytically intractable risk function, leading to the need for approximation methods for the risk function. We develop novel methods to approximate the risk function, and these approximations are then used within an optimisation algorithm to find decision-theoretic optimal designs for a range of examples that have important applications in the design of experiments. Comparison between these approximations will be considered in terms of performance and computing time.

Contents

List of Figures

List of Tables

Declaration of Authorship

I, Meshayil M. Alsolmi , declare that the thesis entitled *Frequentist Decision-theoretic Optimal Design and its Applications to Generalised Linear Models* and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- none of this work has been published before submission.

Signed:.....

Date:.....

Acknowledgements

Foremost, all my praise and thanks go to God for the strengths and abilities he has given me to complete this thesis. Special appreciation goes to my supervisor, Dr. Antony Overstall, for his supervision, particularly his time and his ideas, advice and assistance. My appreciation also goes to my co-supervisor, Prof. Dave Woods, for his support and advice throughout my PhD studies.

I'm thankful to the University of Jeddah and Saudi Arabian Cultural Bureau in the UK for the funding opportunities and their continuous support during my academic journey. I am also grateful to all the members of the Design Group at the University of Southampton who have been a good source of information that helped me build my knowledge in the field of experimental design.

Finally and most importantly, I would like to thank my family for the love and inspiration, particularly my husband, Faris Alsolami, for his unwavering support.

Chapter 1

Introduction

1.1 Design of experiments and optimal experimental design

The statistical design of experiments is essential for experimenters in various branches of science, including physical and social. It is one of the most important topics in the field of statistics due to the role it plays in improving scientific research, where an experiment is an efficient way of learning about the world. There are a number of issues to be resolved in conducting an experiment such as which treatments to study and control, how many experimental units we need to study, how many experimental runs we should allocate to each treatment, all of which fall under the concept of statistical design (Chaloner and Verdinelli, 1995). An effective experimental design is important to present accurate and valid answers to the research questions that the experimenter has identified (Atkinson et al., 2007). In this thesis, we will have a function $\Psi(d)$ which measures a particular design's effectiveness in achieving its aim. Thus the optimal design is given by optimising Ψ over the space of all possible designs. In Section 1.2 we review the standard approach to optimal design: alphabetical optimality.

1.2 Alphabetical optimal designs

Minimising the variability of the estimators aids to improve the statistical inference about the quantities of interest that require estimation and contain the best explanation and conclusion about a set of data. The precision of the estimators is quantified by the Fisher information matrix which is a large sample approximation to the inverse variance matrix of the maximum likelihood estimators (MLEs). The Fisher information matrix measures all information regarding the unknown parameters in the observations (Casella and Berger, 2002, Chapter 7).

An alphabetical optimal design is a standard approach in the design of experiments and was introduced by Kiefer (1959). Under this approach, the optimal design d^* is defined by comparison with the set of all possible designs with respect to a particular criterion. The criterion for design is given by an objective function reflecting the aim of the experiment and is to be minimised or maximised for the design to be optimal within the set of all possible designs. The objective function is defined as a scalar function of the inverse Fisher information matrix for parameters under a specified statistical model. Different scalar functions are referred to using different letters from the alphabet, and collectively, these designs are known as alphabetical optimal (Atkinson et al., 1993; Pukelsheim, 2006). A D-optimal design is a well-known example and is found by minimising the log determinant of the inverse Fisher information matrix, where the scalar function is given by the log determinant. Further details regarding design criteria for choosing the optimal design under frequentist alphabetical optimal designs and their mathematical formulas are given in Chapter 2.

1.3 Decision-theoretic optimal designs

An alternative decision-theoretic basis for frequentist design is proposed and investigated in this thesis. Frequentist decision-theoretic optimal designs are found by minimising a risk function defined as the expectation of an appropriate loss function representing the aim of an experiment (e.g. parameter estimation, prediction or hypothesis testing). The expectation is taken with respect to the distribution of the data given by the statistical model (or models), given quantities of interest (e.g. the parameters that require estimation).

To illustrate, suppose the aim of the experiment is to estimate the parameter by an estimator. Whether the estimator is a good choice depends on the quantity of the loss for estimating the parameter by the estimator. Hence, the loss specifies the loss of estimating the parameters with the estimators. The choice of the loss function depends on the experimental aims. The squared error loss (SE) is an example for the loss function, and used under the experimental aim of parameter estimation. The risk function under this particular loss is known as the expected SE or mean square error (MSE).

The decision-theoretic approach is more suitable for small sample sizes, since it does not rely on the large sample approximation to the variance matrix given by the inverse Fisher information matrix. In addition, the decision-theoretic approach allows the consideration of bespoke experimental aims which are not satisfied by the often default use of alphabetical optimal criteria. For example, in clinical trials, when the aim of the experiment is to investigate the probability of the toxicity of a drug with respect to

increasing dose. In this experiment, the sign loss (SL) is an appropriate choice as a loss function to identify when the probability of the toxicity increases or decreases, based on a chosen design, where this loss concerns the sign of the quantities of interest. These are the advantages of using decision-theoretic optimal designs over alphabetical optimal designs. Further details regarding the loss function and examples are given in Chapter 3.

1.4 Problem statement

There are three hurdles to finding frequentist decision-theoretic optimal designs, and these are:

- The risk function is often analytically intractable and of high dimension. This hinders finding exact frequentist decision-theoretic optimal designs, leading to the need for approximation methods for the risk function.
- The risk function depends on unknown parameters leading to difficulties in finding frequentist decision-theoretic optimal designs. Local and pseudo-Bayesian approaches are proposed for defining an objective function in cases where the risk depends on the parameters.
- The objective function will need to be optimised over a high-dimensional design space.

1.5 Research contributions

The contributions to this research are

1. Extension of the second-order Taylor series (ST) approximation to the risk function that is proposed for non-linear models (Pronzato and Pázman, 2013), to other types of models including generalised linear models (GLMs).
2. Development of novel methods to approximate the risk function for generalised linear models, including quasi-Monte Carlo (QMC), quasi-importance sampling (QIMPS) and quadrature (Quad). These are compared to an asymptotic (Asym) approximation and the exact risk (where possible). Comparison between these approximations will be considered in terms of performance and computing time.
3. Application of local and pseudo-Bayesian approaches to overcome the dependence of the risk on the parameters. Under the pseudo-Bayesian approach, a quadrature rule is applied to approximate the prior expectation of the risk.

4. Application of the ACE algorithm for finding designs by optimising the approximate objective function.

5. Extension of decision-theoretic approach to model selection for generalised linear models.

1.6 Thesis organisation

In Chapter 2, we review the basic concepts of the design of experiments and some essential concepts in estimation theory. Properties of choosing estimators are illustrated, and a detailed description of the most widely applicable approach to finding estimators is introduced, namely maximum likelihood estimation. Under GLMs, maximum likelihood estimates may be infinite or non-unique for certain observations. Therefore, we consider the maximum penalised likelihood estimation as an alternative estimation method to estimate the parameters. Alphabetic optimality, with three popular examples of D-, A- and E-optimality, is described. An introduction to GLMs and examples that have important applications in the design of experiments are given. Last, we review two commonly used approaches to address the challenges in alphabetical optimal design for GLMs, namely local and pseudo-Bayesian optimality.

The framework of frequentist decision-theoretic optimal designs is introduced in Chapter 3, and some examples of loss functions tailored to the experimental aim of parameter estimation are provided. The objective function used for finding such designs under normal linear models is introduced and illustrated with examples of loss functions given in this chapter. Finding decision-theoretic optimal designs under GLMs is complicated by the need to numerically approximate an analytically intractable objective function. Therefore, we introduce quadrature approximation to the Bayesian risk, which is used to find the optimal design under the pseudo-Bayesian approach. In addition, some approximations to the intractable risk function are proposed in Chapter 4, namely asymptotic, quadrature, quasi-Monte Carlo and quasi-importance sampling. Asymptotic approximation is illustrated under the considered loss function given in the previous chapter. An extension of the second-order Taylor series (ST) approximation to the risk function proposed for non-linear models to other types of models including GLMs is derived.

The asymptotic normality of maximum likelihood estimators (MLEs) and maximum penalised likelihood estimators (MPLEs) is investigated in Chapter 5 under four examples of GLMS. Decision-theoretic optimal designs with the aim of estimating unknown parameters under squared error matrix and sign losses are found in Chapter 6. A model discrimination procedure that involves a model selection criterion to choose the best model from a set of candidate models is introduced in Chapter 7. Furthermore,

decision-theoretic optimal designs with the aim of determining the appropriate model of a set of candidate models under 0-1 and factor specific losses are found for the logistic regression model. In the final chapter, a summary of this thesis and an outline of future work are provided.

Chapter 2

Background and Literature Review

2.1 Experiment

This section introduces the basic concepts and preliminaries of experimentation, and these are taken from Dean and Voss (2008) and Ryan and Morgan (2007). Initially, an experiment is the way through which data is collected to answer pre-specified scientific questions.

An experiment measures an outcome (the response variable) after making changes in the settings of controllable factors, or the level of factors, which are assumed to cause changes in the response variable. Design of experiments is concerned with the selection of settings of factors in an experiment to increase the capability or effectiveness of the experiment in attaining its aims. A designed experiment is different from an observational study in that a researcher has control over the settings of the factors. For example, consider an experiment to compare two drugs, the experimenter has control to assign participants one of the two drugs to gain the most accurate inference about the difference in mean response from the two drugs.

To formalise, the response variable (Y) is the measured outcome in an experiment and made under the settings of k controllable factors, $\mathbf{z} = (z_1, \ldots, z_k)^T$, $\mathbf{z} \in \chi \subset \mathbb{R}^k$. The N responses, $Y_1, ..., Y_N$, are from N runs. For the ith run, the jth factor is set as z_{ij}, for $i = 1, ..., N$, and $j = 1, ..., k$, and these are called the factor levels. Optimal design is concerned with choosing the design points $d = \{\mathbf{z}_1, \ldots, \mathbf{z}_N\}$ to best attain the experimental aim, where $\mathbf{z}_i = (z_{i1}, \ldots, z_{ik})$. Support points or treatments are defined as the distinct design points.

2.2 Estimation methods

To answer a scientific question or achieve the experimental goals, we begin with a statistical model. The statistical model is a type of mathematical model, used to approximately represent the process by which the experimental data is generated. Let $\mathbf{y} = (y_1, ..., y_N)$ be the responses where $\mathbf{y}$ is the realisation of the vector of random variables $\mathbf{Y} = (Y_1, ..., Y_N)$. It is assumed that $\mathbf{Y}$ has a probability distribution given by $f(\mathbf{Y}; \boldsymbol{\theta}, d)$; a probability mass function (pmf) if $\mathbf{Y}$ is a discrete variable or a probability density function (pdf) if $\mathbf{Y}$ is a continuous variable. The probability distribution is completely specified except for a $p \times 1$ vector of parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^T \in \Theta$, where Θ denotes the parameter space.

In statistical inference, the best explanation and conclusion about a set of data is given by the values of unknown parameters $\boldsymbol{\theta}$. Some concepts associated with the aim of estimating $\boldsymbol{\theta}$ are now briefly outlined (see Garthwaite et al., 2002, Chapter 2 and Casella and Berger, 2002, Chapter 7 for more details). Let $t(\mathbf{y})$ be a function of $\mathbf{y}$ only, then $t(\mathbf{Y})$ is known as a statistic. If $\hat{\boldsymbol{\theta}}(\mathbf{y}) = (\hat{\theta}_1(\mathbf{y}), \ldots, \hat{\theta}_p(\mathbf{y}))$ is used to estimate $\boldsymbol{\theta}$ then it is known as an estimate. The corresponding statistic $\hat{\boldsymbol{\theta}}(\mathbf{Y})$ is known as an estimator. There are various possible estimators for any situation, and it is necessary to choose between them. The suitability of an estimator is determined based on the properties of its probability or sampling distribution, notably its expectation and variance. When the estimator's expectation is equal to the true value of $\boldsymbol{\theta}$, it is then an unbiased estimator. Mathematically,

$$E_{\mathbf{Y}|\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}(\mathbf{Y})) = \boldsymbol{\theta},$$

where $E_{\mathbf{Y}|\boldsymbol{\theta}}$ is with respect to the distribution of responses given by $f(\mathbf{Y}; \boldsymbol{\theta}, d)$. In addition, if an unbiased estimator has the smallest possible variance, it is the minimum variance unbiased estimator. The minimum variance for all possible unbiased estimators is given by the Cramér–Rao lower bound (CRLB). The bound is also known as the Cramér–Rao inequality, and it states that the variance of any unbiased estimator $\hat{\theta}_i(\mathbf{Y})$, the ith element of the estimators, $i = 1, ..., p$, under certain regularity conditions (Garthwaite et al., 2002, Chapter 2, page 11) on $f(\mathbf{Y}; \boldsymbol{\theta}, d)$ should be at least equal to $[I(\boldsymbol{\theta}; \mathbf{d})^{-1}]_{ii}$, i.e.

$$Var_{\mathbf{Y}|\boldsymbol{\theta}}\left[\hat{\theta}_i(\mathbf{Y})\right] \geq [I(\boldsymbol{\theta}; d)^{-1}]_{ii},$$

where $I(\boldsymbol{\theta}; d)$ is the Fisher information matrix defined as

$$I_{ij}(\boldsymbol{\theta}; d) = E_{\mathbf{Y}|\boldsymbol{\theta}}\left[\frac{\partial \log f(\mathbf{Y}; \boldsymbol{\theta}, d)}{\partial \theta_i} \frac{\partial \log f(\mathbf{Y}; \boldsymbol{\theta}, d)}{\partial \theta_j}\right], \tag{2.1}$$

for $i, j = 1, \ldots, p$. The $E_{\mathbf{Y}|\boldsymbol{\theta}}$ is with respect to the distribution of $\mathbf{Y}$ given by $f(\mathbf{Y}; \boldsymbol{\theta}, d)$. The MSE of the estimator is also a criterion that we may consider in choosing the

estimator, where it incorporates both the estimator's variance and bias. The MSE is defined as

$$R_{SE}(d;\boldsymbol{\theta}) = \mathrm{E}_{\mathbf{Y}|\boldsymbol{\theta}}\left[(\hat{\boldsymbol{\theta}}(\mathbf{Y}) - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}}(\mathbf{Y}) - \boldsymbol{\theta})^T\right]. \tag{2.2}$$

It can be written as

$$R_{SE}(d;\boldsymbol{\theta}) = Var_{\mathbf{Y}|\boldsymbol{\theta}}\left[\hat{\boldsymbol{\theta}}(\mathbf{Y})\right] + b(\hat{\boldsymbol{\theta}}(\mathbf{Y}))b^T(\hat{\boldsymbol{\theta}}(\mathbf{Y})), \tag{2.3}$$

where $Var_{\mathbf{Y}|\boldsymbol{\theta}}\left[\hat{\boldsymbol{\theta}}(\mathbf{Y})\right]$ denotes the variance matrix and $b(\hat{\boldsymbol{\theta}}(\mathbf{Y}))$ corresponds to the bias of the estimator, $b(\hat{\boldsymbol{\theta}}(\mathbf{Y})) = \mathrm{E}_{\mathbf{Y}|\boldsymbol{\theta}}\left[\hat{\boldsymbol{\theta}}(\mathbf{Y})\right] - \boldsymbol{\theta}$. For unbiased estimators,

$$R_{SE}(d;\boldsymbol{\theta}) = Var_{\mathbf{Y}|\boldsymbol{\theta}}\left[\hat{\boldsymbol{\theta}}(\mathbf{Y})\right].$$

2.2.1 Maximum likelihood estimation

There are several approaches to finding estimators. These include maximum likelihood (ML) estimation, method of moments, method of least squares, method of minimum χ^2, and estimating equations (Garthwaite et al., 2002). However, ML estimation is the most widely applicable and has many desirable properties (Garthwaite et al., 2002, Chapter 3, page 49) as the sample size increases to ∞. The properties of the maximum likelihood estimator (MLE) are

- *Asymptotically unbiased.* The bias of the MLE tends to zero as the sample size N tends to ∞.
- *Asymptotically consistent.* When the sample size N moves toward ∞, the variance and bias of the MLE tend to zero.
- *Asymptotically normally distributed.* The MLE has a sampling distribution. When the sample size increases, the MLE tends to a Normal distribution, i.e.

$$\hat{\boldsymbol{\theta}}(\mathbf{Y}) \sim N(\boldsymbol{\theta}, I(\boldsymbol{\theta};d)^{-1}).$$

- *Asymptotically Efficient.* This means that MLE has lower asymptotic mean squared error than any other estimator, as it achieves the CRLB.

In order to define the MLE, the likelihood function should be first defined. The likelihood function corresponds to the joint probability density function of $\mathbf{Y}$ if continuous or the probability mass function if discrete, i.e.

$$L(\boldsymbol{\theta};\mathbf{Y},d) = f(\mathbf{Y};\boldsymbol{\theta},d).$$

The MLE is defined as the values of $\boldsymbol{\theta}$ that maximise the likelihood function $L(\boldsymbol{\theta}; \mathbf{Y}, d)$. Mathematically,

$$\hat{\boldsymbol{\theta}}(\mathbf{Y}) = \arg\max_{\boldsymbol{\theta}} L(\boldsymbol{\theta}; \mathbf{Y}, d).$$

Typically, it is easier to use the log-likelihood function rather than $L(\boldsymbol{\theta}; \mathbf{Y}, d)$, i.e.

$$l(\boldsymbol{\theta}; \mathbf{Y}) = \log L(\boldsymbol{\theta}; \mathbf{Y}, d).$$

The log-likelihood has the same maximum as $L(\boldsymbol{\theta}; \mathbf{Y}, d)$, since log is a convex function (Garthwaite et al., 2002). The score function is defined by

$$U(\boldsymbol{\theta}; \mathbf{Y}, d) = \begin{pmatrix} \frac{\partial l(\boldsymbol{\theta}; \mathbf{Y}, d)}{\partial \theta_1} \\ \vdots \\ \frac{\partial l(\boldsymbol{\theta}; \mathbf{Y}, d)}{\partial \theta_p} \end{pmatrix}.$$

The MLEs are obtained by solving

$$U(\boldsymbol{\theta}; \mathbf{Y}, d) = \mathbf{0}, \tag{2.4}$$

with respect to $\boldsymbol{\theta}$.

Finding the maximum of $L(\boldsymbol{\theta}; \mathbf{Y}, d)$ is an optimisation problem, and iterative procedures are often required to solve this problem. This is because the solution to (2.4) is typically analytically intractable. There are several useful optimisation methods for computing MLEs, such as the Newton-Raphson technique, Fisher scoring, the simplex method, and the EM algorithm (Garthwaite et al., 2002).

2.2.2 Maximum penalised likelihood estimation

The log-likelihood can be modified to incorporate a subsidiary criterion, and with the resulting penalised log-likelihood written as

$$l^*(\boldsymbol{\theta}; \mathbf{Y}, d) = l(\boldsymbol{\theta}; \mathbf{Y}, d) - p(\boldsymbol{\theta}), \tag{2.5}$$

where $p(\boldsymbol{\theta})$ is the penalty function. The context of penalised likelihood is often found in smoothing a curve or density function, where $p(\boldsymbol{\theta})$ gives the measure of roughness (Garthwaite et al., 2002, Chapter 3, page 58). The penalty function $p(\boldsymbol{\theta})$ can also be used to represent prior information, where $\exp(-p(\boldsymbol{\theta}))$ can be considered as being proportional to a prior distribution for $\boldsymbol{\theta}$. Therefore, maximising $l^*(\boldsymbol{\theta}; \mathbf{Y}, d)$ is equivalent to maximising the posterior density of $\boldsymbol{\theta}$. The values of $\hat{\boldsymbol{\theta}}^*(\mathbf{Y})$ maximises (2.5), and is

called the maximum penalised likelihood estimator (MPLE). Mathematically,

$$\hat{\boldsymbol{\theta}}^*(\mathbf{Y}) = \arg\max_{\boldsymbol{\theta}} l^*(\boldsymbol{\theta}; \mathbf{Y}, d).$$

The MPLEs share the same asymptotic properties of MLEs, and

$$\hat{\boldsymbol{\theta}}^*(\mathbf{Y}) \to \hat{\boldsymbol{\theta}}(\mathbf{Y}),$$

as $N \to \infty$. The penalised score function is defined

$$U^*(\boldsymbol{\theta}; \mathbf{Y}, d) = \begin{pmatrix} \frac{\partial l^*(\boldsymbol{\theta}; \mathbf{Y}, d)}{\partial \theta_1} \\ \vdots \\ \frac{\partial l^*(\boldsymbol{\theta}; \mathbf{Y}, d)}{\partial \theta_p} \end{pmatrix}.$$

The MPLE is obtained by solving

$$U^*(\boldsymbol{\theta}; \mathbf{Y}, d) = \mathbf{0},$$

with respect to $\boldsymbol{\theta}$ using iterative procedures. This is an alternative estimation method, used in this thesis, to estimate the parameters under generalised linear models. The reason for using this approach rather than maximum likelihood estimation is that under some models and for certain observations, the maximum likelihood estimates may be infinite or non-unique. In addition, the estimates might not be asymptotically efficient (Garthwaite et al., 2002, Chapter 3, page 52). This means that the estimates do not achieve the CRLB, and the log-likelihood in this case might be flat or has mutli-modal. Penalising the log-likelihood function is an approach to resolve this problem and resulting in bias and variance reduction of maximum likelihood estimates for some statistical models (Russell et al., 2009a).

2.3 Introduction to generalised linear models (GLMs)

Generalised Linear Models are a generalisation of linear models to response types that have a distribution other than normal (Nelder and Baker, 1972; Dobson and Barnett, 2008; Dunteman and Ho, 2005). They are widely-used in practice (McGree and Eccleston, 2012; Russell et al., 2009b; Wang et al., 2006; Myers et al., 1994). They can also be used for analysing a variety of data that can not be described by the linear model. For example, the range of $\mathbf{Y}$ is restricted (e.g., binary, count) and the variance of $\mathbf{Y}$ depends on the mean $\boldsymbol{\mu}$.

GLMs consist of three components:

i *Distribution of the response*: it is a random component, and refers to the distribution of the response Y_i, where the ith element of Y is independent and has a distribution given by the probability distribution $f(Y_i; \boldsymbol{\theta})$ from the exponential family of distributions. The exponential family has many popular distributions such as normal, Poisson, Bernoulli, binomial, and gamma. It (McCullagh, 1984; Atkinson et al., 2012) is a wide class of distributions that all share the following form:

$$f(Y_i; \phi_i, \gamma) = \exp\left[(Y_i\phi_i - a(\phi_i))/\gamma + b(Y_i, \gamma)\right], \tag{2.6}$$

where ϕ_i is the natural or canonical parameter and γ, which is constant for all $i = 1, \ldots, N$, is the scale parameter. The $a(\phi_i)$ and $b(Y_i, \gamma)$ are known functions. This is called the canonical form or standard form of the exponential family of distributions. The mean and variance of the responses $\mathbf{Y}$ can be derived, i.e.

$$E(Y_i) = \mu_i = a'(\phi_i) \quad \text{and} \quad Var(Y_i) = \gamma a''(\phi_i) = \gamma V(\mu_i), \tag{2.7}$$

respectively; where $V(\mu_i)$ is the variance function.

ii *Linear predictor*: it forms the systematic part of the model and is defined as

$$\boldsymbol{\eta} = X\boldsymbol{\theta}, \tag{2.8}$$

where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^T$ is a $p \times 1$ vector of unknown parameters, X is an $N \times p$ matrix of known function h of the k controllable variables (i.e. $h(\mathbf{z}) = (1, \mathbf{z})$ gives an intercept and first-order terms) and $\boldsymbol{\eta}$ is an $N \times 1$ vector, where $\eta_i = \mathbf{x}_i^T\boldsymbol{\theta}$.

iii *Link function*: the mean response $E(Y_i) = \mu_i$ is related to the linear predictor through the link function $g(.)$, i.e.

$$g(\mu_i) = \eta_i = \mathbf{x}_i^T\boldsymbol{\theta},$$

where the choice of the link function depends on the distribution of the response. If the link function connects the linear predictors $\boldsymbol{\eta}$ and natural parameters $\boldsymbol{\phi}$ via the mean $\boldsymbol{\mu}$, this link is then called canonical and has the form $g(\mu_i) = (a(\mu_i)')^{-1}$. Hence,

$$g(\mu_i) = (a(\mu_i)')^{-1} = \phi_i = \eta_i.$$

Using the canonical link leads to desirable properties for GLMs; thus, they are used by default. These properties are that (I) the Newton Raphson method, for finding the MLEs, coincides with the Fisher scoring method, (II) the derivation of the MLEs is simplified by using these canonical links, and (III) some properties of the linear regression model (e.g., the sum of the residuals is 0) are ensured to hold or $\boldsymbol{\mu}$ are ensured to lie within the range of $\mathbf{Y}$.

The score function using the canonical link (McCullagh, 1984) is given by

$$U(\boldsymbol{\theta}, \mathbf{Y}, d) = X^T(\mathbf{Y} - \boldsymbol{\mu}). \tag{2.9}$$

The MLEs $\hat{\boldsymbol{\theta}}(\mathbf{Y})$ are then obtained by solving

$$U(\boldsymbol{\theta}, \mathbf{Y}, d) = \mathbf{0},$$

with respect to $\boldsymbol{\theta}$ using the iterative weighted least squares (IWLS) algorithm. The approximate distribution of $\hat{\boldsymbol{\theta}}(\mathbf{Y})$, under GLMs, is given by

$$\hat{\boldsymbol{\theta}}(\mathbf{Y}) \sim N(\boldsymbol{\theta}, I(\boldsymbol{\theta}; d)^{-1}).$$

The Fisher information is

$$I(\boldsymbol{\theta}; d) = X^T W X, \tag{2.10}$$

where W is an $N \times N$ diagonal matrix, with elements

$$w_{ii} = \frac{1}{Var_{\mathbf{Y}|\boldsymbol{\theta}}[Y_i]} \left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2, \tag{2.11}$$

where the weights depend on both the distribution of $\mathbf{Y}$ and the link function.

As mentioned previously, maximum likelihood estimates for certain observations can be infinite. This is commonly found under GLMs (Firth, 1993), particularity in logistic regression for a binary response. Firth (1993) demonstrates that the maximum penalised likelihood estimators, with Jeffreys prior as the penalty function, in a GLM with a canonical link function, have reduced bias and variance over the MLEs. The penalty, Jeffreys prior for $\boldsymbol{\theta}$, is equal to minus one half of the log of the determinant of the Fisher information matrix (Firth, 1993; Russell et al., 2009a), which is given by $|I(\boldsymbol{\theta}, d)|^{-1/2}$. The penalised log likelihood function with Jeffreys prior as a penalty function can be written as

$$l^*(\boldsymbol{\theta}; \mathbf{Y}, d) = l(\boldsymbol{\theta}; \mathbf{Y}, d) + \frac{1}{2} \log |I(\boldsymbol{\theta}; d)|. \tag{2.12}$$

This can be applied to log-linear models for count data, reciprocal linear models for gamma-distributed data, and logistic regression models for binary data (Section 2.3.2). The penalised score function (Firth, 1992) can be then written as

$$U^*(\boldsymbol{\theta}, \mathbf{Y}, d) = U(\boldsymbol{\theta}, \mathbf{Y}, d) + \frac{1}{2} \frac{\partial}{\partial \boldsymbol{\theta}} \log |I(\boldsymbol{\theta}; d)|. \tag{2.13}$$

Thus, the MPLEs can be obtained (Firth, 1992) by solving

$$U^*(\boldsymbol{\theta}, \mathbf{Y}, d) = \mathbf{0},$$

with respect to $\boldsymbol{\theta}$ using the adjusted IWLS algorithm.

Some examples of GLMs are introduced in the forthcoming sections.

2.3.1 Linear models

Linear models are a special case of GLMs. It is assumed

$$Y_i \sim N(\mu_i, \sigma^2),$$

and the canonical link is the identity link, i.e. $g(\mu_i) = \mu_i$. Hence, the linear predictor

$$\eta_i = \mathbf{x}_i^T \boldsymbol{\theta} = h(\mathbf{z}_i)\boldsymbol{\theta} = \rho_0 + \sum_{j=1}^{k} \theta_j z_{ij}$$

is linked to the mean of the response as follows

$$\mu_i = \rho_0 + \sum_{j=1}^{k} \theta_j z_{ij}.$$

Under a linear model, the MLE has a closed form solution given by

$$\hat{\boldsymbol{\theta}}(\mathbf{Y}) = (X^T X)^{-1} X^T \mathbf{Y}. \tag{2.14}$$

The Fisher information matrix is given by

$$I(\boldsymbol{\theta}; d) = \frac{1}{\sigma^2} X^T X. \tag{2.15}$$

The expectation and variance of $\hat{\boldsymbol{\theta}}(\mathbf{Y})$ are

$$\begin{aligned} E\left(\hat{\boldsymbol{\theta}}(\mathbf{Y})\right) &= \boldsymbol{\theta}, \\ Var\left(\hat{\boldsymbol{\theta}}(\mathbf{Y})\right) &= \sigma^2 (X^T X)^{-1}, \end{aligned}$$

i.e. the asymptotic results of Section 2.2.1 are exact.

2.3.2 Logistic regression models

The logistic regression model is an example of a GLM, and widely-studied (see McCullagh, 1984, Russell et al., 2009a and references therein). It is used for modelling the observations of a categorical response variable, where the response variable is the number of successes observed in a pre-specified number of trials. It is assumed

$$Y_i \sim Bin(n_i, \pi_i),$$

with

$$\pi_i = 1/[1 + \exp(-\eta_i)],$$

where π_i refers to the probability of a success, and n_i corresponds to the number of trials. For the binomial distribution, the logit or logistic link (the natural logarithm of the odds of the probability of a success) is the canonical link

$$g(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right).$$

Hence, the linear predictor

$$\eta_i = \mathbf{x}_i^T\boldsymbol{\theta} = h(\mathbf{z}_i)\boldsymbol{\theta} = \rho_0 + \sum_{j=1}^{k}\theta_j z_{ij}$$

is linked to the mean of the response through the logit function, i.e.

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \rho_0 + \sum_{j=1}^{k}\theta_j z_{ij}, \tag{2.16}$$

where there are $p = k + 1$ unknown parameters that require estimation, $\boldsymbol{\theta} = (\rho_0, \theta_1, \ldots, \theta_k)^T$, and the quantity ρ_0 refers to the intercept term.

The simple linear logistic regression model is a special case of a logistic regression model with $k = 1$ and $N = 2$. In this case, the MLEs of ρ_0 and θ_1 have a closed form, and are given by

$$\hat{\rho}_0(\mathbf{Y}) = \frac{T_1 z_2 - T_2 z_1}{z_1 - z_2} \qquad \text{and} \qquad \hat{\theta}_1(\mathbf{Y}) = \frac{T_2 - T_1}{z_1 - z_2}, \tag{2.17}$$

where $T_i = \log\left[(n_i - Y_i)/(Y_i)\right]$, $i = 1, 2$.

Using the MPL estimation illustrated in Section 2.2.2, with Jeffreys prior as the penalty function, results in estimators with reduced bias compared with the MLEs (Russell et al., 2009a). Hence, the MPLEs are given by

$$\hat{\rho}_0^*(\mathbf{Y}) = \frac{T_1^* z_2 - T_2^* z_1}{z_1 - z_2} \qquad \text{and} \qquad \hat{\theta}_1^*(\mathbf{Y}) = \frac{T_2^* - T_1^*}{z_1 - z_2}, \tag{2.18}$$

where $T_i^* = \log\left[(n_i - Y_i + 0.5)/(Y_i + 0.5)\right]$, $i = 1, 2$. Here, the MPLEs are shrinkage estimators and have smaller variance than MLEs. When $Y_i = 0$ or $Y_i = n_i$, for either $i = 1$ or $i = 2$, the MLEs have infinite estimates of ρ_0 and θ_1, see equation (2.17). However, the MPL estimation cannot give infinite estimates, see equation (2.18), which provides it with an additional property.

The simple linear logistic regression model and the $k = 4$ factors logistic regression

model (with $n_i = 1$) are considered in Chapter 6 to investigate the decision-theoretic optimal designs for the experimental aim of parameter estimation.

2.3.3 Poisson regression model

The Poisson regression model can be written as

$$Y_i \sim \text{Poisson}(\mu_i), \tag{2.19}$$

for $i = 1, \ldots, N$. For count data (Poisson responses), the log link function is the canonical link where

$$g(\mu_i) = \log(\mu_i).$$

Hence, the linear predictor

$$\eta_i = \mathbf{x}_i^T \boldsymbol{\theta} = h(\mathbf{z}_i)\boldsymbol{\theta} = \rho_0 + \sum_{j=1}^{k} \theta_j z_{ij},$$

is linked to the mean of the response through the natural logarithm function, i.e

$$\log(\mu_i) = \rho_0 + \sum_{j=1}^{k} \theta_j z_{ij}, \tag{2.20}$$

where there are $p = k + 1$ unknown parameters that require estimation, $\boldsymbol{\theta} = (\rho_0, \theta_1, \ldots, \theta_k)^T$. The Poisson regression model with $k = 2$ factors is considered in Chapter 6 for investigating the decision-theoretic optimal designs for the experimental aim of parameter estimation.

2.4 Alphabetic optimality

It is known that finding frequentist alphabetical optimal designs is accomplished by minimising an objective function $\psi(d; \boldsymbol{\theta})$, given by a scalar function S of the inverse Fisher information matrix $I(\boldsymbol{\theta}; d)^{-1}$. In Chapter 1, we gave a brief introduction to the frequentist alphabetical approach to design that is used as design criteria for finding optimal designs. There are many different examples of alphabet optimality criteria, see, for example, Goos (2012, page 13), Atkinson et al. (2007, page 135) and Silvey (2013, page 10). In this chapter, we review three popular examples of alphabetical optimal designs, and these are D-, A- and E- optimal designs.

- D-optimal. A D-optimal design will minimise the volume of an approximate confidence ellipsoid for the parameters (Chernoff, 1953; Myers et al., 1994; Atkinson, 1996; Atkinson et al., 2007). Here the objective function is

$$\psi_D(d; \boldsymbol{\theta}) = \log |I(\boldsymbol{\theta}; d)^{-1}|, \tag{2.21}$$

i.e. $S(I(\boldsymbol{\theta}; d)^{-1}) = \log |I(\boldsymbol{\theta}; d)^{-1}|$.

- A-optimal. An A-optimal design minimises the average of the variances of the parameter estimators (Atkinson et al., 2007, page 135) . Here the objective function is

$$\psi_A(d; \boldsymbol{\theta}) = \text{tr}\{I(\boldsymbol{\theta}; d)^{-1}\}, \tag{2.22}$$

  i.e. $S(I(\boldsymbol{\theta}; d)^{-1}) = \text{tr}\{I(\boldsymbol{\theta}; d)^{-1}\}$.

- E-optimal. An E-optimal design minimises the variance of the least well-estimated contrast $\boldsymbol{a}^T \boldsymbol{\theta}$ which is subject to the constraint $\boldsymbol{a}^T \boldsymbol{a} = 1$ (Pukelsheim, 2006). Here the objective function is

$$\psi_E(d; \boldsymbol{\theta}) = \lambda_{\max}\{I(\boldsymbol{\theta}; d)^{-1}\}, \tag{2.23}$$

  i.e. $S(I(\boldsymbol{\theta}; d)^{-1}) = \lambda_{\max}\{I(\boldsymbol{\theta}; d)^{-1}\}$. The expression $\lambda_{max}(A)$ in (2.23) refers to the maximum eigenvalue of A.

2.4.1 Alphabetical optimal designs for linear models

Finding optimal design for the class of linear models is straightforward. This is because the inverse of the Fisher information matrix is independent of $\boldsymbol{\theta}$, see equation (2.15). However, linear models often contain more parameters than nonlinear models for fitting the same data. Hence, increasing the number of parameters in the model may also lead to an increase in the experimental runs (Atkinson et al., 2007, page 36).

There is an extensive study in the literature concerning the investigation of optimal designs for linear models which have been widely studied in theory and practice, see, for example, Atkinson et al. (2007). Linear regression models with uncorrelated errors are often studied in optimal design of experiments, see, for example, the studies of Pázman (1986), Pukelsheim (2006) and Harman and Štulajter (2010). The challenging part of studying optimal experimental design for linear models lies in models with correlated responses, leading to the need for complex numerical approaches (Brimkulov et al., 1980; Müller and Pázman, 1999, 2003). In most applications, there exists a correlation between responses, which makes this challenge of particular practical interest.

2.4.2 Alphabetical optimal designs for GLMs

Many authors have studied alphabetical optimal designs for GLMs, with focus on models with $k = 1$ experimental factor. Ford et al. (1992) investigated D- and c-optimality designs. Dette and Haines (1994) studied E-optimal designs.

Dette et al. (2005) derived optimal design for a restricted design space based on a geometric approach with respect to Kiefer's ψ_k-criteria for regression models with two parameters. Russell et al. (2009b) found a local D-optimal design under a Poisson regression model for any value of independent factors using the natural algorithm as a link function. Niaparast (2009) found locally D-optimal design for a Poisson regression model for a single experimental factor with random intercept in the linear predictor. The construction of local D-optimal designs for binary response with one experimental factor with the link function $g(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right)$ has been discussed by Abdelbasit and Plackett (1983). Russell et al. (2009a) also work on simple logistic regression model with two support points to obtain locally optimal designs using the integrated mean square error as a criterion for design selection. Most research on the generation of optimal designs tailored for GLMs has focused mainly on simple models (see, e.g Chaloner and Larntz (1989); Atkinson et al. (1995); Hedayat et al. (1997); Mathew and Sinha (2001)).

In practice, the response is commonly affected by multiple factors. For large numbers of experimental factors, Dror and Steinberg (2005) found approximate multivariate local D-optimal designs for logistic models with main effects and interactions. Noticeable work was done by Dror and Steinberg (2006) and Woods et al. (2006) for investigating robust designs. The former found robust designs based on clustering a set of locally optimal designs, whereas the latter found exact design for experiments with a specified number of runs and several experimental factors, using a criterion that allows uncertainty in the link function, linear predictor or model parameters.

2.4.3 Difficulties in finding alphabetical optimal designs for GLMs

Finding alphabetical optimal designs for GLMs is hindered by the dependance of the design on the unknown parameters which enter the optimality criteria. In other words, the information matrix for a GLM is a function of the unknown parameters $\boldsymbol{\theta}$; hence, the optimal design is a function of these parameters as well. Accordingly, knowledge of the values of the model parameters is required prior to designing an optimal experiment to estimate them. There are many approaches given in the literature for defining an objective function in cases where the Fisher information depends on the parameters. These are locally optimal (Chernoff, 1953), pseudo-Bayesian (Chaloner and Verdinelli, 1995), minimax (Welch, 1983), compromise (Woods et al., 2006) and sequential (Abdelbasit and Plackett, 1983). In this thesis, we consider the first two common approaches to deal with this problem, a detailed description of them is given in Section 2.4.3.1 and 2.4.3.2, respectively .

2.4.3.1 Locally optimal designs

The locally optimal design is found by minimising the objective function $\Psi(d) = \psi(d;\boldsymbol{\theta}_*)$ (Chernoff, 1953; Chaloner and Verdinelli, 1995; Chen et al., 2013), where $\boldsymbol{\theta}_*$ is an initial guess at the parameters. For examples, a locally D-optimal design minimises

$$\Psi_D(d) = \psi_D(d;\boldsymbol{\theta}_*) = \log|I(\boldsymbol{\theta}_*;d)^{-1}|. \tag{2.24}$$

This is the most simple and widely used approach to design an experiment based on the best guess for the unknown parameters. It is comparatively easy to construct; however, they are only optimal for the given parameter values. In addition, the resulting optimal design under this approach is applicable only for the parameters values used. Hence, if two experiments have the same model with different values of the unknown parameters, this will typically lead to different designs (Dror and Steinberg, 2006, page 520). This approach is useful in evaluating a proposed design for a given choice of parameters (Gotwalt et al., 2009), where it can be compared with the optimal design in terms of its efficiency (how well it does for a particular choice of parameters). The disadvantage of using this approach is if the prior chosen values $\boldsymbol{\theta}_*$ are far from the true parameters $\boldsymbol{\theta}$, the resulting design will be far from optimal (Yang et al., 2012).

2.4.3.2 Pseudo-Bayesian optimal designs

Under the pseudo-Bayesian approach, the point estimates for the unknown parameters used in locally optimal designs are replaced by a prior distribution, where the prior uncertainty about the values of unknown parameters is incorporated into the model. Optimal designs under this approach minimise the expectation of the objective function with respect to a prior distribution of the model parameters. This approach is referred to as "pseudo-Bayesian", and has an asymptotic Bayesian justification, where it is based on a normal approximation to the posterior distribution (Chaloner and Verdinelli, 1995). The word "pseudo" is used alongside Bayesian because the design is still assessed under the frequentist variance of parameter estimators. To formalise, a pseudo-Bayesian optimal design minimises the objective function

$$\Psi(d) = E_{\boldsymbol{\theta}}\left[\psi(d;\boldsymbol{\theta})\right] = \int_{\boldsymbol{\Theta}} \psi(d;\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta},$$

where the expectation is taken with respect to the prior distribution of the parameters, $\pi(\boldsymbol{\theta})$ (Van De Ven and Woods, 2014). For example, pseudo-Bayesian D-optimal designs (Chaloner and Verdinelli, 1995; Atkinson, 1996; Woods et al., 2006) minimise

$$\Psi_D(d) = E_{\boldsymbol{\theta}}\left[\log|I(\boldsymbol{\theta};d)^{-1}|\right] = \int_{\boldsymbol{\Theta}} \log|I(\boldsymbol{\theta};d)^{-1}|\pi(\boldsymbol{\theta})d\boldsymbol{\theta}. \tag{2.25}$$

The prior distribution gives an idea, before collecting the data, about the probabilities of the possible values of $\boldsymbol{\theta}$ (Garthwaite et al., 2002, Chapter 6). Typically, prior information is often available from previous experiments. The major challenge that we typically encounter is that the calculation of the expectation, which results from a multidimensional integral, tends to be difficult or intractable, and approximation methods must be used. In addition, algorithms for finding optimal design under this approach evaluate the integral many times, leading to expensive computing time.

2.5 Summary

In this chapter, we introduced key concepts of designing experiments and some underlying concepts in estimation theory. We reviewed the properties of choosing estimators and gave a detailed description of the most widely applicable approach to finding estimators, namely maximum likelihood estimation. Maximum likelihood estimates under GLMs may be non-unique or infinite for certain observations. Consequently, the maximum penalised likelihood estimation was considered as an alternative estimation method to estimate the parameters. An introduction to GLMs and examples that have essential applications in the design of experiments were given. We demonstrated mathematical frequentist alphabetic optimality with three popular examples of alphabetical optimal designs, namely D-, A- and E-optimal designs. In addition, we gave an introduction to alphabetical optimal design for linear models, and illustrated challenges that could be arisen in design for linear models. Last, we reviewed alphabetical optimal designs for GLMs and its challenges, with two commonly used approaches to handle the challenges in design for GLMs, namely local and pseudo-Bayesian optimality.

Chapter 3

Frequentist Decision-theoretic Optimal Designs

3.1 Introduction

We gave a brief introduction to the alternative decision-theoretic basis for frequentist optimal designs in Chapter 1. A frequentist decision-theoretic optimal design is found by minimising the risk function over a design space $\mathcal{X}$. The risk function is defined as the expectation of a loss function $\lambda(\hat{\boldsymbol{\theta}}(\mathbf{Y}), \boldsymbol{\theta}, d)$, with respect to the distribution of $\mathbf{Y}$. The loss, $\lambda(\hat{\boldsymbol{\theta}}(\mathbf{Y}), \boldsymbol{\theta}, d)$, specifies the loss of estimating the parameters $\boldsymbol{\theta}$ with the estimators $\hat{\boldsymbol{\theta}}(\mathbf{Y})$, given the observed data $\mathbf{Y}$ obtained from the design d. Hence, the risk function is given by

$$R(d; \boldsymbol{\theta}) = \int \lambda(\hat{\boldsymbol{\theta}}(\mathbf{Y}), \boldsymbol{\theta}, d) f(\mathbf{Y}; \boldsymbol{\theta}, d) d\mathbf{Y}, \tag{3.1}$$

where $\mathbf{Y}$ is a continuous variable and the model parameters $\boldsymbol{\theta}$ are considered to be fixed. In the case of $\mathbf{Y}$ is a discrete variable, a sum is used instead of an integral. The risk function is often a matrix; hence, a decision-theoretic (DT) design is found by minimising an objective function $\psi_{DT}(d; \boldsymbol{\theta})$ given by a scalar function of the risk function $S(R(d; \boldsymbol{\theta}))$.

The advantages of the decision theoretic approach are that it can be applied for bespoke experimental aims and it is suitable for small sample sizes. These advantages enhance the use of the decision-theoretic framework to frequentist optimal designs rather than the use of the standard approach of alphabetical optimal designs. The frequentist decision-theoretic framework will be described for the experimental aims of parameter estimation (see Chapter 6) and model discrimination (see Chapter 7), represented by the specification of the loss function. Some examples of loss functions are given in Section 3.2.

The disadvantages of the decision theoretic approach are that finding optimal designs under this approach is hindered by some challenges. First, similar to the Fisher information in alphabetical optimal designs, the risk function depends on unknown parameters. Therefore, local and pseudo-Bayesian approaches are considered and applied to overcome the dependence of the risk on the parameters for finding decision-theoretic optimal designs in GLMs. The pseudo-Bayesian decision-theoretic optimal designs minimise the objective function

$$\Psi_{DT}(d) = E_{\boldsymbol{\theta}}\left[\psi_{DT}(d;\boldsymbol{\theta})\right] = \int_{\boldsymbol{\Theta}} \psi_{DT}(d;\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}, \tag{3.2}$$

over the design space $\mathcal{X}$. This gives the Bayesian risk, whereas the locally decision-theoretic optimal design is a special case for a point mass prior. Second, the risk function is often analytically intractable and of high dimension, leading to the need for approximation to the risk function. A number of approximation methods is considered in Chapter 4. Third, the objective function will need to be optimised over a high-dimensional design space. This can be addressed throughout the use of the recently-proposed approximate coordinate exchange (ACE) algorithm (Overstall and Woods, 2017), see "Appendix A" for more details.

3.2 Examples of loss functions

The loss function represents the aim of the experiment. It should be chosen after careful discussion with the experimenter. In this section, some examples of default loss functions tailored to the experimental aim of parameter estimation are given.

3.2.1 Squared error matrix

The squared error matrix (SE) is defined (Chaloner and Verdinelli, 1995) as:

$$\lambda_{SE}(\hat{\boldsymbol{\theta}}(\mathbf{Y}), \boldsymbol{\theta}, d) = (\hat{\boldsymbol{\theta}}(\mathbf{Y}) - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}}(\mathbf{Y}) - \boldsymbol{\theta})^T, \tag{3.3}$$

and is also known as quadratic loss. It is used when the experimental aim is to obtain a point estimate of the parameters (e.g investigating the effect of a new treatment). It gives larger penalty when the estimate is far from the true parameter. The risk function under the squared error matrix is the MSE and given in (2.2). Here, under the SE loss, the pseudo-Bayesian decision-theoretic D-optimal design minimises the objective function

$$\Psi_{DTD}(d) = E_{\boldsymbol{\theta}}\left[\log|R_{SE}(\boldsymbol{\theta};d)|\right] = \int_{\boldsymbol{\Theta}} \log|R_{SE}(\boldsymbol{\theta};d)|\pi(\boldsymbol{\theta})d\boldsymbol{\theta}. \tag{3.4}$$

In this thesis, we consider two scalar functions of $R_{SE}(\boldsymbol{\theta}; d)$ in addition to the log determinant $\log|R_{SE}(\boldsymbol{\theta}; d)|$, namely the trace $\mathrm{tr}\{R_{SE}(\boldsymbol{\theta}; d)\}$ and maximum eigenvalue $\lambda_{max}\{R_{SE}(\boldsymbol{\theta}; d)\}$, where the resulting designs that minimise these functions are pseudo-Bayesian decision-theoretic A- and E-optimal designs, respectively.

3.2.2 The sign loss

The novel sign loss (SL) is

$$\lambda_{SL}(\hat{\boldsymbol{\theta}}(\mathbf{Y}), \boldsymbol{\theta}, d) = \frac{1}{2}\sum_{i=1}^{p}\left|\mathrm{sgn}\left(\hat{\theta}_i(\mathbf{Y})\right) - \mathrm{sgn}\left(\theta_i\right)\right|, \tag{3.5}$$

where the $\mathrm{sgn}(A)$ refers to the sign function of A such that

$$\mathrm{sgn}(A) = \begin{cases} -1 & \text{if } A < 0, \\ +1 & \text{if } A > 0. \end{cases} \tag{3.6}$$

It can be also expressed as

$$\lambda_{SL}(\hat{\boldsymbol{\theta}}(\mathbf{Y}), \boldsymbol{\theta}, d) = \sum_{i=1}^{p}\mathbb{1}\left(\mathrm{sgn}(\hat{\theta}_i(\mathbf{Y})) = \mathrm{sgn}(\theta_i)\right), \tag{3.7}$$

where $\mathbb{1}(.)$ is the indicator function. The loss gives the number of times the estimate of the parameter has the wrong sign.

3.3 The risk function under linear models

It is known that the risk function typically depends on the unknown parameters, so the objective function that we aim to minimise for finding decision-theoretic optimal designs can be formed using one of the local or pseudo-Bayesian approaches. As the asymptotic results of MLEs are exact under linear models, e.g. $\hat{\boldsymbol{\theta}}(\mathbf{Y}) \sim N(\boldsymbol{\theta}, \sigma^2(X^TX)^{-1})$, the objective function that we aim to minimise for finding decision-theoretic optimal designs, for the class of linear models, does not often depend on the parameters.

3.3.1 The risk function under the squared error matrix for linear models

Under the squared error matrix, the risk function can be written as

$$R_{SE}(d; \boldsymbol{\theta}) = Var_{\mathbf{Y}|\boldsymbol{\theta}}\left[\hat{\boldsymbol{\theta}}(\mathbf{Y})\right] + b(\hat{\boldsymbol{\theta}}(\mathbf{Y}))b^T(\hat{\boldsymbol{\theta}}(\mathbf{Y})).$$

Using the fact that the asymptotic results of MLEs are exact under the normal linear models, the MLEs are unbiased estimators and the risk function is given by

the inverse Fisher information matrix

$$R_{SE}(d;\boldsymbol{\theta}) = Var_{\mathbf{Y}|\boldsymbol{\theta}}\left[\hat{\boldsymbol{\theta}}(\mathbf{Y})\right],$$

where $Var_{\mathbf{Y}|\boldsymbol{\theta}}\left[\hat{\boldsymbol{\theta}}(\mathbf{Y})\right] = \sigma^2(X^TX)^{-1}$.

Hence, a decision-theoretic D-, A- and E-optimal design, under linear models, minimises the objective function

$$\Psi_{DTD}(d) = \log|(X^TX)^{-1}|,$$

$$\Psi_{DTA}(d) = \text{tr}\{(X^TX)^{-1}\}$$

and

$$\Psi_{DTE}(d) = \lambda_{\max}\{(X^TX)^{-1}\},$$

respectively.

3.3.2 The risk function under the sign loss for linear models

Now consider the SL in (3.7), and note that $\hat{\theta}_i(\mathbf{Y}) \sim N(\theta_i, \sigma^2[(X^TX)^{-1}]_{ii})$, where $[(X^TX)^{-1}]_{ii}$ is ith diagonal elements of $(X^TX)^{-1}$. Given the definition of the sign function in (3.6), the SL can then be written as follows

$$\lambda_{SL}(\hat{\boldsymbol{\theta}}(\mathbf{Y}),\boldsymbol{\theta},d) = \sum_{i=1}^{p} \mathbb{1}\left(\hat{\theta}_i(\mathbf{Y}) < 0\right)\mathbb{1}\left(\theta_i < 0\right) + \mathbb{1}\left(\hat{\theta}_i(\mathbf{Y}) > 0\right)\mathbb{1}\left(\theta_i > 0\right). \quad (3.8)$$

The risk function is given by the expectation of (3.8) with respect to the distribution of $\hat{\boldsymbol{\theta}}(\mathbf{Y})$ as follows:

$$R_{SL}(d;\boldsymbol{\theta}) = \sum_{i=1}^{p} P\left(\hat{\theta}_i(\mathbf{Y}) < 0\right)\mathbb{1}\left(\theta_i < 0\right) + P\left(\hat{\theta}_i(\mathbf{Y}) > 0\right)\mathbb{1}\left(\theta_i > 0\right). \quad (3.9)$$

Following that, we standardise the normally distributed random variable $\hat{\theta}_i(\mathbf{Y})$ in (3.9), i.e.

$$R_{SL}(d;\boldsymbol{\theta}) = \sum_{i=1}^{p} P\left(\frac{\hat{\theta}_i(\mathbf{Y}) - \theta_i}{\Sigma_{ii}} < \frac{-\theta_i}{\Sigma_{ii}}\right)\mathbb{1}\left(\theta_i < 0\right) + P\left(\frac{\hat{\theta}_i(\mathbf{Y}) - \theta_i}{\Sigma_{ii}} > \frac{-\theta_i}{\Sigma_{ii}}\right)\mathbb{1}\left(\theta_i > 0\right), \quad (3.10)$$

where $\Sigma_{ii} = \sqrt{\sigma^2[(X^TX)^{-1}]_{ii}}$. The right hand side of (3.10) can be simplified to

$$\sum_{i=1}^{p} \Phi\left(\frac{-\theta_i}{\Sigma_{ii}}\right)\mathbb{1}\left(\theta_i < 0\right) + \left(1 - \Phi\left(\frac{-\theta_i}{\Sigma_{ii}}\right)\right)\left(1 - \mathbb{1}(\theta_i < 0)\right),$$

where $\Phi\left(\frac{-\theta_i}{\Sigma_{ii}}\right)$ refers to the cumulative distribution function of the standard normal distribution. Then, the risk function under the SL is given by

$$R_{SL}(d;\boldsymbol{\theta}) = \sum_{i=1}^{p} \left(2\mathbb{1}(\theta_i < 0) - 1\right) \Phi\left(\frac{-\theta_i}{\Sigma_{ii}}\right) - \mathbb{1}\left(\theta_i < 0\right) + 1.$$

3.4 Summary

We reviewed frequentist decision-theoretic optimal designs, and the objective function (risk function) under the frequentist decision-theoretic framework was defined. Furthermore, we outlined the hurdles to finding frequentist decision-theoretic optimal designs for GLMs, and these were:

- The risk function depends on unknown parameters, leading to difficulties in finding decision-theoretic optimal designs.We will use the local and pseudo-Bayesian approaches.
- The risk function is often analytically intractable, and this hinders finding exact frequentist decision theoretic optimal designs, leading to the need for approximation methods to the risk function.
- The objective function will need to be optimised over a multi-dimensional design space.

Some common loss functions used in finding decision-theoretic optimal designs for parameter estimation were described. Last, the risk function under the class of linear models was also reviewed and illustrated under the considered loss functions in this chapter.

Chapter 4

Methods for Approximating the Bayesian Risk and Risk Function

4.1 Introduction

The Bayesian risk that is used to find pseudo-Bayesian decision-theoretic optimal designs, defined in (3.2), is not available in closed form and results, typically, from a low-dimensional (p-dimensional) integral. The primary difficulty encountered when minimising the Bayesian risk is the need to accurately compute the integral. Typically, this requires a numerical evaluation of the p-dimensional integral. In addition, computational algorithms for finding optimal designs require many evaluations of the integral. This makes an approximation such as Monte-Carlo integration rule painfully slow (Gotwalt et al., 2009). We used a deterministic quadrature approximation technique to the Bayesian risk, described in Gotwalt et al. (2009), where the p-dimensional integral in (3.2) can be approximated using a radial-spherical integration rule (Monahan and Genz, 1997). Here, the integral is reparameterized into a radial component τ and $p-1$ spherical surface components $\mathbf{v}$. The radial integral is performed using generalised Gauss-Laguerre quadrature with abscissas and weights $\{\tau_i, w_{R_i}\}_{i=0}^{n_R}$, where n_R refers to the number of radial abscissas, whereas the spherical integral is approximated using Mysovskikh (1980) extended simplex rule with abscissas and weights $\{\mathbf{v}_j, w_{S_k}\}_{k=0}^{n_s}$, where n_s refers to the number of points in the extended simplex rule. For each nonzero radius, random orthogonal matrices n_Q, $\{\mathbf{Q}_{ij}\}_{j=1}^{n_Q}$, are created, and the value of the integrand is averaged over these orthogonal rotations.

Here, we illustrate how to form the generalised Gauss-Laguerre abscissas and the Mysovskikh extended simplex points and their corresponding weights. The extended simplex method is performed by creating a $p+1$ vertex-centered simplex in $\mathbb{R}^p$,

$$\mathbf{v}_{ij} = \begin{cases} -\sqrt{\dfrac{p+1}{p(p-j+2)(p-j+1)}}, & j < i, \\ \sqrt{\dfrac{(p+1)(p-i+1)}{p(p-i+2)}}, & j = i, \\ 0, & j > i. \end{cases} \tag{4.1}$$

The midpoints of the simplex vertices are created to proceed the construction that is completed by adding the negatives of all vertices and midpoints. The weights of the extended simplex only takes two values. The simplex vertex points and the midpoints and their corresponding negatives receive a weight of $\frac{p(7-p)}{2(p+1)^2(p+2)}$ and $\frac{2(p-1)^2}{p(p+1)^2(p+2)}$, respectively. The algorithm generates $(p+1)(p+2)$ abscissas, and the weights of these abscissa are then summed. The radial integral is approximated using generalised Gauss-Laguerre quadrature with a point added at 0 that causes the generalised Laguerre shape parameter to be incremented by one. Without adding these zeros, the abscissas would be twice the roots of the generalised Laguerre polynomial with parameter $p/2 - 1$

Approximating the Bayesian risk via the quadrature approximation is illustrated by Gotwalt et al. (2009) for use with normal prior distributions. In addition, they describe this approach for use with non-normal priors by applying a transformation to integrate optimality criteria, provided independent prior distributions on the parameters. This means that for each considered θ_i it has its own prior density $\pi_i(\boldsymbol{\theta_i})$, and it is differentiable. Let G_i refers to the cumulative distribution function (cdf) for θ_i. For this situation, the Bayesian criterion can be written as

$$\Psi_{DT}(d) = \int \psi_{DT}(d; \theta_1, \ldots, \theta_p) \prod_{i=1}^{p} \pi_i(\boldsymbol{\theta_i}) d\theta_i.$$

Let $t_i \ldots, t_p$ be a sample from a standard normal distribution. Then $G^{-1}(\Phi(t_i))$ has the same distribution as θ_i due to the transformation from $\mathbf{t}$ to $\boldsymbol{\theta}$, hence the Bayesian criterion is

$$\Psi_{DT}(d) = \int \psi_{DT}(d; G_1^{-1}(\Phi(t_1)), \ldots, G_p^{-1}(\Phi(t_p))) \prod_{i=1}^{p} \phi(t_i) dt_i.$$

From this an extension to non-normal priors is proposed. Let $\mathbf{G}^{-1}(\boldsymbol{\Phi}(\mathbf{t}))$ refers to the vector mapping of $\mathbf{t}$ to the $\boldsymbol{\theta}$. Then the quadrature approximation with non-constant weights and abscissas to the Bayesian risk is

$$\Psi_{DT}(d) \approx w_{R_0} \boldsymbol{\psi}_{DT}(d; \mathbf{G}^{-1}(\boldsymbol{\Phi}(\mathbf{0}))) + \Sigma_{i=1}^{n_R} \Sigma_{j=1}^{n_Q} \Sigma_{k=1}^{n_s} \frac{w_{R_i} w_{S_k}}{n_Q} \boldsymbol{\psi}_{DT}(d; \mathbf{G}^{-1}(\boldsymbol{\Phi}(\sqrt{\tau}_i \mathbf{Q}_{ij} \mathbf{v}_k))). \tag{4.2}$$

A large number of n_R and n_Q aid to obtain an accurate approximation. In this thesis we used $n_R = 3$ and $n_Q = 2$.

This approximation can be also written such that

$$\Psi_{DT}(d) \approx \sum_{h=1}^{H} w_h \psi_{DT}(d; \boldsymbol{\theta}_h), \tag{4.3}$$

where w_h refers to the vector of weights given in (4.2), $\left[w_{R_0}, \frac{w_{R_i} w_{S_k}}{n_Q}\right]^T$, and $\boldsymbol{\theta}_h$ refers to the abscissas, a $H \times p$ matrix of zero abscissas $\mathbf{G}^{-1}(\boldsymbol{\Phi}(\mathbf{0}))$ and the other abscissas $\mathbf{G}^{-1}(\boldsymbol{\Phi}(\sqrt{\tau}_i \mathbf{Q}_{ij} \mathbf{v}_k))$.

The objective function $\psi_{DT}(d; \boldsymbol{\theta}_h)$ was defined in Chapter 3 and given by a scalar function of the risk function $S(R(d; \boldsymbol{\theta}))$. The risk function (defined in (3.1)) is analytically intractable, and results from a multidimensional (N-dimensional) integral. In order to handle the intractability of the risk function, an approximation method to the risk function is required. In this chapter, some approximation techniques are described, and these are asymptotic (Asym), second-order Taylor series (ST), quadrature (Quad), quasi-Monte Carlo (QMC) and quasi-importance sampling (QIMPS).

4.2 Asymptotic approximation

First note that the risk function can be written as follows

$$R(d; \boldsymbol{\theta}) = \int \lambda(\hat{\boldsymbol{\theta}}(\mathbf{Y}), \boldsymbol{\theta}; d) f(\hat{\boldsymbol{\theta}}(\mathbf{Y}); \boldsymbol{\theta}, d) \mathrm{d}\hat{\boldsymbol{\theta}}(\mathbf{Y}),$$

where $f(\hat{\boldsymbol{\theta}}(\mathbf{Y}); \boldsymbol{\theta}, d)$ is the pdf of the estimator $\hat{\boldsymbol{\theta}}(\mathbf{Y})$. Here, the estimator $\hat{\boldsymbol{\theta}}(\mathbf{Y})$ is a random variable. In the case of discrete $\hat{\boldsymbol{\theta}}(\mathbf{Y})$, the sum replaces the integral. If $\hat{\boldsymbol{\theta}}(\mathbf{Y})$ is the maximum likelihood estimator, then due to the tractability of the normal distribution, the approximate distribution of the MLEs can be used to provide a simple approximation to the risk function. Therefore, the risk function is written in this form because the asymptotic approximation is based on the normal approximation to the MLEs.

4.2.1 Asymptotic approximation to the risk under the squared error matrix

In the case of the SE loss, the risk function is given by (2.3). The MLE is approximately unbiased, and the variance is approximately the inverse Fisher information matrix, $I(\boldsymbol{\theta}; d)^{-1}$, which is a large sample approximation to the variance matrix of MLEs. Therefore, the risk function is given by the inverse Fisher information, and an approximation to the objective function under decision-theoretic optimal

designs corresponds to the objective function under alphabetical optimal designs. For example, if we choose the scalar function S to be the log determinant, then the asymptotic approximation to the risk function for the decision-theoretic design is given by the objective function for finding the D-optimal design (see equation (2.21)).

4.2.2 Asymptotic approximation to the risk under the sign loss

Now consider the SL in (3.7), and note that $\hat{\theta}_i(\mathbf{Y}) \sim N(\theta_i, [I(\boldsymbol{\theta}; d)^{-1}]_{ii})$, where $[I(\boldsymbol{\theta}; d)^{-1}]_{ii}$ is ith diagonal elements of $I(\boldsymbol{\theta}; d)^{-1}$, and $I(\boldsymbol{\theta}; d)$ is the Fisher information matrix given in (2.1). Given the definition of the sign function in (3.6), the SL can be then written as follows

$$\lambda_{SL}(\hat{\boldsymbol{\theta}}(\mathbf{Y}), \boldsymbol{\theta}, d) = \sum_{i=1}^{p} \mathbb{1}\left(\hat{\theta}_i(\mathbf{Y}) < 0\right) \mathbb{1}\left(\theta_i < 0\right) + \mathbb{1}\left(\hat{\theta}_i(\mathbf{Y}) > 0\right) \mathbb{1}\left(\theta_i > 0\right). \quad (4.4)$$

The risk function is given by the expectation of (4.4) with respect to the distribution of $\hat{\boldsymbol{\theta}}(\mathbf{Y})$ as follows:

$$R_{SL}(d; \boldsymbol{\theta}) = \sum_{i=1}^{p} P\left(\hat{\theta}_i(\mathbf{Y}) < 0\right) \mathbb{1}\left(\theta_i < 0\right) + P\left(\hat{\theta}_i(\mathbf{Y}) > 0\right) \mathbb{1}\left(\theta_i > 0\right). \quad (4.5)$$

Following that, we standardise the normally distributed random variable $\hat{\theta}_i(\mathbf{Y})$ in (4.5), i.e.

$$R_{SL}(d; \boldsymbol{\theta}) = \sum_{i=1}^{p} P\left(\frac{\hat{\theta}_i(\mathbf{Y}) - \theta_i}{\Sigma_{ii}} < \frac{-\theta_i}{\Sigma_{ii}}\right) \mathbb{1}\left(\theta_i < 0\right) + P\left(\frac{\hat{\theta}_i(\mathbf{Y}) - \theta_i}{\Sigma_{ii}} > \frac{-\theta_i}{\Sigma_{ii}}\right) \mathbb{1}\left(\theta_i > 0\right), \quad (4.6)$$

where $\Sigma_{ii} = \sqrt{[I^{-1}(\boldsymbol{\theta}; d)]_{ii}}$. The right hand side of (4.6) can be simplified to

$$\sum_{i=1}^{p} \Phi\left(\frac{-\theta_i}{\Sigma_{ii}}\right) \mathbb{1}\left(\theta_i < 0\right) + \left(1 - \Phi\left(\frac{-\theta_i}{\Sigma_{ii}}\right)\right) \left(1 - \mathbb{1}(\theta_i < 0)\right),$$

where $\Phi\left(\frac{-\theta_i}{\Sigma_{ii}}\right)$ refers to the cumulative distribution function of the standard normal distribution. Then, the asymptotic approximation to the risk function under the SL is given by

$$R_{SL}(d; \boldsymbol{\theta}) \approx \sum_{i=1}^{p} \left(2\mathbb{1}(\theta_i < 0) - 1\right) \Phi\left(\frac{-\theta_i}{\Sigma_{ii}}\right) - \mathbb{1}\left(\theta_i < 0\right) + 1.$$

4.3 Second-order Taylor series approximation

We approximate the risk function by the second-order Taylor series approximation. The development in this section is a generalisation of the approach used by Pronzato and Pázman (2013) to any type of models, not just non-linear models. In order to apply this approximation technique, the loss function must be twice differentiable. Suppose the loss function is a matrix, then define $\lambda_{qr}(\hat{\boldsymbol{\theta}}(\mathbf{Y}), \boldsymbol{\theta}, d)$ to be the qrth element. We use the second-order Taylor series expansion in the neighbourhood of the point $\boldsymbol{\mu}$, where $\boldsymbol{\mu}$ denotes the mean response, $\boldsymbol{\mu} = E_{\mathbf{Y}|\boldsymbol{\theta}}(\mathbf{Y})$. That is

$$\lambda_{qr}(\hat{\boldsymbol{\theta}}(\mathbf{Y}), \boldsymbol{\theta}, d) \approx \lambda_{qr}(\boldsymbol{\mu}, \boldsymbol{\theta}, d) + \left.\frac{\partial \lambda_{qr}(\hat{\boldsymbol{\theta}}(\mathbf{Y}), \boldsymbol{\theta}, d)}{\partial \mathbf{Y}}\right|_{\mathbf{Y}=\boldsymbol{\mu}} (\mathbf{Y} - \boldsymbol{\mu}) + \frac{1}{2}(\mathbf{Y} - \boldsymbol{\mu})^T \left.\frac{\partial^2 \lambda_{qr}(\hat{\boldsymbol{\theta}}(\mathbf{Y}), \boldsymbol{\theta}, d)}{\partial \mathbf{Y} \partial \mathbf{Y}^T}\right|_{\mathbf{Y}=\boldsymbol{\mu}} (\mathbf{Y} - \boldsymbol{\mu}). \tag{4.7}$$

An approximate risk function is given by the expectation of (4.7) with respect to the distribution of $\mathbf{Y}$ as follows:

$$R_{qr}(d; \boldsymbol{\theta}) \approx \lambda_{qr}(\boldsymbol{\mu}, \boldsymbol{\theta}, d) + \frac{1}{2} \mathrm{E}_{\mathbf{Y}|\boldsymbol{\theta}} \left[(\mathbf{Y} - \boldsymbol{\mu})^T \left.\frac{\partial^2 \lambda_{qr}(\hat{\boldsymbol{\theta}}(\mathbf{Y}), \boldsymbol{\theta}, d)}{\partial \mathbf{Y} \partial \mathbf{Y}^T}\right|_{\mathbf{Y}=\boldsymbol{\mu}} (\mathbf{Y} - \boldsymbol{\mu}) \right], \tag{4.8}$$

where $R_{qr}(d; \boldsymbol{\theta})$ denotes the qrth element of the risk function. The expectation in the second term of (4.8) (Mathai and Provost, 1992) is given by

$$\mathrm{E}_{\mathbf{Y}|\boldsymbol{\theta}} \left[(\mathbf{Y} - \boldsymbol{\mu})^T \left.\frac{\partial^2 \lambda_{qr}(\hat{\boldsymbol{\theta}}(\mathbf{Y}), \boldsymbol{\theta}, d)}{\partial \mathbf{Y} \partial \mathbf{Y}^T}\right|_{\mathbf{Y}=\boldsymbol{\mu}} (\mathbf{Y} - \boldsymbol{\mu}) \right] = \mathrm{tr} \left[\left.\frac{\partial^2 \lambda_{qr}(\hat{\boldsymbol{\theta}}(\mathbf{Y}), \boldsymbol{\theta}, d)}{\partial \mathbf{Y} \partial \mathbf{Y}^T}\right|_{\mathbf{Y}=\boldsymbol{\mu}} \Sigma \right] + \boldsymbol{\kappa}^T \left.\frac{\partial^2 \lambda_{qr}(\hat{\boldsymbol{\theta}}(\mathbf{Y}), \boldsymbol{\theta}, d)}{\partial \mathbf{Y} \partial \mathbf{Y}^T}\right|_{\mathbf{Y}=\boldsymbol{\mu}} \boldsymbol{\kappa},$$

where $\boldsymbol{\kappa}$ and Σ are the mean (expected value) and variance matrix of $(\mathbf{Y} - \boldsymbol{\mu})$, respectively. The mean and variance are taken with respect to the distribution of $\mathbf{Y}$, i.e

$$\boldsymbol{\kappa} = E_{\mathbf{Y}|\boldsymbol{\theta}} [\mathbf{Y} - \boldsymbol{\mu}] = \mathbf{0}, \tag{4.9}$$

$$\Sigma = Var_{\mathbf{Y}|\boldsymbol{\theta}} [\mathbf{Y} - \boldsymbol{\mu}] = Var_{\mathbf{Y}|\boldsymbol{\theta}} [\mathbf{Y}], \tag{4.10}$$

respectively. The mean response is given by $\boldsymbol{\mu} = E_{\mathbf{Y}|\boldsymbol{\theta}} [\mathbf{Y}] = g^{-1}(\boldsymbol{\eta}(\boldsymbol{\theta})) = g^{-1}(X\boldsymbol{\theta})$, the ith element of $X\boldsymbol{\theta}$ is $\mathbf{x}_i^T \boldsymbol{\theta}$. Given (4.9) and (4.10), the expectation in (4.8) can be simplified. Hence, the simplified approximation to the risk function is

$$R_{qr}(d;\boldsymbol{\theta}) \approx \lambda_{qr}(\boldsymbol{\mu},\boldsymbol{\theta},d) + \frac{1}{2}\,\mathrm{tr}\left[\left.\frac{\partial^2\lambda_{qr}(\hat{\boldsymbol{\theta}}(\mathbf{Y}),\boldsymbol{\theta},d)}{\partial\mathbf{Y}\partial\mathbf{Y}^T}\right|_{\mathbf{Y}=\boldsymbol{\mu}}\mathrm{Var}_{\mathbf{Y}|\boldsymbol{\theta}}\left[\mathbf{Y}\right]\right]. \tag{4.11}$$

Assuming independence of the elements of $\mathbf{Y}$, the second-term on the right-hand side of (4.11) can be simplified to

$$\frac{1}{2}\sum_{i=1}^{N}\left.\frac{\partial^2\lambda_{qr}(\hat{\boldsymbol{\theta}}(\mathbf{Y}),\boldsymbol{\theta},d)}{\partial Y_i^2}\right|_{\mathbf{Y}=\boldsymbol{\mu}}\mathrm{Var}_{\mathbf{Y}|\boldsymbol{\theta}}\left[Y_i\right].$$

Therefore, the second-order Taylor series approximation to the risk function is

$$R_{qr}(d;\boldsymbol{\theta}) \approx \lambda_{qr}(\boldsymbol{\mu},\boldsymbol{\theta},d) + \frac{1}{2}\sum_{i=1}^{N}\left.\frac{\partial^2\lambda_{qr}(\hat{\boldsymbol{\theta}}(\mathbf{Y}),\boldsymbol{\theta},d)}{\partial Y_i^2}\right|_{\mathbf{Y}=\boldsymbol{\mu}}\mathrm{Var}_{\mathbf{Y}|\boldsymbol{\theta}}\left[Y_i\right]. \tag{4.12}$$

The second partial derivatives of $\lambda_{qr}(\hat{\boldsymbol{\theta}}(\mathbf{Y}),\boldsymbol{\theta},d)$ are obtained by applying the chain rule as follows:

$$\begin{aligned}\left.\frac{\partial^2\lambda_{qr}(\hat{\boldsymbol{\theta}}(\mathbf{Y}),\boldsymbol{\theta},d)}{\partial Y_i^2}\right|_{\mathbf{Y}=\boldsymbol{\mu}} = \Bigg\{&\left.\frac{\partial\lambda_{qr}(\hat{\boldsymbol{\theta}}(\mathbf{Y}),\boldsymbol{\theta},d)}{\partial\hat{\boldsymbol{\theta}}(\mathbf{Y})}\,\frac{\partial^2\hat{\boldsymbol{\theta}}(\mathbf{Y})}{\partial Y_i^2}\right|_{\mathbf{Y}=\boldsymbol{\mu}}\\ &+\left.\frac{\partial\hat{\boldsymbol{\theta}}(\mathbf{Y})^T}{\partial Y_i}\right|_{\mathbf{Y}=\boldsymbol{\mu}}\left.\frac{\partial^2\lambda_{qr}(\hat{\boldsymbol{\theta}}(\mathbf{Y}),\boldsymbol{\theta},d)}{\partial\hat{\boldsymbol{\theta}}(\mathbf{Y})\partial\hat{\boldsymbol{\theta}}(\mathbf{Y})^T}\,\frac{\partial\hat{\boldsymbol{\theta}}(\mathbf{Y})}{\partial Y_i}\right|_{\mathbf{Y}=\boldsymbol{\mu}}\Bigg\}.\end{aligned} \tag{4.13}$$

For squared error loss,

$$\begin{aligned}\left.\frac{\partial^2\lambda_{qr}(\hat{\boldsymbol{\theta}}(\mathbf{Y}),\boldsymbol{\theta},d)}{\partial Y_i^2}\right|_{\mathbf{Y}=\boldsymbol{\mu}} = &-\left(\theta_r-\hat{\theta}_r(\mathbf{Y})\right)\left.\frac{\partial^2\hat{\theta}_q(\mathbf{Y})}{\partial Y_i^2}\right|_{\mathbf{Y}=\boldsymbol{\mu}}-\left(\theta_q-\hat{\theta}_q(\mathbf{Y})\right)\left.\frac{\partial^2\hat{\theta}_r(\mathbf{Y})}{\partial Y_i^2}\right|_{\mathbf{Y}=\boldsymbol{\mu}}\\ &+2\left.\frac{\partial\hat{\theta}_r(\mathbf{Y})}{\partial Y_i}\right|_{\mathbf{Y}=\boldsymbol{\mu}}\left.\frac{\partial\hat{\theta}_q(\mathbf{Y})}{\partial Y_i}\right|_{\mathbf{Y}=\boldsymbol{\mu}},\end{aligned} \tag{4.14}$$

for $q,r=1,...,p$.

4.3.1 Second-order Taylor Series for GLMs

Approximating the risk function via second-order Taylor series method can be derived for GLMs. Suppose that $\mathbf{Y}=(Y_1,\ldots,Y_N)$ is a vector of random variables and follows a probability distribution function which is a member of the exponential family of distributions. Hence, the approximation to the risk function in

(4.12) can be obtained by calculating the terms $\frac{\partial\hat{\boldsymbol{\theta}}(\mathbf{Y})}{\partial Y_i}$ and $\frac{\partial^2\hat{\boldsymbol{\theta}}(\mathbf{Y})}{\partial Y_i^2}$ required for the term $\frac{\partial^2\lambda_{qr}(\hat{\boldsymbol{\theta}}(\mathbf{Y}),\boldsymbol{\theta},d)}{\partial Y_i^2}\Big|_{\mathbf{Y}=\boldsymbol{\mu}}$ in (4.14). This is accomplished using the implicit function theorem (Krantz and Parks, 2012). Assume a canonical link has been used, then the score function for a GLM is as follows:

$$U(\boldsymbol{\theta},\mathbf{Y}) = X^T(\mathbf{Y}-\boldsymbol{\mu}).$$

Then, the term $\frac{\partial\hat{\boldsymbol{\theta}}(\mathbf{Y})}{\partial Y_i}$ is

$$\frac{\partial\hat{\boldsymbol{\theta}}(\mathbf{Y})}{\partial Y_i} = -\left[\frac{\partial U(\boldsymbol{\theta},\mathbf{Y})}{\partial\boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}(\mathbf{Y})}\right]^{-1}\frac{\partial U(\boldsymbol{\theta},\mathbf{Y})}{\partial Y_i}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}(\mathbf{Y})}. \tag{4.15}$$

The derivative of the score function is calculated with respect to each of $\hat{\boldsymbol{\theta}}(\mathbf{Y})$ and $\mathbf{Y}$ as follows:

$$\frac{\partial U(\boldsymbol{\theta},\mathbf{Y})}{\partial\boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}(\mathbf{Y})} = -X^T\frac{\partial\boldsymbol{\mu}}{\partial\boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}(\mathbf{Y})}, \tag{4.16}$$

and

$$\frac{\partial U(\boldsymbol{\theta},\mathbf{Y})}{\partial\mathbf{Y}}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}(\mathbf{Y})} = X^T. \tag{4.17}$$

Note that,

$$\frac{\partial\mu_i}{\partial\theta_j} = \frac{\partial\mu_i}{\partial\eta_i}\frac{\partial\eta_i}{\partial\theta_j},$$

where $\eta_i = \mathbf{x}_i^T\boldsymbol{\theta}$. It follows that

$$\frac{\partial\mu_i}{\partial\theta_j} = \frac{x_{ij}}{g'(\mu_i)} = \mathrm{Var}_{\mathbf{Y}|\boldsymbol{\theta}}[Y_i]\,x_{ij},$$

where $g'(\mu_i) = 1/\mathrm{Var}_{\mathbf{Y}|\boldsymbol{\theta}}[Y_i]$ and $\frac{\partial\mu_i}{\partial\theta_j}$ can be expressed in a matrix form as

$$\frac{\partial\boldsymbol{\mu}}{\partial\boldsymbol{\theta}} = \mathrm{Var}_{\mathbf{Y}|\boldsymbol{\theta}}[\mathbf{Y}]\,X.$$

Hence,

$$\frac{\partial\hat{\boldsymbol{\theta}}(\mathbf{Y})}{\partial Y_i} = \left[X^T\,\mathrm{Var}_{\mathbf{Y}|\boldsymbol{\theta}}[\mathbf{Y}]\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}(\mathbf{Y})}X\right]^{-1}\mathbf{x}_i^T. \tag{4.18}$$

The second partial derivatives of the estimators in (4.13), $\frac{\partial^2\hat{\boldsymbol{\theta}}(\mathbf{Y})}{\partial Y_i^2}$, is obtained by differentiating (4.18) as follows:

$$\frac{\partial^2\hat{\boldsymbol{\theta}}(\mathbf{Y})}{\partial Y_i^2} = -\left[X^T\,\mathrm{Var}_{\mathbf{Y}|\boldsymbol{\theta}}[\mathbf{Y}]\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}(\mathbf{Y})}X\right]^{-1}X^T\frac{\partial\mathrm{Var}_{\mathbf{Y}|\boldsymbol{\theta}}[\mathbf{Y}]}{\partial Y_i}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}(\mathbf{Y})}X\left[X^T\,\mathrm{Var}_{\mathbf{Y}|\boldsymbol{\theta}}[\mathbf{Y}]\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}(\mathbf{Y})}X\right]^{-1}\mathbf{x}_i^T \tag{4.19}$$

The term $\frac{\partial \text{Var}_{\mathbf{Y}|\boldsymbol{\theta}}[\mathbf{Y}]}{\partial Y_i}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}(\mathbf{Y})}$ in (4.19) is a diagonal matrix, as we need to differentiate with respect to Y_i. Then,

$$\frac{\partial}{\partial Y_i}\text{Var}_{\mathbf{Y}|\boldsymbol{\theta}}\left[Y_i\right]\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}(\mathbf{Y})} = \frac{\partial \text{Var}_{\mathbf{Y}|\boldsymbol{\theta}}\left[Y_i\right]}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}(\mathbf{Y})} \frac{\partial \hat{\boldsymbol{\theta}}(\mathbf{Y})}{\partial Y_i}. \tag{4.20}$$

The term $\frac{\partial \text{Var}_{\mathbf{Y}|\boldsymbol{\theta}}[Y_i]}{\partial \boldsymbol{\theta}}$ in (4.20) is obtained by applying the chain rule as follows

$$\frac{\partial \text{Var}_{\mathbf{Y}|\boldsymbol{\theta}}\left[Y_i\right]}{\partial \boldsymbol{\theta}} = \frac{\partial \text{Var}_{\mathbf{Y}|\boldsymbol{\theta}}\left[Y_i\right]}{\partial \mu_i}\frac{\partial \mu_i}{\partial \eta_i}\frac{\partial \eta_i}{\partial \boldsymbol{\theta}} = -\frac{g^{''}(\mu_i)}{g^{'}(\mu_i)^2}\frac{1}{g^{'}(\mu_i)}\mathbf{x}_i.$$

Hence,

$$\frac{\partial \text{Var}_{\mathbf{Y}|\boldsymbol{\theta}}\left[Y_i\right]}{\partial \boldsymbol{\theta}} = -\frac{g^{''}(\mu_i)}{g^{'}(\mu_i)^3}\mathbf{x}_i \tag{4.21}$$

By substituting (4.21) and (4.18) in (4.20), we obtain

$$\frac{\partial}{\partial Y_i}\text{Var}_{\mathbf{Y}|\boldsymbol{\theta}}\left[Y_i\right]\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}(\mathbf{Y})} = -\frac{g^{''}(\mu_i)}{g^{'}(\mu_i)^3}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}(\mathbf{Y})} \mathbf{x}_i \left[X^T \text{Var}_{\mathbf{Y}|\boldsymbol{\theta}}\left[\mathbf{Y}\right]\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}(\mathbf{Y})} X\right]^{-1} \mathbf{x}_i^T. \tag{4.22}$$

Thus, the risk function can be approximated via second-order Taylor series for problems involving GLMs.

Under a Poisson regression model, for example, the link function is $g(\mu_i) = \log(\mu_i)$, where $\mu_i = \exp(\eta_i)$. Hence, the quantities $g^{'}(\mu_i)$ and $g^{''}(\mu_i)$ in (4.22) are given by $\frac{1}{\mu_i}$ and $\frac{-1}{\mu_i^2}$, respectively. It is known that, under Poisson regression, the variance of the response $\text{Var}_{\mathbf{Y}|\boldsymbol{\theta}}\left[\mathbf{Y}\right]\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}(\mathbf{Y})}$, which is a diagonal matrix, is equal to the mean of the response $\boldsymbol{\mu} = \exp(X\boldsymbol{\theta})\mid_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}(\mathbf{Y})}$. Given the term $\text{Var}_{\mathbf{Y}|\boldsymbol{\theta}}\left[\mathbf{Y}\right]\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}(\mathbf{Y})}$ allow us to calculate the term $\frac{\partial \hat{\boldsymbol{\theta}}(\mathbf{Y})}{\partial Y_i}$ in (4.18). Substituting the quantities of $g^{'}(\mu_i)$, $g^{''}(\mu_i)$ and $\text{Var}_{\mathbf{Y}|\boldsymbol{\theta}}\left[\mathbf{Y}\right]\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}(\mathbf{Y})}$ in (4.22) allow us to calculate the term $\frac{\partial^2 \hat{\boldsymbol{\theta}}(\mathbf{Y})}{\partial Y_i^2}$ in (4.19). Hence, after calculating these equations, we are able to obtain the term $\frac{\partial^2 \lambda_{qr}(\hat{\boldsymbol{\theta}}(\mathbf{Y}),\boldsymbol{\theta},d)}{\partial Y_i^2}\Big|_{\mathbf{Y}=\boldsymbol{\mu}}$ in (4.14).

4.4 Quadrature approximation

Here, the quadrature approximation is applied to the risk function using a weighted sum, i.e.,

$$R(d;\boldsymbol{\theta}) \approx \sum_{q=1}^{Q} w_q \lambda(\hat{\boldsymbol{\theta}}(\mathbf{Y}(\mathbf{a}_q)), \boldsymbol{\theta}, d), \tag{4.23}$$

where $\mathbf{a}_q$ and w_q are quadrature abscissae and weights, respectively. Here, the responses $\mathbf{Y}$ are a function of the abscissas $\mathbf{a}_q$ where they are generated from a

deterministic function, that is the inverse cumulative distribution function of a given distribution $f(\mathbf{Y}(\mathbf{a}); \boldsymbol{\theta}, d)$. The steps that were used to apply quadrature approximation to the risk function are as follows:

1. Generate $\mathbf{a}_1, \ldots, \mathbf{a}_Q$ abscissae from the quadrature rule that are based on the distribution of $\mathbf{Y}$.
2. Specify true values for unknown parameters $\boldsymbol{\theta}$.
3. Generate a design $d = \{\mathbf{z}_1, \ldots, \mathbf{z}_N\}$ from a given design space.
4. Generate $\mathbf{Y}_1, \ldots, \mathbf{Y}_Q$ responses from the inverse cumulative distribution function of a given distribution $f(\mathbf{Y}(\mathbf{a}); \boldsymbol{\theta}, d)$, given $\mathbf{a}_1, \ldots, \mathbf{a}_Q$, $\boldsymbol{\theta}$ and d.
5. Estimate $\boldsymbol{\theta}$ via MPL estimation given $\mathbf{Y}_1, \ldots, \mathbf{Y}_Q$ and d.
6. Apply the approximation in (4.23) to $R(d; \boldsymbol{\theta})$.

In step 2 of the above algorithm, the true values under the local optimality approach are chosen by guessing the values of $\boldsymbol{\theta}$. For the pseudo-Bayesian approach, those values are the quadrature abscissae $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_H$ from the quadrature rule and based on the prior distribution of $\boldsymbol{\theta}$, see Section 4.1. Note that under the pseudo-Bayesian approach, $Q \times H$ responses are generated from $f(\mathbf{Y}(\mathbf{a}); \boldsymbol{\theta}, d)$ since there are H sets of the true values (i.e. responses are generated for each given set of the true values for the parameters). This results in calculating $Q \times H$ MPL estimates of the parameters.

4.5 Quasi-Monte Carlo approximation

Monte Carlo integration rule is a stochastic approximation (Robert, 2004), simple to implement in practice and a widely-used approximation in a variety of disciplines. In this thesis, we use this approach, where the responses are generated from a deterministic function given by the inverse cumulative distribution function of a given distribution. Hence, the risk function is approximated via quasi-Monte Carlo, i.e.

$$R(d; \boldsymbol{\theta}) \approx \frac{1}{B} \sum_{b=1}^{B} \lambda(\hat{\boldsymbol{\theta}}(\mathbf{Y}(\mathbf{u}_b)), \boldsymbol{\theta}, d), \tag{4.24}$$

where quasi-Monte Carlo approximation would sample $\{\mathbf{u}_b\}_{b=1}^{B}$, $\mathbf{u}_b = (u_1, \ldots, u_N)^T$, of size B randomly from a uniform distribution $U[0, 1]$. The generated sample from $U[0, 1]$ is then fixed. Following that, this sample and a design d are used to generate responses $\mathbf{Y}$ from the inverse cumulative distribution function of a given distribution $f(\mathbf{Y}(\mathbf{u}); \boldsymbol{\theta}, d)$, for each given set of the true values of $\boldsymbol{\theta}$. These modifications are made to minimise Monte Carlo errors, making the approximation more robust. The detailed steps of this approximation method are

1. Generate a sample $\mathbf{u}_1, \ldots, \mathbf{u}_B$ from the uniform distribution $U(0,1)$, and make it fixed.
2. Specify true values for $\boldsymbol{\theta}$.
3. Generate a design $d = \{\mathbf{z}_1, \ldots, \mathbf{z}_N\}$ from a given design space.
4. Generate $\mathbf{Y}_1, \ldots, \mathbf{Y}_B$ responses from the inverse cumulative distribution function of a given distribution $f(\mathbf{Y}(\mathbf{u}); \boldsymbol{\theta}, d)$, given $\mathbf{u}_1, \ldots, \mathbf{u}_B$, $\boldsymbol{\theta}$ and d.
5. Estimate $\boldsymbol{\theta}$ via MPL estimation given $\mathbf{Y}_1, \ldots, \mathbf{Y}_B$ and d.
6. Approximate $R(d; \boldsymbol{\theta})$ via quasi-Monte Carlo given in (4.24).

Similar to quadrature approximation to the risk function, in step 2 the true values under the local optimality approach are chosen by guessing the values of $\boldsymbol{\theta}$, whereas under the pseudo-Bayesian approach those values are the quadrature abscissae $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_H$ from the quadrature rule and based on the prior distribution of $\boldsymbol{\theta}$. Hence, we generate $B \times H$ responses from $f(\mathbf{Y}(\mathbf{u}); \boldsymbol{\theta}, d)$ for the pseudo-Bayesian approach which results in calculating $B \times H$ MPL estimates of the parameters (i.e. again responses are generated for each given set of the true values for the parameters).

A quasi-Monte Carlo 2 is an approximation technique, and uses the fact that the trace is a linear function. It is applied to approximate $R(d; \boldsymbol{\theta})$ when it is a matrix, and we aim to approximate the trace of that matrix. For example, if we aim to approximate the trace of $R_{SE}(\boldsymbol{\theta}, d)$, it is then approximated via quasi-Monte Carlo 2, i.e

$$\begin{aligned} R_{SE}(d; \boldsymbol{\theta}) &= \operatorname{tr}\left\{ \mathrm{E}_{\mathbf{Y}|\boldsymbol{\theta}} \left[(\hat{\boldsymbol{\theta}}(\mathbf{Y}) - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}}(\mathbf{Y}) - \boldsymbol{\theta})^T \right] \right\}, \\ &\approx \frac{1}{B} \sum_{b=1}^{B} \sum_{i=1}^{p} \left(\hat{\theta}_{ii}(\mathbf{Y}(\mathbf{u}_b)) - \theta_{ii} \right)^2. \end{aligned}$$

This approximation is more efficient than QMC when it is used to approximate the trace of the risk function, where there is no need to calculate the approximated risk $R(\boldsymbol{\theta}, d)$, and then calculate the trace of this matrix.

4.6 Quasi-importance sampling approximation

The quasi-importance sampling approximation is used to reduce the computational intensity required in using quasi-Monte Carlo approximation under the pseudo-Bayesian approach, especially for large experimental runs. The consuming computational part of using quasi-Monte Carlo approximation is in estimating the parameters via the MPL method, where $B \times H$ sets of parameters are required to be estimated, under the pseudo-Bayesian approach (see Section 4.5).

Quasi-importance sampling reduces the computational part required to estimate those parameters. In order to apply this approximation, the first three steps of quasi-Monte Carlo approximation remains, and we then generate true values for $\tilde{\boldsymbol{\theta}}_1, \ldots, \tilde{\boldsymbol{\theta}}_B$ from a given prior distribution $\pi(\boldsymbol{\theta})$ of the parameters. Following that, given the fixed uniform sample (see step 1 of quasi-Monte Carlo approximation in Section 4.5), a design d and the generated true values for $\tilde{\boldsymbol{\theta}}_1, \ldots, \tilde{\boldsymbol{\theta}}_B$, the responses are then generated from the inverse cumulative distribution function of a given distribution $f(\mathbf{Y}(\mathbf{u}); \tilde{\boldsymbol{\theta}}, d)$. Thus, the risk function is approximated via quasi-Monte Carlo integration rule (see equation (4.24)), and a multiplicative adjustment to the integrand is needed to account for having sampled from this other distribution $f(\mathbf{Y}(\mathbf{u}); \tilde{\boldsymbol{\theta}}, d)$; we compensate for sampling from the distribution of the responses given $\mathbf{u}_1, \ldots, \mathbf{u}_B$ and $\tilde{\boldsymbol{\theta}}$ instead of $f(\mathbf{Y}(\mathbf{u}); \boldsymbol{\theta}, d)$. The detailed steps of quasi-importance sampling approximation to the risk function are

1. Generate a sample $\mathbf{u}_1, \ldots, \mathbf{u}_B$ from the uniform distribution $U(0, 1)$, and make it fixed.
2. Specify true values for $\boldsymbol{\theta}$, under the pseudo-Bayesian approach those values are the quadrature abscissae $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_H$ that generated from the quadrature rule.
3. Generate true values for $\tilde{\boldsymbol{\theta}}_1, \ldots, \tilde{\boldsymbol{\theta}}_B$ from the prior distribution $\pi(\boldsymbol{\theta})$ of the parameters.
4. Generate $\mathbf{Y}_1, \ldots, \mathbf{Y}_B$ responses from the inverse cumulative distribution function of a given distribution $f(\mathbf{Y}(\mathbf{u}); \tilde{\boldsymbol{\theta}}, d)$, given $\mathbf{u}_1, \ldots, \mathbf{u}_B$, $\tilde{\boldsymbol{\theta}}_1, \ldots, \tilde{\boldsymbol{\theta}}_B$ and d.
5. Estimate $\boldsymbol{\theta}$, a $B \times p$ matrix, via MPL estimation given $\mathbf{Y}_1, \ldots, \mathbf{Y}_B$ and d.
6. Approximate $R(d; \boldsymbol{\theta})$ via quasi-Monte Carlo, and adjust the approximation such that

$$R(d; \boldsymbol{\theta}) \approx \frac{1}{B} \sum_{b=1}^{B} \lambda(\hat{\boldsymbol{\theta}}(\mathbf{Y}(\mathbf{u}_b)), \boldsymbol{\theta}, d) \frac{f(\mathbf{Y}(\mathbf{u}_b); \boldsymbol{\theta}, d)}{f(\mathbf{Y}(\mathbf{u}_b); \tilde{\boldsymbol{\theta}}_b, d)}, \tag{4.25}$$

   where $\frac{f(\mathbf{Y}(\mathbf{u}_b); \boldsymbol{\theta}, d)}{f(\mathbf{Y}(\mathbf{u}_b); \tilde{\boldsymbol{\theta}}_b, d)}$ is the likelihood ratio that makes adjustment to the integrand in the quasi-Monte Carlo approximation.

Note that under the pseudo-Bayesian approach B sets of the parameters are required to be estimated (see step 5) by applying quasi-importance sampling approximation compared to $B \times H$ sets of the parameters for the quasi-Monte Carlo approximation.

4.7 Summary

In this chapter, we introduced the quadrature rule to approximate the Bayesian risk that minimised to find pseudo-Bayesian decision-theoretic optimal designs. Also, we illustrated that the risk function is, typically, analytically intractable, and results from a multidimensional (N-dimensional) integral. In order to handle the intractability of the risk function, some approximations to the risk function were developed. These were quasi-Monte Carlo (QMC), quasi-importance sampling (QIMPS), quadrature (Quad), and second-order Taylor series (ST). The latter approach was generalised to any statistical model, and derived under generalised linear models. We also illustrated the asymptotic approximation to the risk function and demonstrated under the squared error and sign losses. The quadrature approximation to the prior expectation of different approximations to the risk function are demonstrated and summarised in Table 4.1.

Table 4.1: Different approximations to the risk function and the quadrature approximation to the prior expectation of these approximated risk functions.

$\widetilde{R}(d;\boldsymbol{\theta})$	$\widetilde{E}_{\boldsymbol{\theta}}\{\widetilde{R}(d;\boldsymbol{\theta})\}$
$\widetilde{R}_D(d;\boldsymbol{\theta}) = \log\lvert I^{-1}(\boldsymbol{\theta};d)\rvert$	$\sum_{h=1}^{H} w_h \log\lvert I^{-1}(\boldsymbol{\theta}_h;d)\rvert$
$\widetilde{R}_A(d;\boldsymbol{\theta}) = \mathrm{tr}\{I^{-1}(\boldsymbol{\theta};d)\}$	$\sum_{h=1}^{H} w_h \mathrm{tr}\{I^{-1}(\boldsymbol{\theta}_h;d)\}$
$\widetilde{R}_E(d;\boldsymbol{\theta}) = \lambda_{\max}\{I^{-1}(\boldsymbol{\theta};d)\}$	$\sum_{h=1}^{H} w_h \lambda_{\max}\{I^{-1}(\boldsymbol{\theta}_h;d)\}$
$\widetilde{R}_{ST}(d;\boldsymbol{\theta})$ (see equation (4.12))	$\sum_{h=1}^{H} w_h \widetilde{R}_{ST}(d;\boldsymbol{\theta}_h)$
$\widetilde{R}_{Quad}(d;\boldsymbol{\theta}) = \sum_{q=1}^{Q} w_q \lambda(\hat{\boldsymbol{\theta}}(\mathbf{Y}(\mathbf{a}_q)),\boldsymbol{\theta},d)$	$\sum_{h=1}^{H} w_h \sum_{q=1}^{Q} w_q \lambda(\hat{\boldsymbol{\theta}}(\mathbf{Y}(\mathbf{a}_q)),\boldsymbol{\theta}_h,d)$
$\widetilde{R}_{QMC}(d;\boldsymbol{\theta}) = \frac{1}{B}\sum_{b=1}^{B} \lambda(\hat{\boldsymbol{\theta}}(\mathbf{Y}(\mathbf{u}_b)),\boldsymbol{\theta},d)$	$\frac{1}{B}\sum_{h=1}^{H} w_h \sum_{b=1}^{B} \lambda(\hat{\boldsymbol{\theta}}(\mathbf{Y}(\mathbf{u}_b)),\boldsymbol{\theta}_h,d)$
$\widetilde{R}_{QMC2}(d;\boldsymbol{\theta}) = \frac{1}{B}\sum_{b=1}^{B}\sum_{i=1}^{p}\left(\hat{\theta}_{ii}(\mathbf{Y}(\mathbf{u}_b)) - \theta_{ii}\right)^2$	$\frac{1}{B}\sum_{h=1}^{H} w_h \sum_{b=1}^{B}\sum_{i=1}^{p}\left(\hat{\theta}_{ii}(\mathbf{Y}(\mathbf{u}_b)) - \theta_{h_{ii}}\right)^2$
$\widetilde{R}_{QIMPS}(d;\boldsymbol{\theta}) = \frac{1}{B}\sum_{b=1}^{B} \lambda(\hat{\boldsymbol{\theta}}(\mathbf{Y}(\mathbf{u}_b)),\boldsymbol{\theta},d)\frac{f(\mathbf{Y}(\mathbf{u}_b);\boldsymbol{\theta},d)}{f(\mathbf{Y}(\mathbf{u}_b);\tilde{\boldsymbol{\theta}}_b,d)}$	$\frac{1}{B}\sum_{h=1}^{H} w_h \sum_{b=1}^{B} \lambda(\hat{\boldsymbol{\theta}}(\mathbf{Y}(\mathbf{u}_b)),\boldsymbol{\theta}_h,d)\frac{f(\mathbf{Y}(\mathbf{u}_b);\boldsymbol{\theta}_h,d)}{f(\mathbf{Y}(\mathbf{u}_b);\tilde{\boldsymbol{\theta}}_b,d)}$

Chapter 5

An investigation into the Normality of the Parameter Estimators of MLEs and MPLEs

5.1 Introduction

The alphabetical optimal design approach is reliant on the normal approximation to the distribution of the parameter estimators. In this chapter we investigate this approximation for common generalised linear models using simulation studies. These are logistic regression and Poisson regression models. For logistic regression models, recall from Section 2.3.2, the response is modelled by $Y_i \sim Bin(n_i, \pi_i)$ with $\pi_i = 1/[1 + \exp(-\eta_i)]$, and the logistic regression is

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \rho_0 + \sum_{j=1}^{k} \theta_j z_{ij}.$$

Here, three examples are considered as special cases of logistic regression. These are simple logistic regression ($k = 1$), $k = 2$ and $k = 4$ factors logistic regression. Note that for the last two examples we draw only one ($n = 1$) Bernoulli trial for each run (i.e. $Y_i \sim Bin(1, \pi_i)$). For Poisson regression, recall from Section 2.3.3, the response is modelled such that $Y_i \sim \text{Poisson}(\mu_i)$ with $\mu_i = \exp(\eta_i)$, and we consider $k = 2$ factors Poisson regression model given by

$$\log(\mu_i) = \rho_0 + \sum_{j=1}^{2} \theta_j z_{ij}.$$

5.2 Algorithm for assessing the normality of MLEs and MPLEs

In this section, we outline the steps for testing the normality of parameter estimates of MLEs $\hat{\boldsymbol{\theta}}(\mathbf{Y})$ and MPLEs $\hat{\boldsymbol{\theta}}^*(\mathbf{Y})$. These are as follows:

1. Choose a sample size (N or n).
2. Specify true values for unknown parameters $\boldsymbol{\theta}$.
3. Generate a design $d = \{\mathbf{z}_1, \ldots, \mathbf{z}_N\}$.
4. For $j = 1, \ldots, B = 10000$, complete steps a and b, where each iteration forms the ith row of $B \times p$ matrix which consists of the parameter estimates.
   (a) Generate a sample of size N from an appropriate probability distribution $f(\mathbf{Y}; \boldsymbol{\theta}, d)$ (see the logistic regression and Poisson regression models), given d and $\boldsymbol{\theta}$.
   (b) Calculate the estimates of the unknown parameters.

For the simple logistic regression model, the estimates of $\hat{\boldsymbol{\theta}}(\mathbf{Y})$ and $\hat{\boldsymbol{\theta}}^*(\mathbf{Y})$ are available in closed form, given by (2.17) and (2.18), respectively. For $k = 2$ and $k = 4$ factors logistic regression and $k = 2$ factors Poisson regression models, the estimates of $\hat{\boldsymbol{\theta}}(\mathbf{Y})$ are calculated by fitting data generated from the corresponding models using the `glm` function in `R` (R Core Team, 2018), while the estimates of $\hat{\boldsymbol{\theta}}^*(\mathbf{Y})$ are obtained by the adjusted IWLS algorithm (Firth, 1992).

We use the Generalized Shapiro-Wilk test (Gonzalez Estrada and Villasenor Alva, 2009) for multivariate normality to investigate the normality of the distribution of $\hat{\boldsymbol{\theta}}(\mathbf{Y})$ and $\hat{\boldsymbol{\theta}}^*(\mathbf{Y})$ (a test of a null hypothesis H_0: the distribution of the parameter estimators is normal). Note that the test cannot be performed when the size of the rows of the parameter estimates matrix exceeds 5000. Instead, the parameter estimates matrix, after completing step 4 of the above algorithm, is divided into ten matrices, and then the test is performed for each matrix, where each matrix consists of 1000 rows. Hence, ten p-values are obtained, one for each matrix. This algorithm is applied for each chosen sample size, where we obtain 10 p-values for each (N or n). In order to investigate the normality of $\hat{\boldsymbol{\theta}}(\mathbf{Y})$ and $\hat{\boldsymbol{\theta}}^*(\mathbf{Y})$, we plot the resulting p-values against (N or n). Note that the normality of $\hat{\boldsymbol{\theta}}(\mathbf{Y})$ and $\hat{\boldsymbol{\theta}}^*(\mathbf{Y})$ reveals when the p-values for each sample size is uniformly distributed.

5.2.1 Simple logistic regression model

The true values of ρ_0 and θ_1, following Russell et al. (2009a), are 0 and 1, respectively. This allows the linear predictors to be expressed by the values of design points z_1 and z_2 for fixed number of trials $n_1 = n_2$, i.e $\eta_i = z_i$. Hence,

$$\pi_i = 1/[1 + \exp(-z_i)], \qquad i = (1, 2).$$

The design is taken as $d = (-1.5434, 1.5434)^T$, where $z_1 = -1.5434$ and $z_2 = 1.5434$, which is the asymptotic D-optimal design found by Russell et al. (2009a) for $n_1 = n_2 = 5$ using the integrated mean square error as an optimality criterion for design selection.

Figure 5.1 and 5.2 show the p-values of the Shapiro-Wilk test for multivariate normality of $\hat{\boldsymbol{\theta}}(\mathbf{Y})$ and $\hat{\boldsymbol{\theta}}^*(\mathbf{Y})$, respectively, plotted against various values of n where $100 \leq n \leq 2000$. It is noted that the p-values are not uniformly distributed for all n under this example. Hence, the estimates of $\hat{\boldsymbol{\theta}}(\mathbf{Y})$ and $\hat{\boldsymbol{\theta}}^*(\mathbf{Y})$ are not normally distributed for small trials n.

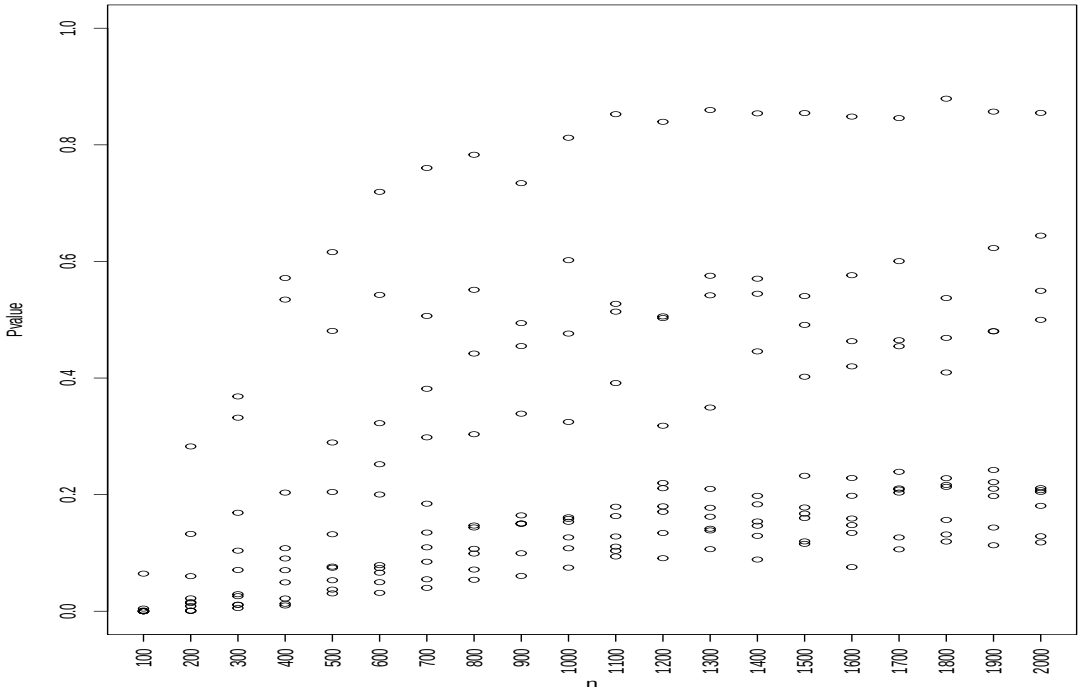

Figure 5.1: Distribution of the p-values for a test of H_0: the MLEs are normally distributed for different sample sizes under simple logistic regression model.

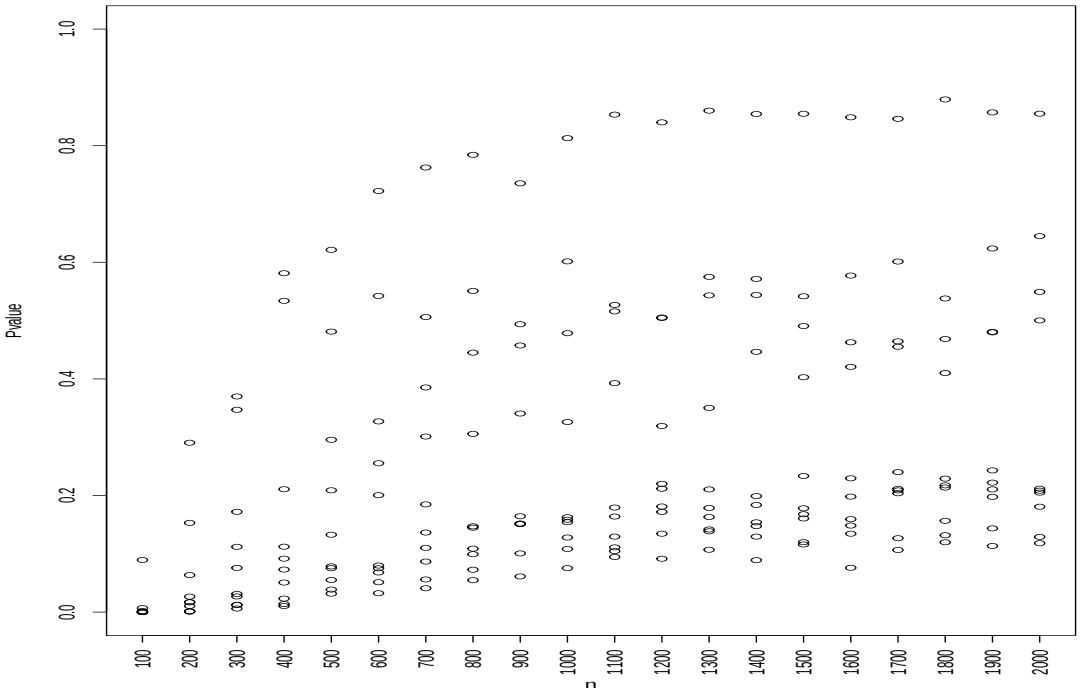

Figure 5.2: Distribution of the p-values for a test of H_0: the MPLEs are normally distributed for different sample sizes under simple logistic regression model.

5.2.2 Two and four factors logistic regression model

Here, for $k = 2$ and $k = 4$ factors logistic regression model, the design d is generated from a uniform distribution $U[-1, 1]$, where the settings z_{ij} of the design d lie in the interval $[-1, 1]$. Following Woods et al. (2006), we assume for each element of unknown parameters an independent uniform prior distribution, i.e

$$\begin{aligned} &\rho_0 \sim U[-3, 3] \qquad \theta_1 \sim U[4, 10] \qquad \theta_2 \sim U[5, 11] \\ &\theta_3 \sim U[-6, 0] \quad \theta_4 \sim U[-2.5, 3.5]. \end{aligned} \tag{5.1}$$

For $k = 2$ and $k = 4$ factors logistic regression, the true values for $\boldsymbol{\theta}$ are the mean of their distributions. These are $\rho_0 = 0, \theta_1 = 7$ and $\theta_2 = 8$ for $k = 2$ factors logistic regression, while for $k = 4$ factors logistic regression the true values for $\boldsymbol{\theta}$ are set to $\rho_0 = 0$, $\theta_1 = 7$, $\theta_2 = 8$, $\theta_3 = -3$ and $\theta_4 = 0.5$.

Under $k = 2$ factors logistic regression, in Figure 5.3 and Figure 5.4, we plot the p-values of the Shapiro-Wilk test for multivariate normality of $\hat{\boldsymbol{\theta}}(\mathbf{Y})$ and $\hat{\boldsymbol{\theta}}^*(\mathbf{Y})$, respectively, against the experimental runs N , where $1000 \leq N \leq 10000$. Similar Figures are plotted in "Appendix B" for $k = 4$ factors logistic regression model. It is concluded that the estimates of $\hat{\boldsymbol{\theta}}(\mathbf{Y})$ and $\hat{\boldsymbol{\theta}}^*(\mathbf{Y})$ are not normally distributed

for small experimental runs N, as the p-values are not uniformly distributed for all N under these examples.

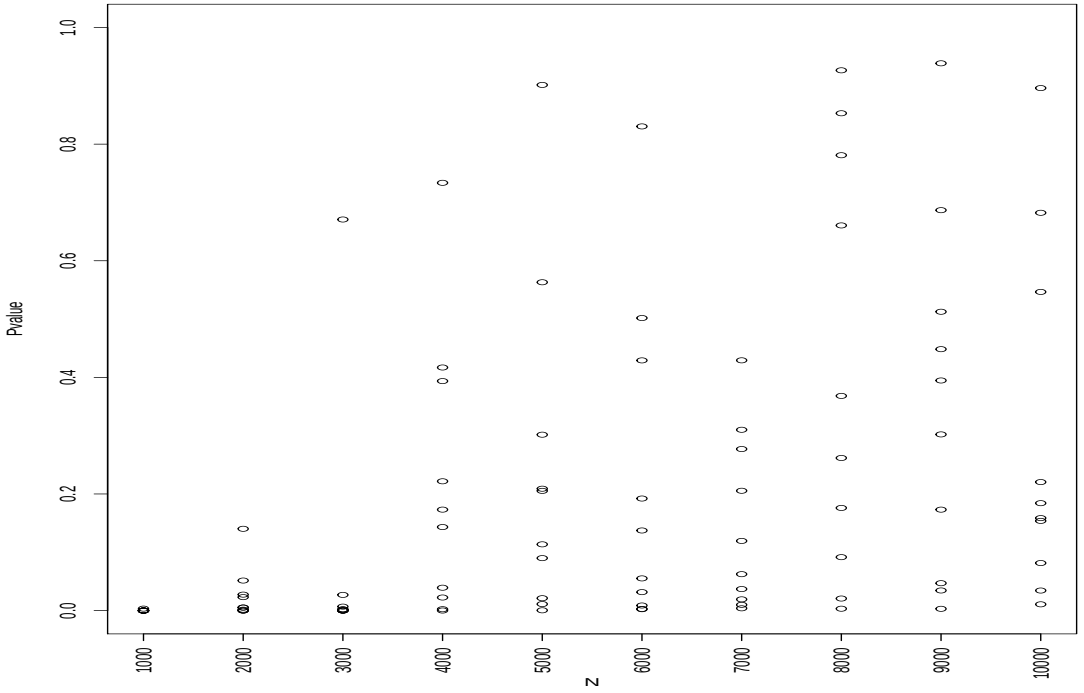

Figure 5.3: Distribution of the p-values for a test of H_0: the MLEs are normally distributed for different sample sizes under two factors logistic regression model.

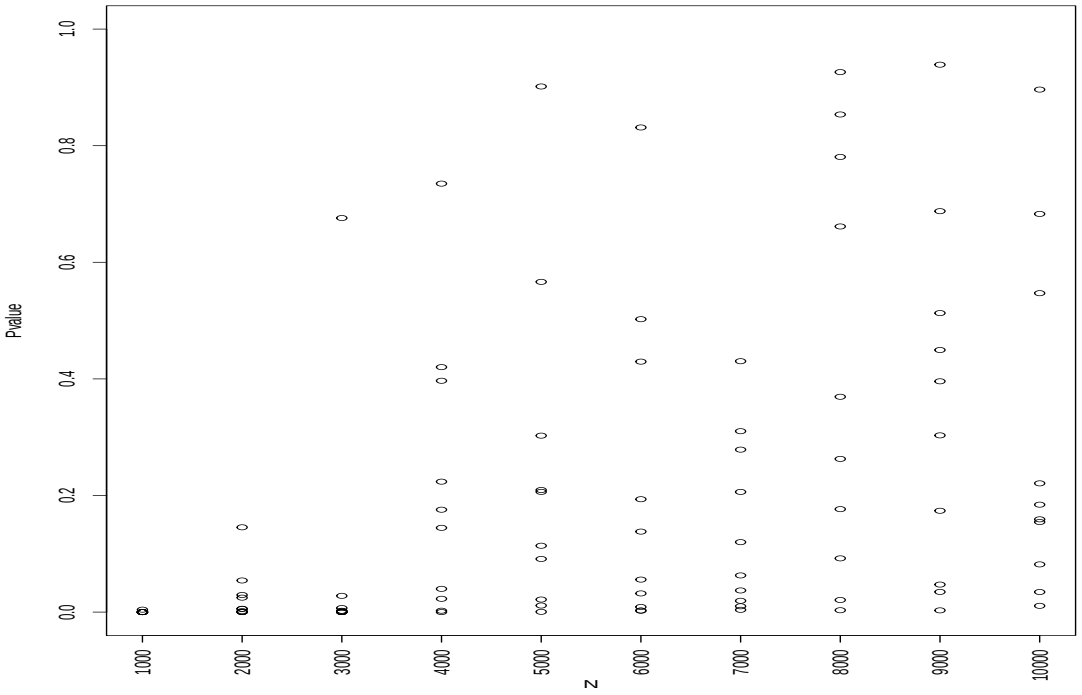

Figure 5.4: Distribution of the p-values for a test of H_0: the MPLEs are normally distributed for different sample sizes under two factors logistic regression model.

5.2.3 Two factors Poisson regression model

The design d under $k = 2$ factors Poisson regression model is also generated from the uniform distribution $U[-1, 1]$. The true values for $\boldsymbol{\theta}$ are set to $\rho_0 = 0.2, \theta_1 = 1.3$ and $\theta_2 = 0.2$ and chosen arbitrary. Figure 5.5 and 5.6 show the p-values of Shapiro-Wilk test for multivariate normality of $\hat{\boldsymbol{\theta}}(\mathbf{Y})$ and $\hat{\boldsymbol{\theta}}^*(\mathbf{Y})$, plotted against different sizes of runs, where $10 \leq N \leq 200$. Similar to two and four factors logistic regression model, the estimates of $\hat{\boldsymbol{\theta}}(\mathbf{Y})$ and $\hat{\boldsymbol{\theta}}^*(\mathbf{Y})$ are not normally distributed for small experimental runs N.

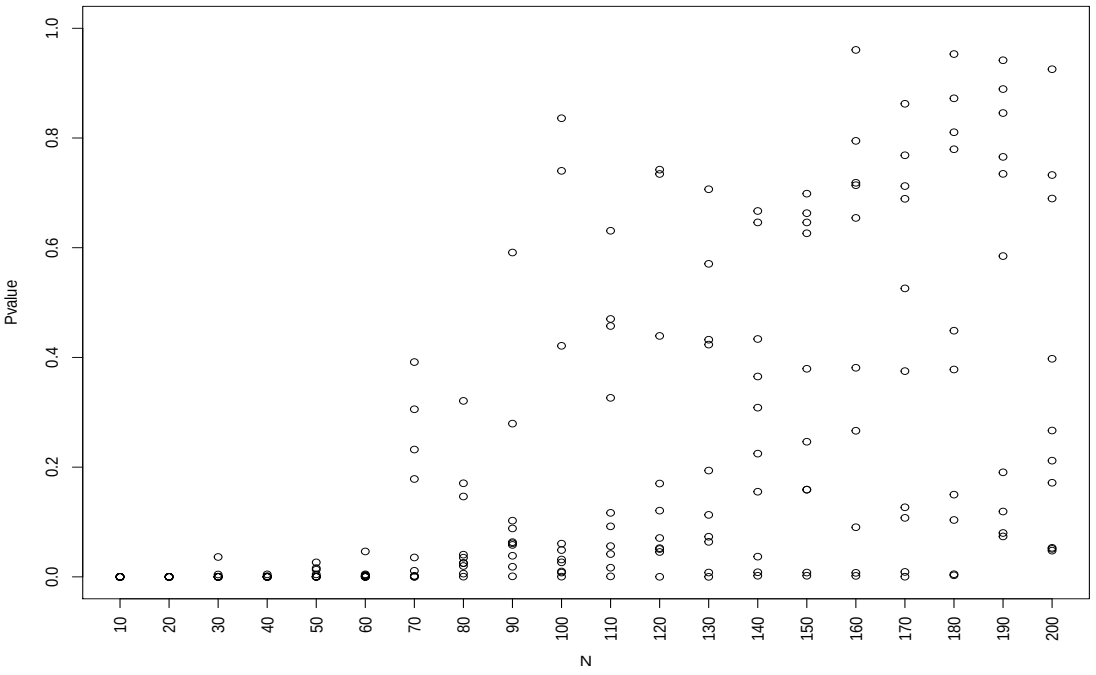

Figure 5.5: Distribution of the p-values for a test of H_0: the MLEs are normally distributed for different sample sizes under Poisson regression model.

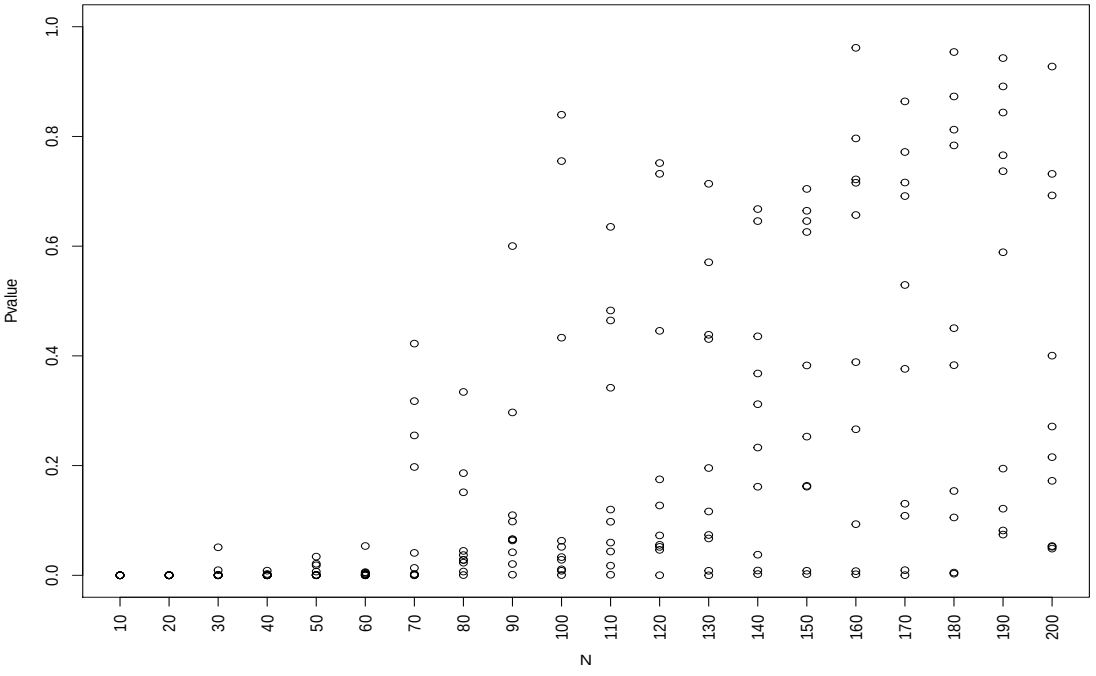

Figure 5.6: Distribution of the p-values for a test of H_0: the MPLEs are normally distributed for different sample sizes under Poisson regression model.

5.2.4 Discussion

From Figure 5.1 to Figure 5.6, we can see that the estimates of $\hat{\boldsymbol{\theta}}(\mathbf{Y})$ and $\hat{\boldsymbol{\theta}}^{*}(\mathbf{Y})$ are not normally distributed for small sample sizes, since the p-values are not uniformly distributed for all (N or n) considered under each example. In a real experiment, we should run a simulation study to make a clear decision allowing us to identify the lowest size of experimental runs. Hence, the recommended size of runs can be considered to find a design under the asymptotic approximation.

We mentioned in the previous chapter that the asymptotic approximation would be used to approximate the analytically intractable risk function in order to find decision-theoretic optimal designs. However, for small sample sizes (either small binomial trials n under simple logistic regression or small experimental runs N under other examples), asymptotic approximation is not suitable to be used to approximate the risk function. This is because the asymptotic approximation depends on the approximate distribution of MLEs $\hat{\boldsymbol{\theta}}(\mathbf{Y})$ that are not normally distributed for small sample sizes; otherwise, it produces poor designs. In the next chapters, we will investigate finding decision-theoretic optimal designs under different approximations to the risk function compared against the asymptotic approximation.

Chapter 6

Decision-theoretic Optimal Designs for Parameter Estimation

6.1 Introduction

In this chapter, we investigate frequentist decision-theoretic optimal designs for experiments involving GLMs, for the experimental aim of parameter estimation under three examples of GLMs, namely simple logistic regression, $k = 4$ factor logistic regression and $k = 2$ factor Poisson regression models. Two losses are considered encapsulating the aim of parameter estimation; these are the squared error (SE) and the sign loss (SL). We find under the SE loss locally and pseudo-Bayesian decision-theoretic optimal designs for the simple logistic regression model (Section 6.3). We also find under the SE and SL pseudo-Bayesian decision-theoretic optimal designs for the $k = 4$ factor logistic regression (Section 6.4) and $k = 2$ factor Poisson regression model (Section 6.5).

For local optimality, the objective function, under the SE loss, is given by a scalar function of the MSE, $S\left(R_{SE}(d;\boldsymbol{\theta}_*)\right)$, whereas for the pseudo-Bayesian optimality the objective function is the expectation of a scalar function of the MSE, $E_{\boldsymbol{\theta}}\{S\left(R_{SE}(d;\boldsymbol{\theta})\right)\}$. We consider three different scalar functions S given in Section 3.2.1; these are the log determinant, trace and maximum eigenvalue, where the resulting designs are locally or pseudo-Bayesian decision-theoretic D-, A- and E-optimal designs, respectively. Under the sign loss, the objective function for the pseudo-Bayesian optimality is given by the expectation of the expected SL, $E_{\boldsymbol{\theta}}\{R_{SL}(d;\boldsymbol{\theta})\}$.

It is known that finding exact designs is currently difficult and an infeasible objective due to the considerable computational challenges of minimising an analytically intractable objective function, given by the optimality criteria, over a

multi-dimensional design space. Therefore, we aim to find a design near the optimal design by approximating the objective function.

We apply the Quad approximation to the prior expectation of the risk function and some approximations to the risk functions (as described in Chapter 4) to find pseudo-Bayesian designs. Comparison between these approximations is introduced in terms of their performance (in terms of the risk function for finding locally optimal designs or the expected risk for pseudo-Bayesian designs) and computing time (CPU time) (in terms of the relative computing time (RT) for finding optimal designs).

6.2 Computational methods to find optimal designs

For the simple logistic regression model, searching for local and pseudo-Bayesian decision-theoretic optimal designs is conducted in two phases. Phase I uses a grid search method, where we place a grid over the design space $\chi = [-6, 6]$ of spacing 0.1 within the design region that determines the row z_1 and column z_2 labels which constitute the design space, referred to as the design grid. For each unique support point pair (z_1, z_2) in the design grid, we evaluate the objective function at each pair. After that, we search for the lowest value of the objective function in the grid and note its row and column coordinates (z_1^*, z_2^*). Phase II considers the row and column coordinates of the lowest value of the objective function found in phase I, (z_1^*, z_2^*), as the start values that would be used to minimise the objective function using the modified quasi-Newton method ("L-BFGS-B") implemented in `optim` function in `R` (R Core Team, 2018).

For $k = 4$ factor logistic regression and $k = 2$ factor Poisson regression models, the objective functions are coded using a low-level programming language (C++) and make use of packages **Rcpp** (Eddelbuettel and François, 2011) and **RcppArmadillo** (Eddelbuettel and Sanderson, 2014). Finding pseudo-Bayesian decision-theoretic optimal designs is accomplished by applying the methodology of the approximate coordinate exchange algorithm (ACE) (see "Appendix A" for details).

6.3 Simple logistic regression model

For the simple logistic regression model with $N = 2$ design points, it is possible to calculate the exact MSE for such simple example, for a moderate fixed number of

binomial trials, which is given by

$$R_{SE}(\boldsymbol{\theta}, d) = \sum_{Y_2=0}^{n_2} \sum_{Y_1=0}^{n_1} \lambda_{SE}(\hat{\boldsymbol{\theta}}(Y_1, Y_2), \boldsymbol{\theta}, d) f(Y_1; \boldsymbol{\theta}, d) f(Y_2; \boldsymbol{\theta}, d), \tag{6.1}$$

where $f(Y_1; \boldsymbol{\theta}, d)$ and $f(Y_2; \boldsymbol{\theta}, d)$ refer to probability mass functions of binomial random variables Y_1 and Y_2, respectively. Calculating the exact MSE, $R_{SE}(\boldsymbol{\theta}, d)$, aids to make a comparison between proposed approximation methods, where the approximation's accuracy is compared by how close the approximation's results is to the exact MSE's results.

We use five different approximations to the MSE, and these are

- Second-order Taylor series

$$\widetilde{R}_{ST}(d; \boldsymbol{\theta}) \text{ (see equation (4.12))}.$$

- Asymptotic

$$\widetilde{R}_{SE}(d; \boldsymbol{\theta}) = S\left(I^{-1}(\boldsymbol{\theta}; d)\right).$$

- Quadrature

$$\widetilde{R}_{Quad}(d; \boldsymbol{\theta}) = \sum_{q=1}^{Q} w_q \lambda(\hat{\boldsymbol{\theta}}(\mathbf{Y}(\mathbf{a}_q)), \boldsymbol{\theta}, d).$$

- Quasi-Monte Carlo with $B = 5,000$

$$\widetilde{R}_{QMC}(d; \boldsymbol{\theta}) = \frac{1}{B} \sum_{b=1}^{B} \lambda(\hat{\boldsymbol{\theta}}(\mathbf{Y}(\mathbf{u}_b)), \boldsymbol{\theta}, d).$$

- Quasi-importance sampling with $B = 5,000$

$$\widetilde{R}_{QIMPS}(d; \boldsymbol{\theta}) = \frac{1}{B} \sum_{b=1}^{B} \lambda(\hat{\boldsymbol{\theta}}(\mathbf{Y}(\mathbf{u}_b)), \boldsymbol{\theta}, d) \frac{f(\mathbf{Y}(\mathbf{u}_b); \boldsymbol{\theta}, d)}{f(\mathbf{Y}(\mathbf{u}_b); \tilde{\boldsymbol{\theta}}_b, d)}.$$

The first four approximations are used for locally decision-theoretic optimal designs. For pseudo-Bayesian decision-theoretic optimal designs these approximations are used in addition to quasi-importance sampling with $B = 5,000$ to approximate the MSE and the prior expectation is approximated via quadrature methods such that

- $\widetilde{E}_{\boldsymbol{\theta}}\{\widetilde{R}_{ST}(d; \boldsymbol{\theta})\} = \sum_{h=1}^{H} w_h \widetilde{R}_{ST}(d; \boldsymbol{\theta}_h)$.
- $\widetilde{E}_{\boldsymbol{\theta}}\{\widetilde{R}_{SE}(d; \boldsymbol{\theta})\} = \sum_{h=1}^{H} w_h S\left(I^{-1}(\boldsymbol{\theta}_h; d)\right)$.
- $\widetilde{E}_{\boldsymbol{\theta}}\{\widetilde{R}_{Quad}(d; \boldsymbol{\theta})\} = \sum_{h=1}^{H} w_h \sum_{q=1}^{Q} w_q \lambda(\hat{\boldsymbol{\theta}}(\mathbf{Y}(\mathbf{a}_q)), \boldsymbol{\theta}_h, d)$.
- $\widetilde{E}_{\boldsymbol{\theta}}\{\widetilde{R}_{QMC}(d; \boldsymbol{\theta})\} = \frac{1}{5,000} \sum_{h=1}^{H} w_h \sum_{b=1}^{5,000} \lambda(\hat{\boldsymbol{\theta}}(\mathbf{Y}(\mathbf{u}_b)), \boldsymbol{\theta}_h, d)$.

- $\widetilde{E}_{\boldsymbol{\theta}}\{\widetilde{R}_{QIMPS}(d;\boldsymbol{\theta})\} = \frac{1}{5{,}000}\sum_{h=1}^{H} w_h \sum_{b=1}^{5{,}000} \lambda(\hat{\boldsymbol{\theta}}(\mathbf{Y}(\mathbf{u}_b)), \boldsymbol{\theta}_h, d)\frac{f(\mathbf{Y}(\mathbf{u}_b);\boldsymbol{\theta}_h,d)}{f(\mathbf{Y}(\mathbf{u}_b);\tilde{\boldsymbol{\theta}}_b,d)}$.

Designs found under these approximations (referred henceforth as ST, Asym, Quad, QMC and QIMPS designs, respectively) are compared against a design found under the exact MSE (henceforth referred to as True designs). The relative efficiency is also used to aid in comparison between designs against design found under asymptotic approximation (alphabetical optimal designs), i.e

$$\text{Efficiency } (\%) = 100 \times \frac{R_{SE}(\boldsymbol{\theta}, d_{Asym})}{R_{SE}(\boldsymbol{\theta}, d)},$$

where d_{Asym} is a design found under the asymptotic approximation, and d refer to other designs found under this example. The relative efficiency is calculated by the exact MSE in (6.1).

In order to find locally decision-theoretic optimal designs, the values of the unknown parameters ρ_0 and θ_1 are 0 and 1, respectively, following Russell et al. (2009a). Each element of the design d is assumed to be within the interval $[-6, 6]$. For pseudo-Bayesian decision-theoretic optimal designs, two different independent prior distributions for ρ_0 and θ_1 are considered, i.e.

$$\pi_1(\boldsymbol{\theta}): \quad \rho_0 \sim U[-0.5, 2] \quad \theta_1 \sim U[0, 2],$$
$$\pi_2(\boldsymbol{\theta}): \qquad \rho_0 \quad \text{and} \quad \theta_1 \sim U[-2, 2].$$

Locally and pseudo-Bayesian decision-theoretic optimal designs will be found for different numbers of binomial trials n made at the jth run ($j = 1, 2$), assuming $n_1 = n_2$. For each size of n, the comparison between the approximated designs is made by evaluating the exact MSE $S\left(R_{SE}(\boldsymbol{\theta}, d)\right)$ at the resulting designs under local optimality, or by evaluating the approximate expectation of the exact MSE, $\widetilde{E}_{\boldsymbol{\theta}}\{S\left(R_{SE}(\boldsymbol{\theta}, d)\right)\}$ at the resulting designs under pseudo-Bayesian optimality. Hence, we compare the approximated decision-theoretic D-, A- or E-optimal design against the corresponding exact design.

We check the validity of the proposed approximations to the MSE, $S\left(\widetilde{R}_{SE}(\boldsymbol{\theta}, d)\right)$, to ensure that it is asymptotically equivalent to $S\left(R_{SE}(\boldsymbol{\theta}, d)\right)$. Following Overstall et al. (2017b), we randomly generate $F = 1000$ designs under SE loss, d_{SE}, and calculate the quantity of $S\left(R_{SE}(\boldsymbol{\theta}, d_{SE})\right)$ and $S\left(\widetilde{R}_{SE}(\boldsymbol{\theta}, d_{SE})\right)$. The fth design, $f = 1, \ldots, 1000$, is generated by perturbing the True design d_T found under the exact MSE $S\left(R_{SE}(\boldsymbol{\theta}, d)\right)$ at each iteration, i.e.

$$d_{SE}^{(f)} = (1 - u_f)d_T + u_f d^{(f)},$$

where $u_f \sim U[0, \frac{1}{2}]$ and $d^{(f)}$ is a design in which each element is randomly generated from $U[-1, 1]$. Then, we plot the quantity $S\left(\widetilde{R}_{SE}(\boldsymbol{\theta}, d_{SE})\right)$ against $S\left(R_{SE}(\boldsymbol{\theta}, d_{SE})\right)$ to investigate the validity of the proposed approximations to the MSE. Similarly, under pseudo-Bayesian optimality, the validity of the Quad approximation to the expectation of an approximation to the MSE, $\widetilde{E}_{\boldsymbol{\theta}}\left\{S\left(\widetilde{R}_{SE}(\boldsymbol{\theta}, d_{SE})\right)\right\}$, is checked to ensure that it is asymptotically equivalent to the Quad approximation to the expectation of the exact MSE, $\widetilde{E}_{\boldsymbol{\theta}}\left\{S\left(R_{SE}(\boldsymbol{\theta}, d_{SE})\right)\right\}$. Hence, these quantities are calculated, where the fth design is generated by perturbing the True design d_T (which here found under $\widetilde{E}_{\boldsymbol{\theta}}\left\{S\left(R_{SE}(\boldsymbol{\theta}, d)\right)\right\}$) as explained above.

The RT (i.e. the CPU time required for each approximation / the CPU time required for the exact MSE) is calculated to aid in the comparison between CPU time required for each approximation. Here, the CPU time required to find an optimal design is obtained by adding the CPU time required for the grid search method to the CPU time required for minimising the objective function using `optim` function.

6.3.1 Locally and pseudo-Bayesian decision-theoretic D-optimal designs

6.3.1.1 Locally decision-theoretic D-optimal designs

We find locally decision-theoretic D-optimal designs for different sizes of n under the exact and different approximations to $R_{SE}(\boldsymbol{\theta}, d)$. Table 6.1 shows the abscissas $\mathbf{a}$ and weights $\mathbf{w}$ used under the quadrature approximation to $R_{SE}(\boldsymbol{\theta}, d)$, given in 4.23, for finding the optimal design with $n = 5$ binomial trials. Note that the number of abscissas for $n_R = 3$ radii and n_Q random rotations used for the quadrature approximation is 73, and in Table 6.1 we introduce the first 40 abscissas and weights.

Figure 6.1 shows plot of the exact $\log|R_{SE}(\boldsymbol{\theta}, d)|$ evaluated at the resulting optimal designs and plotted against n. It is noted that locally D-optimal design found under the Asym approximation (red star), $\widetilde{R}_D(d; \boldsymbol{\theta}) = \log|I^{-1}(\boldsymbol{\theta}; d)|$, is far from the True design (light blue triangle) for small n, indicating inaccurate results under this approximation for small n and the resulting design can be poor. As the size of n increases, locally D-optimal design attains improvement and it is near True design. The ST (black square) and Quad (green plus) designs are poorly performed compared to the other designs for small and moderate n. We can see that the QMC approximation is the most accurate approximation to $\log|R_{SE}(\boldsymbol{\theta}, d)|$, since QMC design (yellow filled circle) is identical to True design for all sizes of n.

Although the QMC design has performed better than others in terms of closeness to the True design, it is noted that a slight increase occurs in the exact $\log|R_{SE}(\boldsymbol{\theta}, d)|$

Table 6.1: The abscissas **a** and weights **w** found under the quadrature approximation to $R_{SE}(\boldsymbol{\theta}, d)$ under the simple logistic regression with $n = 5$ binomial trials.

q	a		w
1	0.50	0.50	0.25
2	0.50	0.91	0.04
3	0.24	0.88	0.04
4	0.88	0.25	0.04
5	0.91	0.50	0.04
6	0.12	0.25	0.04
7	0.25	0.12	0.04
8	0.50	0.09	0.04
9	0.76	0.12	0.04
10	0.12	0.75	0.04
11	0.09	0.50	0.04
12	0.88	0.75	0.04
13	0.75	0.88	0.04
14	0.88	0.75	0.01
15	0.75	0.88	0.01
16	0.12	0.75	0.01
17	0.09	0.50	0.01
18	0.50	0.09	0.01
19	0.76	0.12	0.01
20	0.12	0.25	0.01
21	0.25	0.12	0.01
22	0.88	0.25	0.01
23	0.91	0.50	0.01
24	0.50	0.91	0.01
25	0.24	0.88	0.01
26	0.50	0.99	0.01
27	0.10	0.99	0.01
28	0.99	0.10	0.01
29	0.99	0.51	0.01
30	0.01	0.10	0.01
31	0.10	0.01	0.01
32	0.50	0.01	0.01
33	0.90	0.01	0.01
34	0.01	0.90	0.01
35	0.01	0.49	0.01
36	0.99	0.90	0.01
37	0.90	0.99	0.01
38	0.99	0.90	0.002
39	0.90	0.99	0.002
40	0.01	0.90	0.002

under QMC and True designs, and a sharp increase under ST and Quad designs for small n. This increase requires an investigation to justify the unacceptable increase in the exact $\log|R_{SE}(\boldsymbol{\theta}, d)|$. The relative D-efficiency is plotted in Figure 6.2 against n. From this figure, we can see that the pseudo-Bayesian D-optimal design is less efficient for small n compared with other designs, and it becomes more efficient as the size of n increases. For large n, all designs found under this example have the same efficiency.

The increase that occurred in the exact $\log|R_{SE}(\boldsymbol{\theta}, d)|$ is investigated by plotting the contour plots of all approximations considered to $\log|R_{SE}(\boldsymbol{\theta}, d)|$ in addition to the exact $\log|R_{SE}(\boldsymbol{\theta}, d)|$ for $n = (5, 15, 100)$. Figure 6.3, 6.4 and 6.5 show five

contour plots for $n = 5, 15$ and 100, respectively. These are a contour plot of the exact $\log|R_{SE}(\boldsymbol{\theta}, d)|$ (the first row on left), and the approximated $\log|\widetilde{R}_{SE}(\boldsymbol{\theta}, d)|$ via ST approximation (the first row on right), Asym approximation (the second row on left), Quad approximation (the second row on right) and QMC approximation with $B = 5,000$ (the third row). For each contour plot, there is a plot key on the right hand side of the plot, which explains the contour plot and the values of $\log|R_{SE}(\boldsymbol{\theta}, d)|$ or $\log|\widetilde{R}_{SE}(\boldsymbol{\theta}, d)|$ that we aim to minimise. In addition, there are five different coloured shapes mentioned above, on each plot, which refer to the exact or approximated locally decision-theoretic D-optimal designs that minimise $\log|R_{SE}(\boldsymbol{\theta}, d)|$ or $\log|\widetilde{R}_{SE}(\boldsymbol{\theta}, d)|$, respectively. Each contour plot is symmetric around the line $z_1 = z_2$, and the designs are symmetric as well due to the interchangeability of z_1 and z_2; hence, we restrict our search for the optimal design to $z_1 < z_2$.

In Figure 6.3, the contour plot of the exact $\log|R_{SE}(\boldsymbol{\theta}, d)|$ shows that the resulting designs found under ST, Quad and QMC approximations to $R_{SE}(\boldsymbol{\theta}, d)$ are very similar to the True design for a sample of size $n = 5$, where the True design is $d_T = (-6, 6)^T$ and $\log|R_{SE}(\boldsymbol{\theta}, d_T)| = -5.58$. However, Quad and ST designs change unlike other designs in the first row of Figure 6.4 (left) for $n = 15$, and this change may cause the noticeable sharp increase in the exact $\log|R_{SE}(\boldsymbol{\theta}, d)|$ for each of them at $n = 15$. In addition, these designs, as the size of n gets larger, return to be similar to the True design (see the first row of Figure 6.5 on left).Therefore, we need to check the validity of the proposed approximation to the MSE.

In Figure 6.6, the validity of the QMC, Quad and ST approximations to $\log|R_{SE}(\boldsymbol{\theta}, d)|$ are checked (i.e. large values of n are considered ($n = 100$)) by plotting each of these approximations against the exact $\log|R_{SE}(\boldsymbol{\theta}, d)|$. Apparently, the QMC and Quad approximations to $\log|R_{SE}(\boldsymbol{\theta}, d)|$ (the first and second rows of Figure 6.6, respectively) appear very accurate. We can see that the ST approximation (the third row of Figure 6.6) can be inaccurate, especially for designs close to the minimum, but the ordering of designs in terms of $\log|\widetilde{R}_{SE}(\boldsymbol{\theta}, d)|$ is the same as that for $\log|R_{SE}(\boldsymbol{\theta}, d)|$. This indicates that the ST approximation to $\log|R_{SE}(\boldsymbol{\theta}, d)|$ is minimised for a design near the design that minimises the exact $\log|R_{SE}(\boldsymbol{\theta}, d)|$.

The justification of the increase in the exact $\log|R_{SE}(\boldsymbol{\theta}, d)|$ for small n is still required. In Figure 6.7 (the first row on the left), we investigate the $R_{SE}(\boldsymbol{\theta}, d)$ for each model parameter (ρ_0 and θ_1) that was calculated for the True designs found for each n and that was plotted against n. It is clear that the increase in the exact $\log|R_{SE}(\boldsymbol{\theta}, d)|$ for small n is dominated by the MSE of the intercept, $R_{SE}(\rho_0, d_T)$, which increases for small n and then decreases gradually, whereas the MSE of the

slope parameter, $R_{SE}(\theta_1, d_T)$, decreases as n gets larger. This conclusion is confirmed by plotting the exact $\log|R_{SE}(\boldsymbol{\theta}, d_T)|$ for those True designs against n; see the first row of Figure 6.7 on the right. It is apparent that the increase occurs for small n and then decreases gradually similar to the plot of $R_{SE}(\rho_0, d)$.

With further investigation, it is found that this increase occurs when the responses $\mathbf{Y} = (0, 0)$ or $\mathbf{Y} = (n, n)$, where under these responses, the slope parameter θ_1 is always zero, as we have the same responses for both z_1 and z_2. We calculate the estimates of ρ_0 and θ_1 under these responses, and plot them against a range of size n (see Figure C.1 in "Appendix C"). These estimates cause the unacceptable increase in $\log|R_{SE}(\boldsymbol{\theta}, d)|$.

Plots similar to the first row of Figure 6.7 are plotted in the second row of this figure, but the quantities are calculated for the ST designs found for each n. The same comments apply to these plots, confirming our conclusion about the undesirable increase that occurred in the exact $\log|R_{SE}(\boldsymbol{\theta}, d)|$.

The RT for each approximation is calculated, and found that the QMC approximation is the most expensive approximation to the MSE with RT= 41.40, indicating that the CPU time required for the QMC approximation is approximately 41 times more than the CPU time required for the exact MSE. However, other approximations to the MSE require less executing time than the exact MSE (the CPU time for the exact MSE is 5.084), where the locally D-optimal designs found under the Asym approximation requires the smallest amount of the CPU time with RT= 0.05.

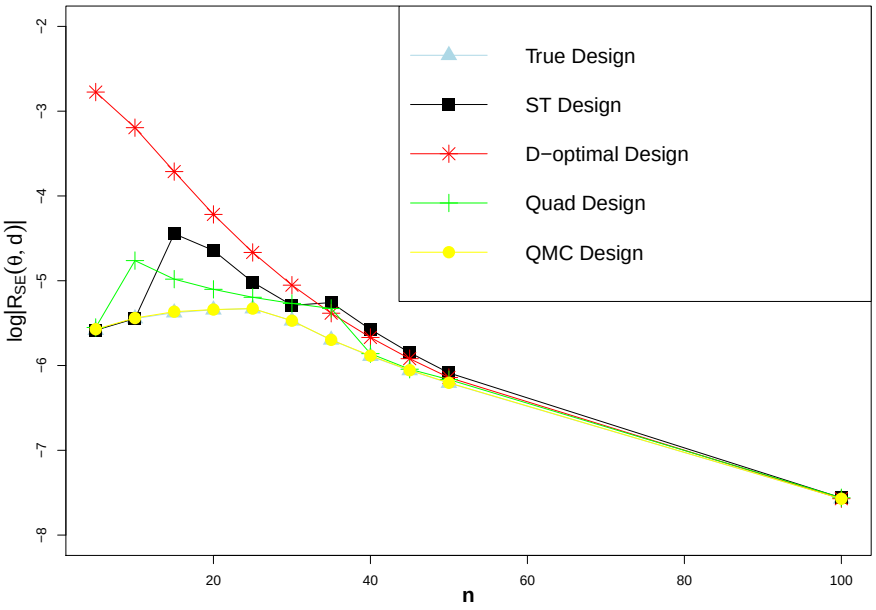

Figure 6.1: Plots of the exact $\log|R_{SE}(\boldsymbol{\theta}, d)|$ evaluated at the True and approximated designs, and plotted against different sizes of n for simple logistic regression model.

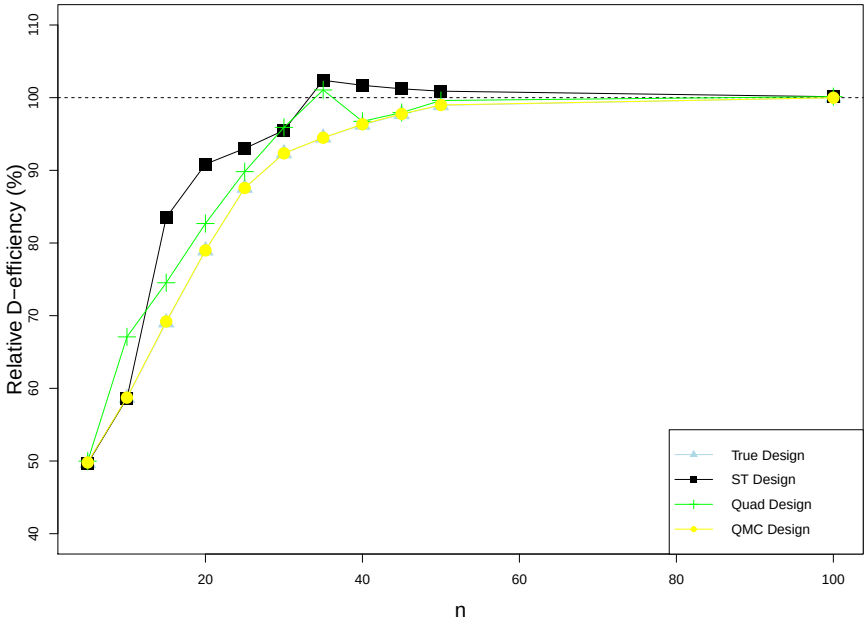

Figure 6.2: Plots of the relative D-efficiency for the True and approximated designs, and plotted against different sizes of n for simple logistic regression model.

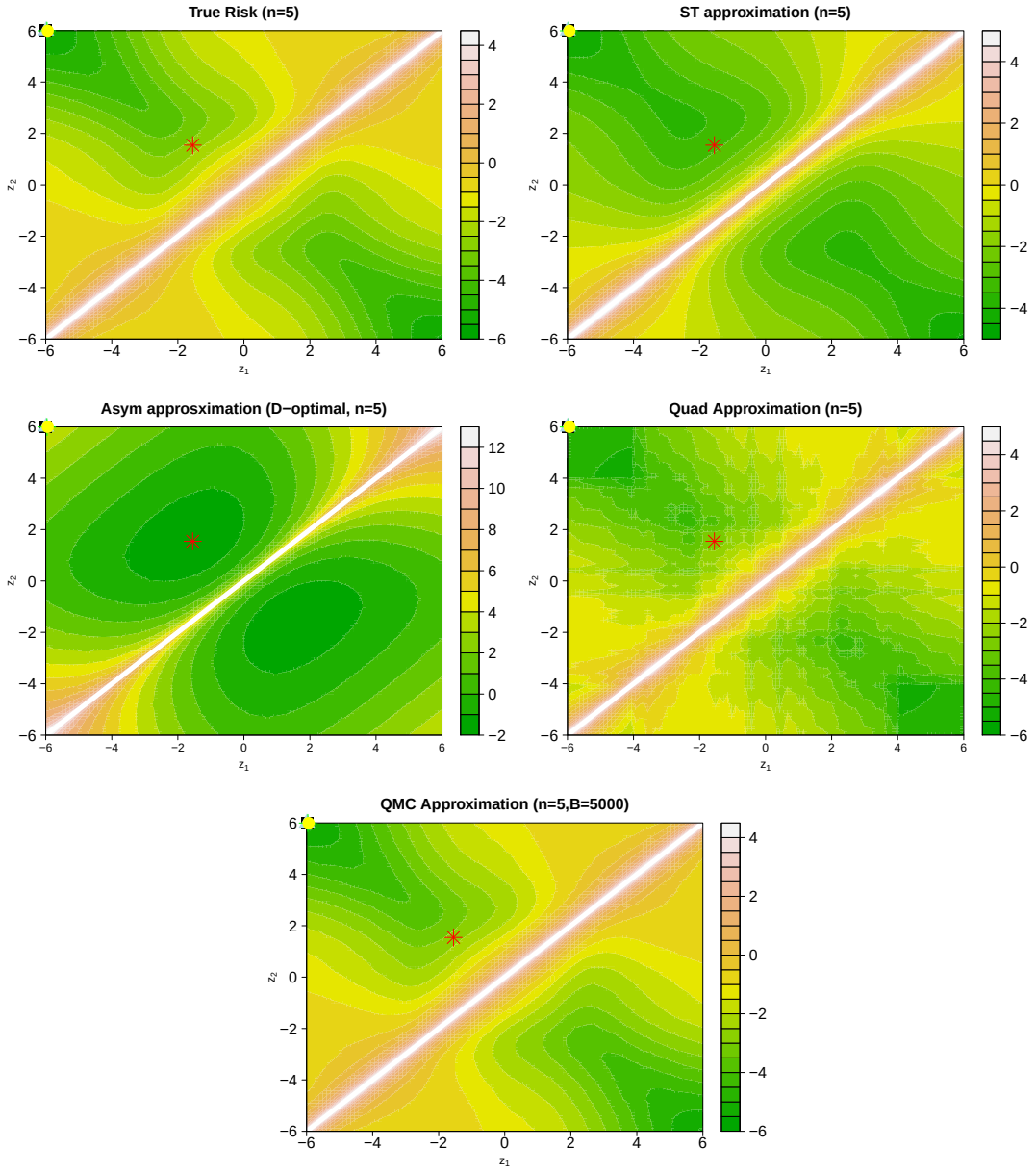

Figure 6.3: Contour plots of the exact $\log|R_{SE}(\boldsymbol{\theta}, d)|$ (the first row on left) and approximated MSE $\log|\widetilde{R}_{SE}(\boldsymbol{\theta}, d)|$ via ST (the first row on right), Asym (the second row on left), Quad (the second row on right) and QMC with size $B = 5,000$ (the third row), under simple logistic regression model for a sample of size $n = 5$. A plot key for each contour is plotted which refers to the values of $\log|R_{SE}(\boldsymbol{\theta}, d)|$ and $\log|\widetilde{R}_{SE}(\boldsymbol{\theta}, d)|$ for the remaining contour plots. In Each plot, there are five coloured-shapes denote the decision-theoretic D-optimal designs found under the exact risk (light blue triangle), ST(black square), Asym (red star), Quad (green plus) and QMC approximation (yellow filled circle). Here, True, ST, Quad and QMC designs are the same, where $d = (-6, 6)$.

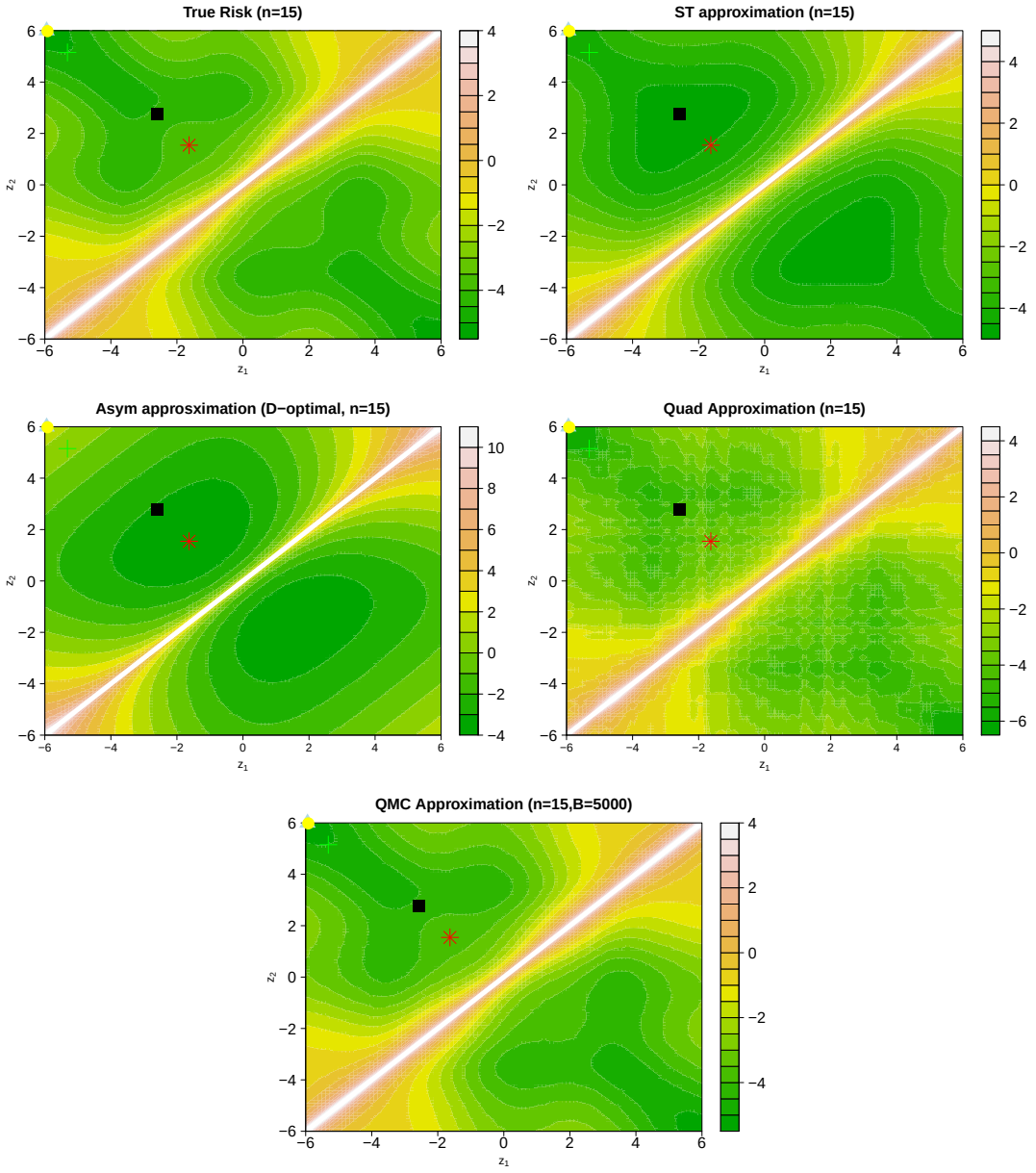

Figure 6.4: Contour plots of the exact $\log|R_{SE}(\boldsymbol{\theta}, d)|$ (the first row on left) and approximated MSE $\log|\widetilde{R}_{SE}(\boldsymbol{\theta}, d)|$ via ST (the first row on right), Asym (the second row on left), Quad (the second row on right) and QMC with size $B = 5,000$ (the third row), under simple logistic regression model for a sample of size $n = 15$. A plot key for each contour is plotted which refers to the values of $\log|R_{SE}(\boldsymbol{\theta}, d)|$ and $\log|\widetilde{R}_{SE}(\boldsymbol{\theta}, d)|$ for the remaining contour plots. In Each plot, there are five coloured-shapes denote the decision-theoretic D-optimal designs found under the exact risk (light blue triangle), ST(black square), Asym (red star), Quad (green plus) and QMC approximation (yellow filled circle). Here, True, ST, Quad and QMC designs are the same, where $d = (-6, 6)$.

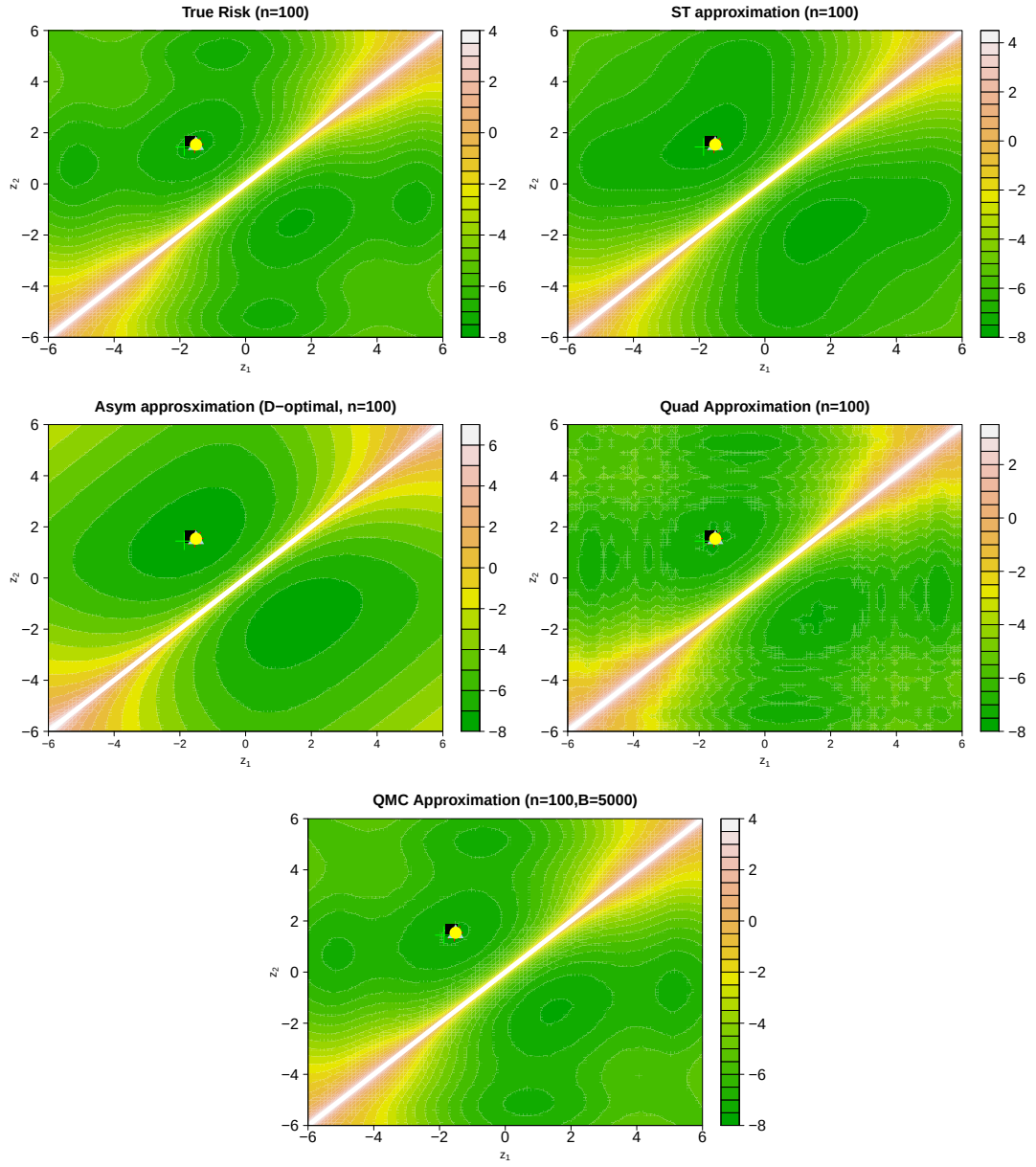

Figure 6.5: Contour plots of the exact $\log|R_{SE}(\boldsymbol{\theta}, d)|$ (the first row on left) and approximated MSE $\log|\widetilde{R}_{SE}(\boldsymbol{\theta}, d)|$ via ST (the first row on right), Asym (the second row on left), Quad (the second row on right) and QMC with size $B = 5,000$ (the third row), under simple logistic regression model for a sample of size $n = 100$. A plot key for each contour is plotted which refers to the values of $\log|R_{SE}(\boldsymbol{\theta}, d)|$ and $\log|\widetilde{R}_{SE}(\boldsymbol{\theta}, d)|$ for the remaining contour plots. In Each plot, there are five coloured-shapes denote the decision-theoretic D-optimal designs found under the exact risk (light blue triangle), ST(black square), Asym (red star), Quad (green plus) and QMC approximation (yellow filled circle). Here, True, ST, Quad and QMC designs are the same, where $d = (-6, 6)$.

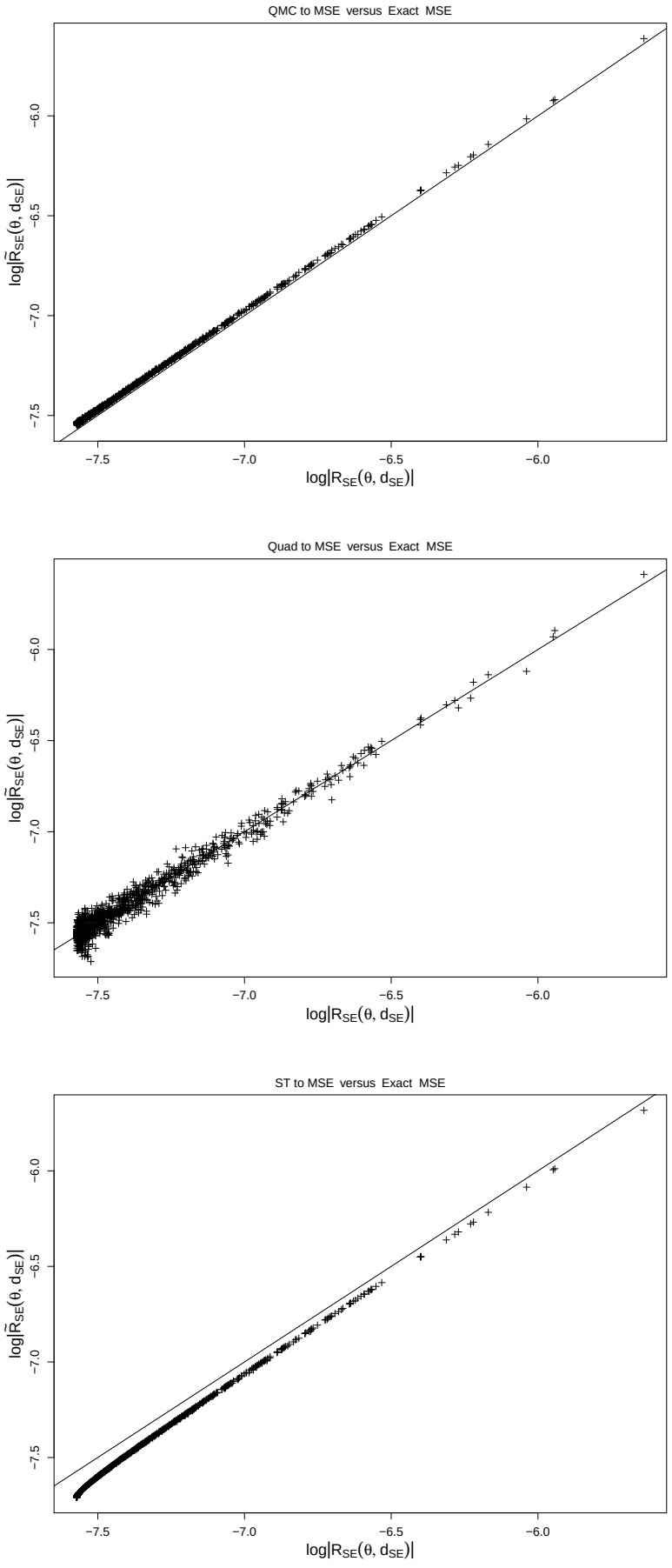

Figure 6.6: Plots of the QMC (the first row), Quad (the second row) and ST (the third row) approximation to the MSE $\log|\widetilde{R}_{SE}(\boldsymbol{\theta}, d)|$ plotted against the exact MSE $\log|R_{SE}(\boldsymbol{\theta}, d)|$ for simple logistic regression model for $n = 100$.

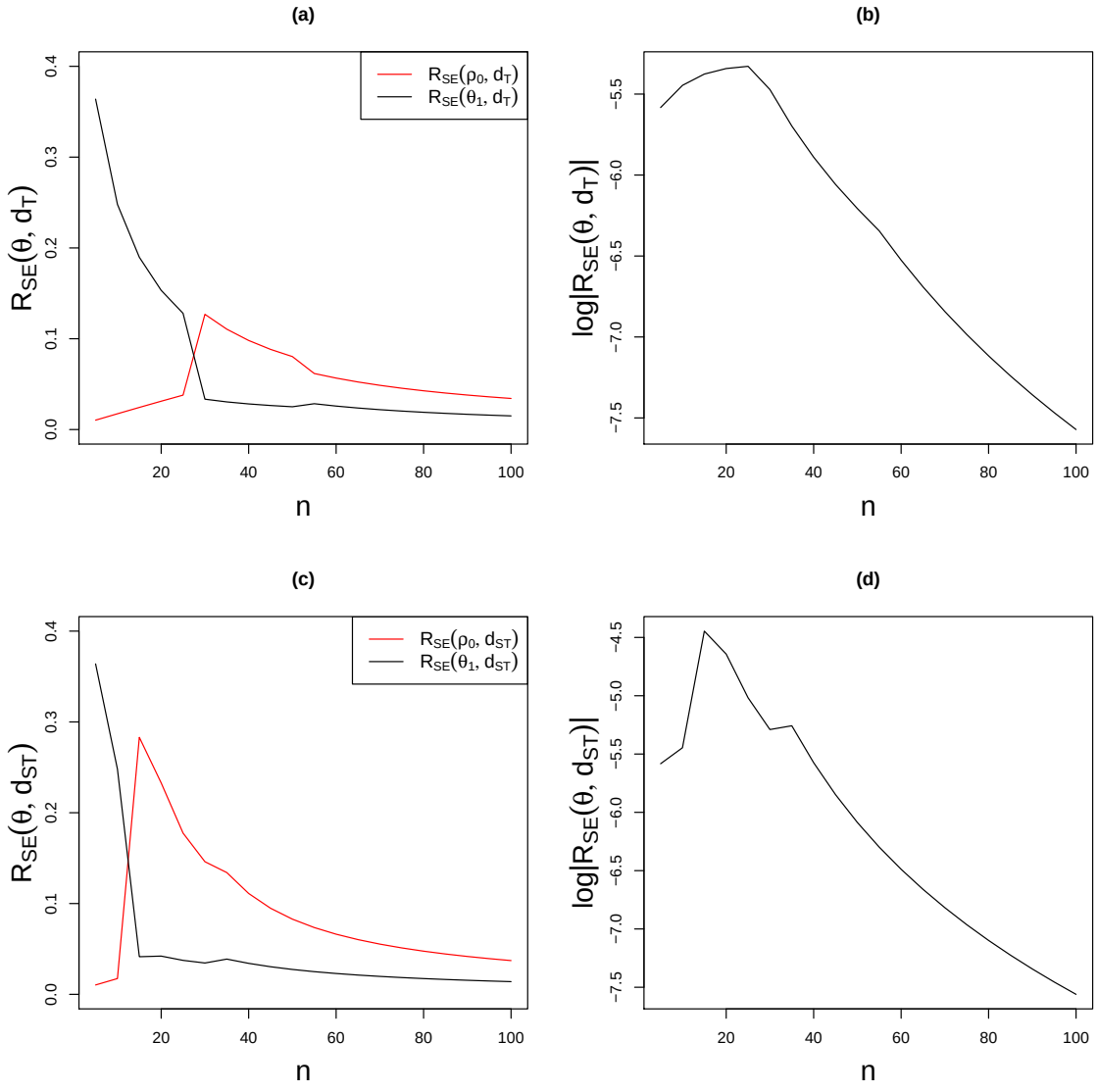

Figure 6.7: Plots of the MSE for $R_{SE}(\rho_0, d)$ and $R_{SE}(\theta_1, d)$, evaluated at d_T **(a)** and d_{ST} **(c)** found under the log determinant of MSE, and the log determinant of $R_{SE}(\boldsymbol{\theta}, d)$ evaluated at d_T **(b)** and d_{ST} **(d)**, plotted against different sizes of n.

6.3.1.2 Pseudo-Bayesian decision-theoretic D-optimal designs

Exact and approximate pseudo-Bayesian decision-theoretic D-optimal designs are found by minimising the approximate expectation of the log determinant of the exact MSE, $\widetilde{E}_{\boldsymbol{\theta}}\{\log|R_{SE}(\boldsymbol{\theta}, d)|\}$, and an approximation to the MSE, $\widetilde{E}_{\boldsymbol{\theta}}\left\{\log|\widetilde{R}_{SE}(\boldsymbol{\theta}, d)|\right\}$, respectively. In Figure 6.8, we evaluate $\widetilde{E}_{\boldsymbol{\theta}}\{\log|R_{SE}(\boldsymbol{\theta}, d)|\}$ at the exact and approximated designs for different sizes of n and two different priors, $\pi_1(\boldsymbol{\theta})$ (left) and $\pi_2(\boldsymbol{\theta})$ (right). It is found that there are some inaccuracies for a small n ($n = 45$), see $\pi_1(\boldsymbol{\theta})$, and a moderate-sized n ($n = 100$), see $\pi_2(\boldsymbol{\theta})$, where the approximate expectation of the exact MSE increases under the ST (black square) and QIMPS

(purple diamond) designs.

However, the approximate expectation of the QMC approximation to MSE appears very accurate compared with others and its design is near True design. Also, as n gets larger all approximated designs get near the exact optimal design since $\widetilde{E}_{\boldsymbol{\theta}}\{\log|R_{SE}(\boldsymbol{\theta}, d)|\}$ evaluated at each approximated design converges to $\widetilde{E}_{\boldsymbol{\theta}}\{\log|R_{SE}(\boldsymbol{\theta}, d_T)|\}$.

Note that some approaches improve when we consider pseudo-Bayesian optimality either under $\pi_1(\theta)$ or $\pi_2(\theta)$. This suggests that the chosen prior distribution of the parameters plays a crucial role in finding pseudo-Bayesian decision-theoretic optimal designs. Under $\pi_2(\boldsymbol{\theta})$, the quantity $\widetilde{E}_{\boldsymbol{\theta}}\{\log|R_{SE}(\boldsymbol{\theta}, d)|\}$ has not been evaluated at the Quad and QMC designs for small n, particularly $5 \leq n \leq 15$. This is because at some initial values that come from the abscissas of the quadrature rule, we always observe responses $\mathbf{Y} = (0, n)$ or $\mathbf{Y} = (n, 0)$, resulting in a singular MSE in which the intercept ρ_0 is always 0. For $n \geq 20$, the probability of observing responses where $\mathbf{Y} \neq (0, n)$ or $\mathbf{Y} \neq (n, 0)$ increases under the Quad and QMC approximations, where at least we obtained one observed response $\mathbf{Y} = (1, n)$ or $\mathbf{Y} = (0, n-1)$.

In the Figure 6.9, we investigate the performance of optimal designs under $\pi_2(\theta)$ if $\pi_1(\theta)$ is correct (a), and the performance of optimal designs under $\pi_1(\theta)$ if $\pi_2(\theta)$ is correct (b). It is noted that the optimal designs are sensitive to the chosen prior, as their performance are different than what we observe in Figure 6.8, especially True and QMC designs. The exact risk evaluated under these designs increases, and this increase happens due to the sensitivity of these designs to the chosen prior distribution.

Figure 6.10 and Figure 6.11 show five contour plots of the approximate expectations of the exact and approximations to the MSE, plotted for $n = 45$ (under $\pi_1(\boldsymbol{\theta})$) and $n = 100$ (under $\pi_2(\boldsymbol{\theta})$), respectively. It appears from the first row (left) of these plots that the ST and QIMPS designs move far from the other optimal designs and become close to the pseudo-Bayesian D-optimal design. This is similar to what happened for locally decision-theoretic D-optimal designs, but here, the QIMPS moves far from other optimal designs instead of Quad design. Thus, we need to check the validity of these approximations.

The validity of the approximate expectations of approximations to the MSE is investigated in Figure 6.12 (again, large values of n are considered), where under $\pi_1(\boldsymbol{\theta})$ is shown on the first two rows and under $\pi_2(\boldsymbol{\theta})$ is shown on the last two rows.

For both priors, it is noted that the approximate expectations of the QMC and Quad approximations to the MSE appear very accurate, unlike the approximate expectations of the ST and QIMPS approximations to the MSE. However, we can see from the ordering of the designs found under these approximations, in terms of the approximate expectation of the MSE, that it is the same as that for the approximate expectation of the exact MSE for each plot. This means that the approximate expectations of the ST and QIMPS approximations to the MSE are minimised for designs close to the design that minimises the approximate expectation of the exact MSE.

Here, the RT is calculated, i.e. the CPU time for the approximate expectation of each approximation / the CPU time for the approximate expectation of the exact MSE. It is found that the RT of the approximate expectation of the QMC approximation to the MSE is 89.40. This means that this approximation is approximately 89 times more expensive than the time required to compute the approximate expectation of the exact MSE. The approximate expectation of the QIMPS approximation is the second most expensive technique with RT=84.45, followed by Quad (RT=0.84), ST (RT=0.25) and pseudo-Bayesian D-optimality (RT=0.02).

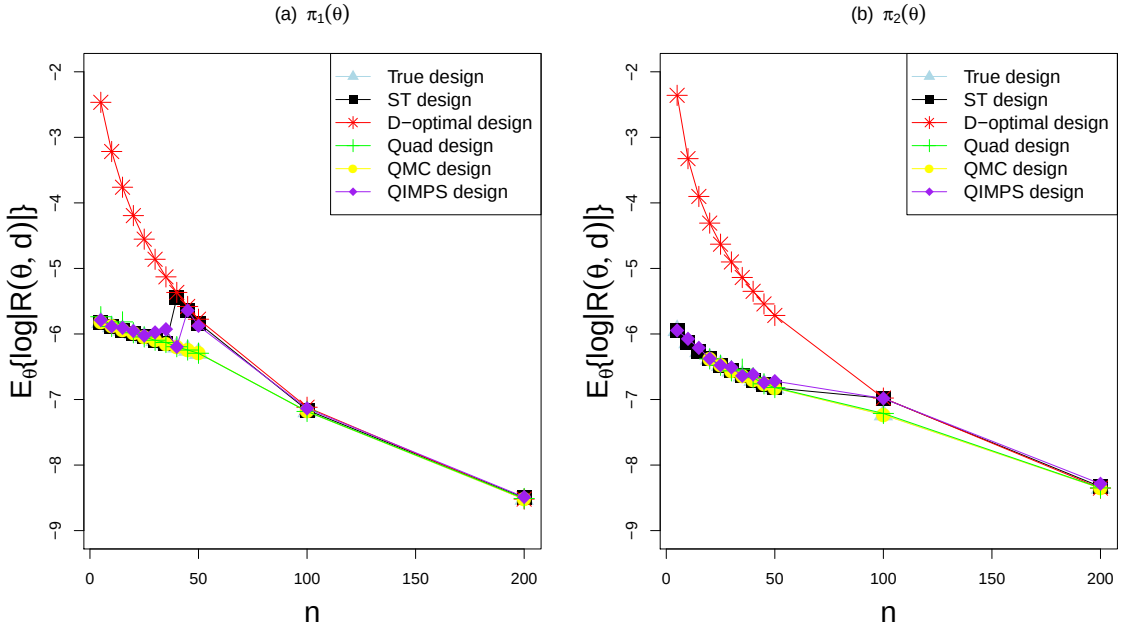

Figure 6.8: Plots of the two approximate prior expectations of the exact $\log|R_{SE}(\boldsymbol{\theta}, d)|$ evaluated at the True and approximated designs, and plotted against different sizes of binomial trials n for simple logistic regression model.

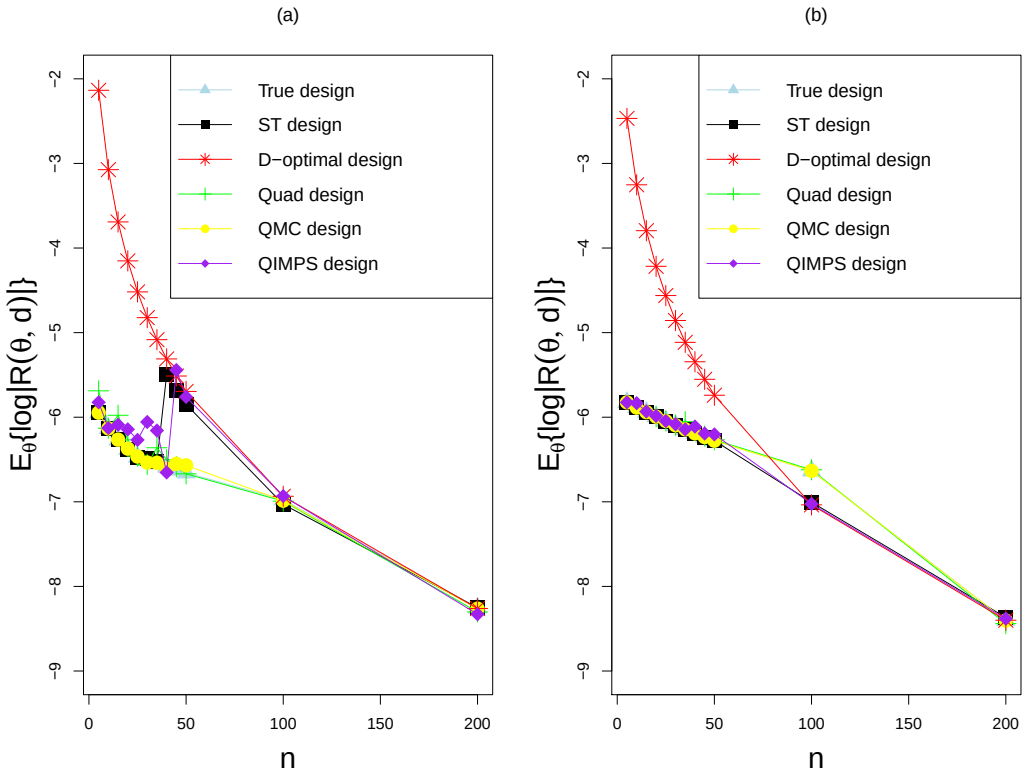

Figure 6.9: Plots of the exact $\log|R_{SE}(\boldsymbol{\theta}, d)|$ evaluated at optimal designs, and plotted against different sizes of binomial trials n. These plots illustrates the misspecification of the prior distributions, where the performance of the optimal designs found under prior 1 are investigated under prior 2 (a) and vice versa (b).

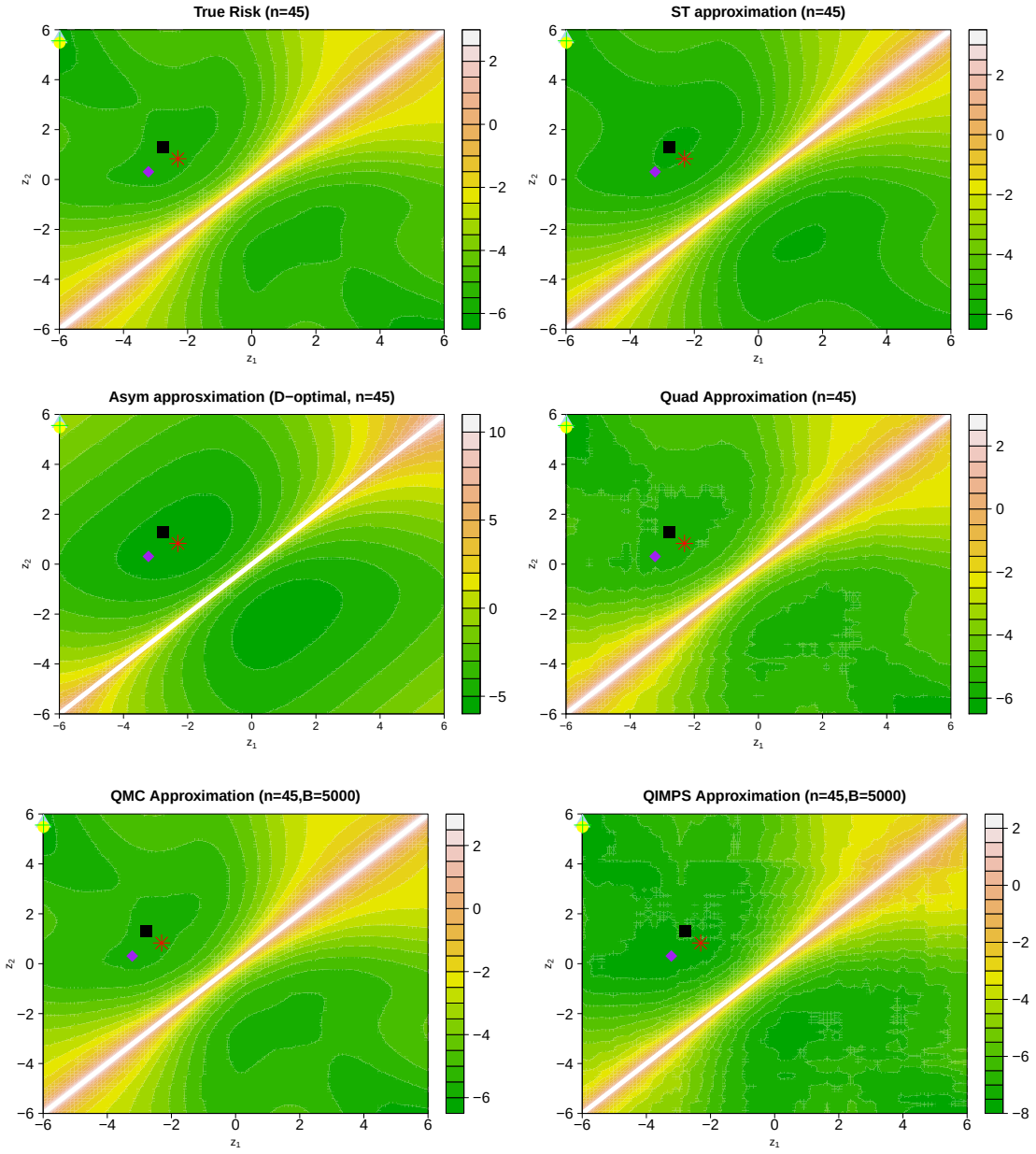

Figure 6.10: Contour plots of the approximate prior expectation, $\pi_1(\boldsymbol{\theta})$, of the exact $\log|R_{SE}(\boldsymbol{\theta}, d)|$ (the first row on left), and of the approximated $\log|\widetilde{R}_{SE}(\boldsymbol{\theta}, d)|$ via second-order Taylor series (ST) (the first row on right), Asymptotic (Asym) (the second row on left), Quadrature (Quad) (the second row on right), quasi-Monte Carlo (QMC) with size $B = 5,000$ (the third row on left) and quasi-importance sampling (QIMPS) with size $B = 5,000$ (the third row on right), under simple logistic regression model of binomial trials of size $n = 45$. A plot key is plotted for each contour which explains the contour plot and refers to the values of $E_{\boldsymbol{\theta}}\{\log|R_{SE}(\boldsymbol{\theta}, d)|\}$ (the first row on left) and $E_{\boldsymbol{\theta}}\left\{\log|\widetilde{R}_{SE}(\boldsymbol{\theta}, d)|\right\}$ for the remaining contour plots. In Each plot, there are six coloured-shapes denote the pseudo-Bayesian decision-theoretic D-optimal design found via the truth risk (light blue triangle), ST (black square), Asym (red star), Quad (green plus), QMC (yellow filled circle) and QIMPS approximation (purple diamond).

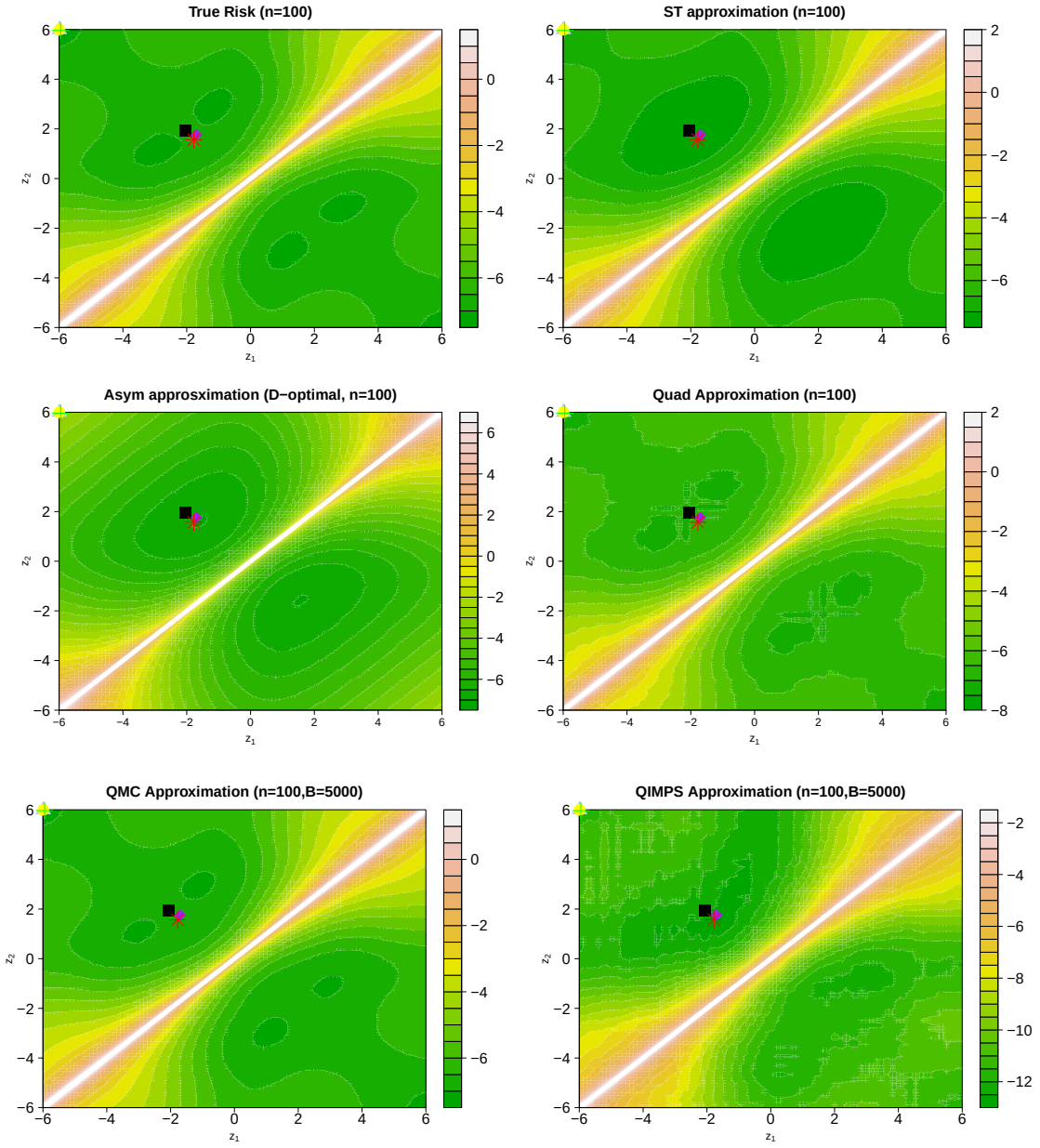

Figure 6.11: Contour plots of the approximate prior expectation, $\pi_2(\boldsymbol{\theta})$, of the exact $\log|R_{SE}(\boldsymbol{\theta}, d)|$ (the first row on left), and of the approximated $\log|\widetilde{R}_{SE}(\boldsymbol{\theta}, d)|$ via second-order Taylor series (ST) (the first row on right), Asymptotic (Asym) (the second row on left), Quadrature (Quad) (the second row on right), quasi-Monte Carlo (QMC) with size $B = 5,000$ (the third row on left) and quasi-importance sampling (QIMPS) with size $B = 5,000$ (the third row on right), under simple logistic regression model of binomial trials of size $n = 100$. A plot key is plotted for each contour which explains the contour plot and refers to the values of $E_{\boldsymbol{\theta}}\{\log|R_{SE}(\boldsymbol{\theta}, d)|\}$ (the first row on left) and $E_{\boldsymbol{\theta}}\left\{\log|\widetilde{R}_{SE}(\boldsymbol{\theta}, d)|\right\}$ for the remaining contour plots. In Each plot, there are six coloured-shapes denote the pseudo-Bayesian decision-theoretic D-optimal design found via the truth risk (light blue triangle), ST (black square), Asym (red star), Quad (green plus), QMC (yellow filled circle) and QIMPS approximation (purple diamond).

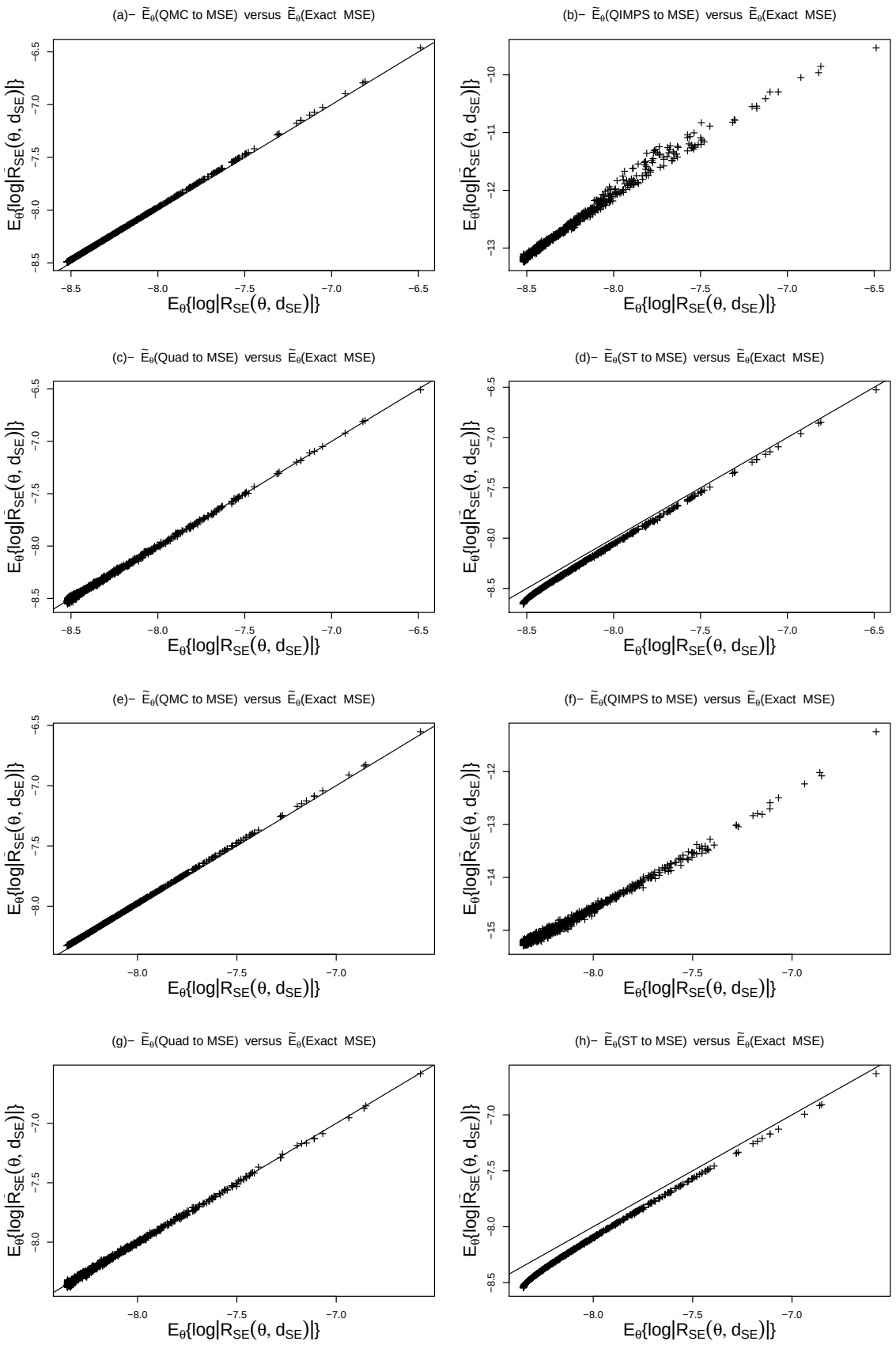

Figure 6.12: Plots of the approximate expectations of approximations to MSE, plotted against the approximate expectations of the exact MSE for $n = 100$ under $\pi_1(\boldsymbol{\theta})$ (the first two rows) and $\pi_2(\boldsymbol{\theta})$ (the last two rows).

6.3.2 Discussion

We found locally and pseudo-Bayesian decision-theoretic optimal designs under SE loss and three different scalar functions of the MSE. The designs found under the scaler function of the trace and the maximum eigenvalue of the MSE are discussed in "Appendix C". Different approximations to the MSE were applied for finding locally approximate designs, and these were Asym, ST, Quad, and QMC with size B=5,000. We used the QIMPS approximation with size $B = 5,000$ in addition to these approximations for finding approximate pseudo-Bayesian designs where the prior expectation was approximated via Quad approximation. It was found that the QMC is the most accurate approximation to the MSE for various values of n, but it requires much more computing time than the other approximations. The same comment applies to the approximate expectation of the QMC approximation to the MSE for finding pseudo-Bayesian designs.

For local optimality, there were inaccuracies that occurred for small n under some approaches; these inaccuracies appeared when the responses $\mathbf{Y} = (0, 0)$ or $\mathbf{Y} = (n, n)$. Intuitively, when the same responses are applied for z_1 and z_2, zero value results for the slope parameter θ_1. Hence, the value of the intercept, ρ_0, dominates the MSE; this was confirmed by plotting the MSE for both ρ_0 and θ_1 separately. This result was generalised to the pseudo-Bayesian approach to justify the inaccuracies that occurred there.

6.4 Four factors logistic regression model

Four factors logistic regression model is an extension to simple logistic regression model, and it has an application in a crystallography experiment (Woods et al., 2006). It is used to model how process variables affect the probability of the formulation of a new product, where there are four variables that affect the formulation of the new product. These are the evaporation rate, temperature, rate of agitation during mixing and volume of composition. Under $k = 4$ factor logistic regression model, there are $p = 5$ unknown parameters. In order to find pseudo-Bayesian decision-theoretic optimal designs, following Woods et al. (2006), we assume for each element of unknown parameters an independent uniform prior distribution, i.e

$$\begin{aligned} &\rho_0 \sim U[-3, 3] \qquad \theta_1 \sim U[4, 10] \qquad \theta_2 \sim U[5, 11] \\ &\theta_3 \sim U[-6, 0] \quad \theta_4 \sim U[-2.5, 3.5]. \end{aligned} \tag{6.2}$$

Each element of the design d is assumed to be generated from $U[-1, 1]$. In this section, under SE loss, we find pseudo-Bayesian decision-theoretic D-, A-, and E-optimal designs by minimising the prior expectation of the log determinant, trace

and maximum eigenvalue of the MSE, respectively. Furthermore, we find pseudo-Bayesian decision-theoretic optimal designs under the SL by minimising the prior expectation of the expected SL.

Finding such designs under this example is complicated, as the prior expectation of the risk function is analytically intractable and results from a $p = 5$-dimensional integral. Therefore, the prior expectation is approximated using the Quad approximation.

The risk functions (the MSE or the expected SL) can only be analytically calculated for a small experiment (a small number of runs); however, a number of approximations to these functions are considered, namely ST, Asym, QMC with $B = 1,000$ and $B = 500$, QIMPS with $B = 1,000$, and Quad. We refer to the design found under the QMC approximation with $B = 1,000$ and $B = 500$ as $d_{QMC_{L1}}$ and $d_{QMC_{L2}}$, respectively. Note that QMC 2 is only applied under SE loss to approximate the trace of the MSE, and the objective function under this approximation is

$$\widetilde{E}_{\boldsymbol{\theta}}\{\widetilde{R}_{QMC}(d;\boldsymbol{\theta})\} = \frac{1}{B}\sum_{h=1}^{H} w_h \sum_{b=1}^{B}\sum_{i=1}^{p}\left(\hat{\theta}_{ii}(\mathbf{Y}(\mathbf{u}_b)) - \theta_{h_{ii}}\right)^2.$$

The optimal designs are found by applying the methodology of the approximate coordinate exchange (ACE) algorithm (see "Appendix A" for details). Finding the optimal designs via the ACE algorithm is an optimisation problem that is common with other searching algorithms for optimal designs in that they converge to the local optima (Overstall et al., 2016). To avoid this convergence, the algorithm is restarted from T different starting designs that are randomly chosen from the design space χ, where these T repetitions of the ACE algorithm are run in parallel. Here, we set $T = 20$ as a reasonable number used to find an optimal design, as the algorithm is computationally expensive. We consider phase I of the ACE algorithm, where we set $N1 = 20$ and $N2 = 0$ (see Section A.4 in "Appendix A"). Hence, the design that has the lowest expected risk is chosen to be the optimal one.

A warm-starting optimisation is a technique which consists of two approaches: asymptotic and QMC approximations (henceforth referred to as WSMC). It can be used to reduce the CPU time required to solve an optimisation problem in phase I of the ACE algorithm. This technique uses $N1 = 15$ out of 20 repetitions for evaluating $\widetilde{E}_{\boldsymbol{\theta}}\left\{\log|\widetilde{R}(\boldsymbol{\theta},d)|\right\}$, as the $R(\boldsymbol{\theta},d)$ is approximated via a cheaper approximation given by the asymptotic approximation. Then, the resulting design is

used as the start design for the remaining 5 repetitions that are used for evaluating $\widetilde{E}_{\boldsymbol{\theta}}\left\{\log|\widetilde{R}(\boldsymbol{\theta},d)|\right\}$ as $R(\boldsymbol{\theta},d)$ is approximated via QMC (the computationally expensive technique). Here, QMC with $B=1,000$ and $B=500$ are considered where the resulting optimal design under the former is referred to as $d_{WSMC_{L1}}$, while under the latter is referred to as $d_{WSMC_{L2}}$.

We compare approximated designs found under the Quad approximation to the prior expectation of the approximated risk function against True designs found under the Quad approximation to the prior expectation of the exact risk (for small N). In order to make the comparison valid, the approximated designs are found under exactly the same implementation of ACE as those used to find the True designs.

We calculate the RT for all resulting designs; note that the objective functions that we aim to minimise to find pseudo-Bayesian decision-theoretic optimal designs are coded using a low-level programming language (C++) and make use of packages **Rcpp** (Eddelbuettel and François, 2011) and **RcppArmadillo** (Eddelbuettel and Sanderson, 2014). The algorithm is run on IRIDIS 4, the super computer facility at the University of Southampton, which has 2.6 GHz processors with 4GB memory.

6.4.1 Pseudo-Bayesian decision-theoretic D-optimal designs

We find pseudo-Bayesian decision-theoretic D-optimal designs, and then evaluate $\widetilde{E}_{\boldsymbol{\theta}}\{\log|R_{SE}(\boldsymbol{\theta},d)|\}$ at all resulting designs to compare their accuracies. Table 6.2 shows the performance of each design in terms of the $\widetilde{E}_{\boldsymbol{\theta}}\{\log|R_{SE}(\boldsymbol{\theta},d)|\}$ that was evaluated at all designs found under different approaches. It also shows between brackets "(.)" that the CPU time required to find each design relative to the CPU time used for $d_{QMC_{L2}}$ (i.e. the CPU time for each design / the CPU time for $d_{QMC_{L2}}$). It can be seen that all designs perform very well under SE loss for $N=6$, except pseudo-Bayesian D-optimal designs that are very far from other designs, indicating poorly produced designs.

When the size of N becomes larger, the approximate expectation of the exact MSE slightly increases or fluctuates at all resulting designs. This unusual scenario needs to be investigated. In order to obtain a better understanding of the problem, we evaluate the log determinant of the exact MSE at each resulting decision-theoretic optimal designs (i.e. the smallest number of runs ($N=6$) considered). For example, True, QMC, Quad and ST designs with $N=6$ runs are given in Table 6.3, 6.4, 6.5 and 6.6, respectively. Interestingly, it is found that under these designs, we get the same response for every single experiment. Hence, we get the same estimates that result in a nearly singular MSE (i.e. it is very close to having

the determinant of the MSE is zero), where the binomial probabilities of observing the response are always such that

$$\begin{pmatrix}
2.17e-09 & 3.30e-06 & 9.12e-14 & 2.22e-10 & 7.91e-15 & 1.85e-11 & 6.87e-18 & 5.96e-14 & 3.30e-06 & 9.999e-01 \\
2.22e-10 & 8.03e-06 & 1.85e-11 & 2.14e-06 & 5.96e-14 & 1.20e-09 & 6.60e-14 & 1.61e-10 & 8.65e-17 & 7.57e-13 \\
4.93e-18 & 4.31e-14 & 5.23e-20 & 5.01e-16 & 1.61e-10 & 6.14e-06 & 7.57e-13 & 1.13e-08 & 4.31e-14 & 8.43e-10 \\
5.01e-16 & 5.69e-12 & 9.72e-14 & 2.44e-10 & 1.38e-16 & 1.20e-12 & 7.84e-18 & 6.81e-14 & 8.34e-20 & 7.97e-16 \\
2.44e-10 & 8.27e-06 & 1.20e-12 & 1.71e-08 & 6.81e-14 & 1.30e-09 & 7.97e-16 & 8.91e-12 & 9.91e-17 & 8.70e-13 \\
1.11e-18 & 1.06e-14 & 6.04e-20 & 5.78e-16 & 7.97e-22 & 7.57e-18 & 8.70e-13 & 1.27e-08 & 1.06e-14 & 1.13e-10 \\
5.78e-16 & 6.51e-12 & 7.57e-18 & 7.54e-14 & & & & & &
\end{pmatrix}$$

As a result, the log determinant is problematic when it is used as a scalar function of the MSE under this particular example.

From Table 6.2, we can also see, under each design that there are stars ($*$); these indicate that the corresponding design cannot be found, or it requires highly expensive time to be found. For example, the design with 8 runs found under QMC method with $B = 1000$ required about 54 hours to be found using the supercomputer, where the maximum limit for running a job using the supercomputer is 60 hours. As $d_{QMC_{L2}}$, found under QMC with $B = 500$, performs well and is found to be computed for a range of N, it is considered as the standard to be compared with against other approximated designs.

We can see that $d_{QMC_{L1}}$ requires approximately twice the CPU time used for $d_{QMC_{L2}}$. The RT of $d_{QMC_{L1}}$, d_Q and d_T increases as N becomes larger. This means that these designs are computationally expensive to be found for large N, especially d_T and d_{Quad}, even in the use of high-performance computing. This is because the number of responses generated under the corresponding approaches for finding these designs increases exponentially with the number of runs. We did not include the performance of the ST design, found under the approximate expectation of the ST approximation to the MSE, in this table, where the RT for finding this design is 11.01 for $N = 6$. This means that the CPU time used for finding this design is approximately 11 times more than the CPU time required for finding $d_{QMC_{L2}}$. However, the pseudo-Bayesian D-optimal design requires the smallest amount of CPU time.

Similar to Section 6.3, we check the validity of the approximate expectation of the used approximations to the MSE, but just for those used to find designs with a moderate-sized N. In the first row of Figure 6.13, we plot both the approximate expectations of QMC with $B = 1,000$ and $B = 500$ (left) and QIMPS with $B = 1,000$ (right) approximation to the MSE against the approximate expectation of the exact MSE for $N = 6$ (i.e. the smallest number of runs considered). We can see that these approximations show inaccuracies, and this is due to the problem

of using the log determinant to the MSE that has been discussed above for this particular example.

Table 6.2: The Quad approximation to the prior expectation of the exact MSE $\widetilde{E}_{\boldsymbol{\theta}}\{\log|R_{SE}(\boldsymbol{\theta},d)|\}$, evaluated at different designs found under different approaches for $k = 4$ factors logistic regression and different numbers of runs, $6 \leq N \leq 16$. The relative computing time (RT) used for finding the optimal design is given between (.) and the exact computing time of seconds is given for QMC with B=500 that is considered as the reference method. The stars $(*)$ under each design mean that the corresponding design cannot be computed or requires highly expensive time to be found.

N	d_T	d_D	d_{Quad}	$d_{QMC_{L1}}$	$d_{QMC_{L2}}$	d_{QIMPS}	$d_{WSMC_{L1}}$	$d_{WSMC_{L2}}$
6	-54.04 (0.15)	-3.17 (0.0002)	-45.61 (0.74)	-38.80 (1.97)	-46.71 (49115)	-38.64 (0.19)	-47.61 (0.52)	-48.20 (0.25)
7	-54.00 (0.18)	-2.74 (0.0002)	-49.25 (0.82)	-42.75 (2.13)	-42.03 (74459)	-38.78 (0.15)	-47.12 (0.43)	-43.59 (0.23)
8	-53.94 (0.39)	-2.74 (0.0002)	-49.23 (1.08)	-37.66 (2.13)	-44.77 (91479)	-38.58 (0.14)	-47.58 (0.55)	-41.80 (0.26)
9	-53.83 (0.74)	-2.46 (0.0002)	-48.43(1.45)	*	-42.53 (117095)	-37.70 (0.13)	-43.29 (0.54)	-45.24 (0.25)
10	*	-1.91 (0.0002)	-48.45 (1.69)	*	-45.07 (148168)	-39.18 (0.13)	-42.38 (0.55)	-43.58 (0.26)
11	*	-1.70 (0.0001)	*	*	-44.24 (204347)	-36.65 (0.10)	-48.86 (0.50)	-44.93 (0.24)
12	*	-1.71 (0.0001)	*	*	-42.52 (242654)	-35.34 (0.10)	-43.35 (0.50)	-42.53 (0.25)
13	*	-1.59 (0.0001)	*	*	-43.20 (277487)	-35.69 (0.10)	-42.85 (0.53)	-43.35 (0.26)
14	*	-1.61 (0.0001)	*	*	-35.02 (362961)	-36.29 (0.09)	*	-41.74 (0.24)
15	*	-1.61 (0.0001)	*	*	-40.58 (397719)	-33.85 (0.08)	*	-41.06 (0.25)
16	*	-1.56 (0.0001)	*	*	-39.31 (473776)	-34.01 (0.08)	*	-41.39 (0.24)

Table 6.3: True Design with $N = 6$ runs found under the exact risk for four factors logistic regression

	k1	k2	k3	k4
1	-1.00	-0.87	1.00	-1.00
2	-0.79	-1.00	1.00	-1.00
3	-1.00	-1.00	1.00	1.00
4	-1.00	-1.00	0.38	-1.00
5	1.00	1.00	-1.00	1.00
6	1.00	1.00	-1.00	1.00

Table 6.4: QMC Design with $N = 6$ runs found under the QMC approximation to the risk for four factors logistic regression

	k1	k2	k3	k4
1	-0.93	-0.84	0.21	0.26
2	0.91	1.00	-1.00	0.77
3	-1.00	-0.95	0.49	-0.39
4	-1.00	-0.91	-0.12	0.58
5	-1.00	-0.96	0.34	0.19
6	0.93	0.74	-0.18	0.42

Table 6.5: Quad Design with $N = 6$ runs found under the Quad approximation to the risk for four factors logistic regression

	k1	k2	k3	k4
1	-1.00	-1.00	1.00	-0.15
2	1.00	1.00	-0.90	0.29
3	-1.00	-1.00	0.70	0.45
4	-1.00	-1.00	1.00	-0.01
5	-0.99	-1.00	1.00	-0.78
6	-1.00	-1.00	1.00	-0.70

Table 6.6: ST Design with $N = 6$ runs found under the ST approximation to the risk for four factors logistic regression

	k1	k2	k3	k4
1	-1.00	-0.85	1.00	-0.89
2	-1.00	-1.00	0.61	-1.00
3	1.00	1.00	-1.00	1.00
4	-1.00	-1.00	0.97	0.10
5	-0.82	-1.00	1.00	-0.83
6	1.00	1.00	-1.00	1.00

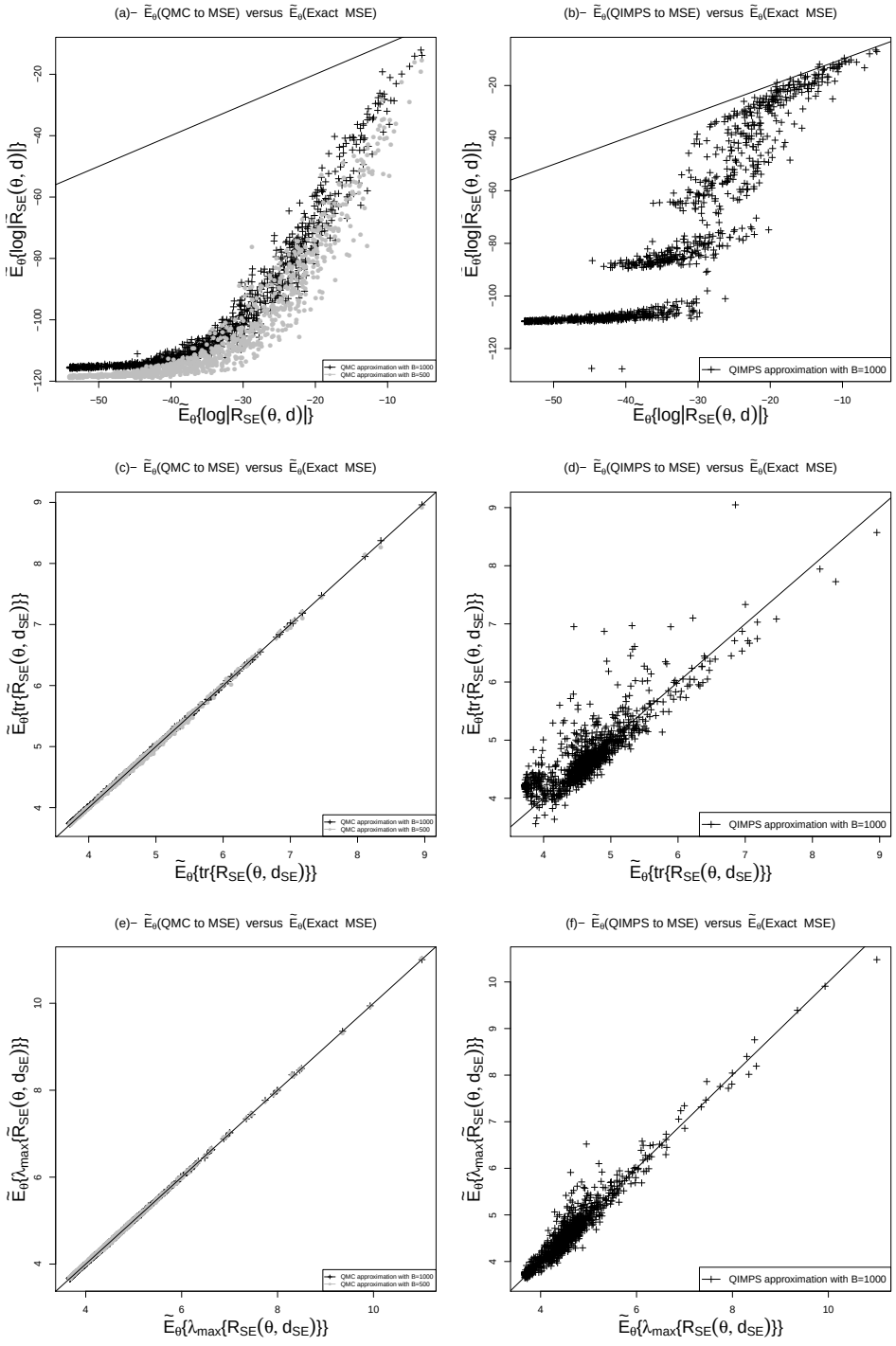

Figure 6.13: Plots of the Quad approximation to the prior expectation of QMC (with $B = 1000$ and $B = 500$) (left) and QIMPS approximations (right) to the MSE, $\widetilde{E}_{\boldsymbol{\theta}}\left\{S\left(\widetilde{R}_{SE}(\boldsymbol{\theta}, d)\right)\right\}$, plotted against the Quad approximation to the prior expectation of the exact MSE, $\widetilde{E}_{\boldsymbol{\theta}}\left\{S\left(R_{SE}(\boldsymbol{\theta}, d)\right)\right\}$, for four factor logistic regression ($N = 6$), where $S(.)$ refers to the scalar function of the log determinant (the first row), trace (the second row) and the maximum eigenvalues (the third row). In order to aid in the comparison, a line through the origin with slope one has been added.

6.4.2 Pseudo-Bayesian decision-theoretic A-optimal designs

In this section, we find pseudo-Bayesian decision-theoretic A-optimal designs. To compare the accuracies of these designs, we evaluate the $\widetilde{E}_{\boldsymbol{\theta}}\left\{\text{tr}\left\{R_{SE}(\boldsymbol{\theta}, d)\right\}\right\}$ at all designs given in Table 6.7. Apparently, $d_{QMC_{L1}}$, $d_{QMC_{L2}}$, d_Q and d_{QMC_2} are near d_T for small N compared to others.

Similar to Section 6.4.1, it is noted that $d_{QMC_{L1}}$ is the most expensive design to be found with RT=1.86 for $N = 6$, while d_T and d_Q are very computationally expensive to be found for large N, where their RTs are increasing as N becomes larger. The pseudo-Bayesian A-optimal design is the least computationally expensive where it can be found in seconds; nevertheless, this should be judged in parallel with the lack of efficiency these designs exhibit when compared with other decision-theoretic optimal designs.

For a moderate-sized N, we only find d_A, $d_{QMC_{L2}}$, d_{QIMPS} and $d_{WSMC_{L2}}$. Hence, we compare the performance of d_A, d_{QIMPS} and $d_{WSMC_{L2}}$ against $d_{QMC_{L2}}$ which is the most accurate design amongst these designs. In Figure 6.14 (left), the quantity $\widetilde{E}_{\boldsymbol{\theta}}\left\{\text{tr}\left\{R_{SE}(\boldsymbol{\theta}, d)\right\}\right\}$ is evaluated at $d_{QMC_{L2}}$ (yellow open circle), $d_{WSMC_{L2}}$ (orange star), d_{QIMPS} (purple diamond) and d_A (red star), and plotted against N. It can be seen that d_{QIMPS} found under the QIMPS approximation is the worst design and shows inaccuracies for a range of N compared with others. The $d_{WSMC_{L2}}$ performs better than d_A for small N; however, these designs gain improvement as N becomes larger where the $\widetilde{E}_{\boldsymbol{\theta}}\left\{\text{tr}\left\{R_{SE}(\boldsymbol{\theta}, d)\right\}\right\}$ under these designs converges to this quantity under $d_{QMC_{L2}}$.

Again similar to Section 6.4.1, we check the validity of the approximate expectation of the used approximations to the MSE. In the second row of Figure 6.13, we plot both the approximate expectations of the QMC with $B = 1,000$ and $B = 500$ (left) and the QIMPS with $B = 1,000$ (right) approximation against the approximate expectation of the exact MSE for $N = 6$. We can see that the QMC approximation (with $B = 500$ and $B = 1000$) to the MSE appear very accurate, whereas the QIMPS approximation can be inaccurate, especially for designs close to the maximum. However, we can see from the ordering of the designs found under the approximate expectation of the QIMPS approximation to the MSE is the same as that for the approximate expectation of the exact MSE, indicating that the approximate expectations of the QIMPS approximation to the MSE are minimised for designs close to the designs that minimise the approximate expectation of the exact MSE.

Table 6.7: The Quad approximation to the prior expectation of the exact MSE $\widetilde{E}_{\boldsymbol{\theta}}\{\mathrm{tr}\{R_{SE}(\boldsymbol{\theta}, d)\}$, evaluated at different designs found under different approaches for $k = 4$ factor logistic regression and different numbers of runs, $6 \leq N \leq 16$. The relative computing time (RT) used for finding the optimal design is given between (.) and the exact computing time of seconds is given for QMC with B=500 that is considered as the reference method. The stars $(*)$ under each design mean that the corresponding design cannot be computed or requires highly expensive time to be found.

N	d_T	d_A	d_{Quad}	$d_{QMC_{L1}}$	$d_{QMC_{L2}}$	d_{QMC_2}	d_{QIMPS}	$d_{WSMC_{L1}}$	$d_{WSMC_{L2}}$
6	41.15 (0.11)	62.20 (0.0004)	41.07 (0.59)	43.21 (1.86)	45.72 (53059)	42.20 (1.80)	65.54 (0.19)	46.43 (0.46)	47.25 (0.23)
7	35.31 (0.23)	53.61 (0.0003)	37.38 (0.97)	35.99 (2.03)	34.59 (71052)	36.76 (2.27)	48.72 (0.19)	43.35 (0.62)	40.76 (0.29)
8	31.85 (0.42)	46.92 (0.0002)	32.55 (1.14)	*	32.71 (111473)	*	55.59 (0.13)	40.21 (0.56)	39.72 (0.29)
9	29.15 (0.74)	41.41 (0.0002)	29.52 (1.29)	*	30.49 (162279)	*	54.08 (0.11)	34.91 (0.49)	39.35 (0.25)
10	*	37.06 (0.0001)	28.30 (1.69)	*	28.41 (206853)	*	45.88 (0.10)	34.78 (0.53)	35.31 (0.25)
11	*	33.38 (0.0001)	*	*	26.64 (257391)	*	45.54 (0.09)	31.66 (0.51)	31.48 (0.24)
12	*	30.00 (0.0001)	*	*	24.45 (323153)	*	38.53 (0.09)	28.68 (0.48)	28.83 (0.24)
13	*	27.47 (0.0001)	*	*	23.51 (372654)	*	39.44 (0.09)	26.48 (0.49)	26.09 (0.25)
14	*	24.74 (0.0001)	*	*	22.39 (443538)	*	29.92 (0.08)	*	23.60 (0.25)
15	*	22.55 (0.0001)	*	*	22.36 (499445)	*	29.11 (0.08)	*	22.43 (0.25)
16	*	20.87 (0.0001)	*	*	20.72 (561247)	*	28.30 (0.07)	*	20.45 (0.26)

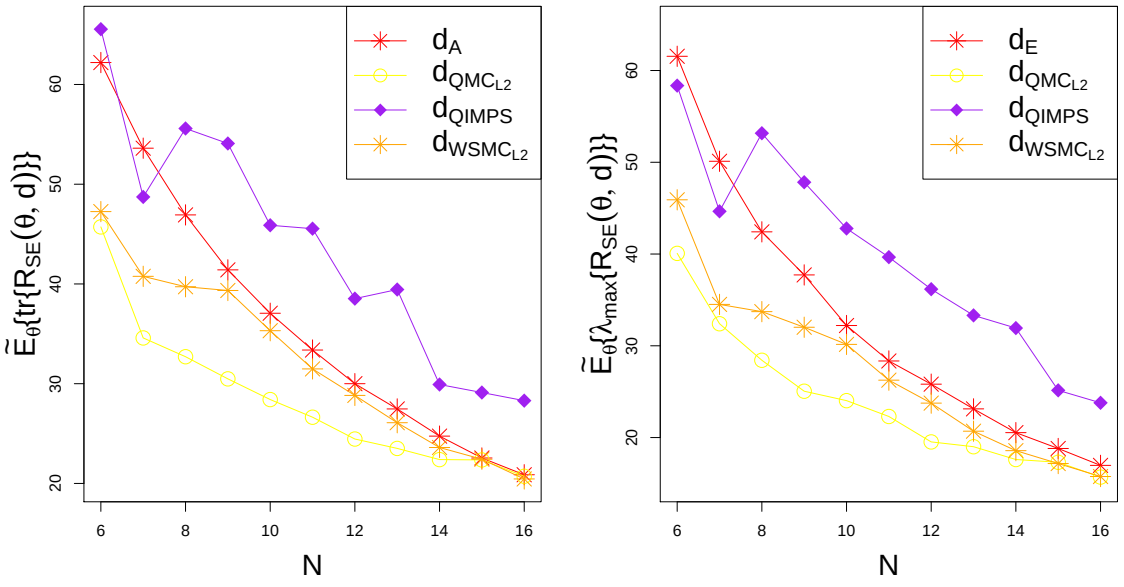

Figure 6.14: Plots of the Quad approximation to the prior expectation of the exact expected SE, where $\widetilde{E}_{\boldsymbol{\theta}}\{\mathrm{tr}\{R_{SE}(\boldsymbol{\theta}, d)\}$ (left) and $\widetilde{E}_{\boldsymbol{\theta}}\{\lambda_{max}\{R_{SE}(\boldsymbol{\theta}, d)\}\}$ (right). There are four different coloured shapes refer to the approximate decision-theoretic optimal designs, and these are d_A or d_E (red star), $d_{QMC_{L2}}$ (yellow open circle), d_{QIMPS} (purple diamond) and $d_{WSMC_{L2}}$ (orange star).

6.4.3 Pseudo-Bayesian decision-theoretic E-optimal designs

We find pseudo-Bayesian decision-theoretic E-optimal designs, and then evaluate the $\widetilde{E}_{\boldsymbol{\theta}}\{\lambda_{max}\{R_{SE}(\boldsymbol{\theta}, d)\}\}$ at all designs given in Table 6.8 to compare the accuracies of the resulting designs. Similar comments apply here as those for the pseudo-Bayesian decision-theoretic A-optimal designs, in terms of the performance and RT that are calculated for the resulting designs.

In Figure 6.14 (right), we compare the designs found for moderate-sized runs against $d_{QMC_{L2}}$, and these are d_E, d_{QIMPS} and $d_{WSMC_{L2}}$. It can be seen that $d_{WSMC_{L2}}$ performs very well as the N becomes larger, in which it gets near $d_{QMC_{L2}}$, followed by d_E found under the approximate expectation of the asymptotic approximation to the MSE. Furthermore, d_{QIMPS} is the worst design for all sizes of runs. These results are similar to what was found under pseudo-Bayesian decision-theoretic A-optimal designs.

We check the validity of the approximations, similar to Section 6.4.2, by plotting the Quad approximation to the prior expectation of QMC (with $B = 500$ and $B = 1,000$)(left) and QIMPS with $B = 1,000$ (right) approximations to the MSE against the Quad approximation to the prior expectation of the exact MSE, see the third row of Figure 6.13. Again, similar comments apply here.

Table 6.8: The Quad approximation to the prior expectation of the exact MSE $\widetilde{E}_{\boldsymbol{\theta}}\{\lambda_{\max}\{R_{SE}(\boldsymbol{\theta}, d)\}\}$, evaluated at different designs found under different approaches for $k = 4$ factor logistic regression and different numbers of runs, $6 \leq N \leq 16$. The relative computing time (RT) used for finding the optimal design is given between (.) and the exact computing time of seconds is given for QMC with B=500 that is considered as the reference method. The stars $(*)$ under each design mean that the corresponding design cannot be computed or requires highly expensive time to be found.

N	d_T	d_E	d_{Quad}	$d_{QMC_{L1}}$	$d_{QMC_{L2}}$	d_{QIMPS}	$d_{WSMC_{L1}}$	$d_{WSMC_{L2}}$
6	39.15 (0.12)	61.55 (0.0003)	40.35 (0.67)	40.55 (1.99)	40.07 (48767)	58.35 (0.20)	46.04 (0.59)	45.91 (0.29)
7	32.45 (0.24)	50.11 (0.0002)	32.90 (0.91)	31.98 (2.08)	32.41 (83634)	44.66 (0.15)	34.55 (0.54)	34.51 (0.27)
8	28.44 (0.41)	42.42 (0.0002)	28.63 (1.08)	*	28.42 (115252)	53.17 (0.12)	33.89 (0.54)	33.72 (0.27)
9	26.21 (0.78)	37.72 (0.0002)	26.24 (1.30)	*	25.05 (165186)	47.82 (0.11)	33.70 (0.50)	32.02 (0.24)
10	*	32.21 (0.0001)	24.43 (1.60)	*	24.02 (216743)	42.78 (0.09)	30.47 (0.48)	30.15 (0.24)
11	*	28.35 (0.0001)	*	*	22.30 (260461)	39.66 (0.09)	26.33 (0.50)	26.25 (0.25)
12	*	25.82 (0.0001)	*	*	19.53 (323199)	36.17 (0.08)	23.02 (0.48)	23.75 (0.24)
13	*	23.13 (0.0001)	*	*	19.00 (384034)	33.31 (0.08)	21.19 (0.48)	20.69 (0.24)
14	*	20.54 (0.0001)	*	*	17.61 (454590)	31.93 (0.07)	*	18.57 (0.24)
15	*	18.81 (0.0001)	*	*	17.31 (514623)	25.15 (0.07)	*	17.17 (0.24)
16	*	16.97 (0.0001)	*	*	15.69 (579429)	23.79 (0.07)	*	15.75 (0.25)

6.4.4 Pseudo-Bayesian decision-theoretic optimal designs under the SL

Pseudo-Bayesian decision-theoretic optimal designs under the SL are found, for moderate-sized runs, by minimising the Quad approximation to the prior expectation of the expected SL, $\widetilde{E}_{\boldsymbol{\theta}}\{R_{SL}(\boldsymbol{\theta}, d)\}$. The exact expected SL can only be found for small runs; hence, a number of approximations to the expected SL are considered, and these are Asym, Quad and QMC with $B = 1,000$ and $B = 500$. Here, the QIMPS approximation to the expected SL is excluded because of the inaccuracies that exhibits this approximation (i.e. noisy data), as shown previously.

However, we compare the designs found under the approximate expectations of the approximations to the expected SL in Table 6.9. Note that we evaluate the approximate expectation of the exact expected SL at all resulting designs to make the comparison valid. It is noted that the $d_{QMC_{L1}}$ and $d_{QMC_{L2}}$, as well as the d_Q design are near d_T for small N, but finding True designs are infeasible for large N under $k = 4$ factor logistic regression model, as stated earlier. The RT is also calculated in this table, and is found to be similar to what was concluded under SE loss.

For medium-sized runs, we are able to find $d_{QMC_{L2}}$ and d_{Asym} (design found under the approximate expectation of the asymptotic approximation to the expected SL) in addition to $d_{WSMC_{L2}}$. In Figure 6.15, we compare d_{Asym} and $d_{WSMC_{L2}}$ against $d_{QMC_{L2}}$. It is noted that $d_{WSMC_{L2}}$ appears accurate, in which it is very close to $d_{QMC_{L2}}$, whereas d_{Asym} can be inaccurate for small N. However, these designs improve as N increases.

In Figure 6.16, the validity of the Quad approximation to the prior expectations of the QMC approximations to the expected SL with ($B = 500$ and $B = 1,000$) is checked to ensure that it is asymptotically equivalent to the Quad approximation to the prior expectation of the exact expected SL. Apparently, the expectations of the QMC approximation with $B = 1,000$ and $B = 500$ appear very accurate.

Table 6.9: The Quad approximation to the prior expectation of the exact expected SL, $\widetilde{E}_{\boldsymbol{\theta}}\{R_{SL}(\boldsymbol{\theta}, d)\}$, evaluated at different designs found under different approaches for $k = 4$ factor logistic regression and different numbers of runs, $6 \leq N \leq 16$. The relative computing time (RT) used for finding the optimal design is given between (.), and the exact computing time of seconds is given for QMC with B=500 that is considered as the reference method. The stars $(*)$ under each design mean that the corresponding design cannot be computed or requires highly expensive time to be found.

N	d_T	d_{Asym}	d_{Quad}	$d_{QMC_{L1}}$	$d_{QMC_{L2}}$	$d_{WSMC_{L1}}$	$d_{WSMC_{L2}}$
6	0.48 (0.33)	0.82 (0.006)	0.49 (0.61)	0.47 (2.19)	0.49(51220)	0.49 (0.51)	0.49 (0.27)
7	0.44 (0.51)	0.65 (0.004)	0.46 (0.91)	0.47 (1.65)	0.46(82418)	0.47 (0.52)	0.46 (0.25)
8	0.39 (0.84)	0.58 (0.003)	0.40 (1.15)	0.41(1.92)	0.42 (114744)	0.46(0.52)	0.44 (0.24)
9	*	0.50 (0.003)	0.37 (1.49)	0.40 (2.04)	0.39(144316)	0.44 (0.56)	0.42 (0.28)
10	*	0.49 (0.002)	*	*	0.37 (203400)	0.40(0.48)	0.41 (0.25)
11	*	0.48 (0.002)	*	*	0.36(269601)	0.40 (0.45)	0.40 (0.23)
12	*	0.45 (0.002)	*	*	0.34(293317)	0.39 (0.50)	0.38 (0.23)
13	*	0.43 (0.002)	*	*	0.34 (341710)	0.37 (0.52)	0.37 (0.26)
14	*	0.40 (0.002)	*	*	0.33 (414343)	0.37 (0.52)	0.36 (0.27)
15	*	0.40 (0.001)	*	*	0.33 (488875)	*	0.36 (0.25)
16	*	0.38 (0.001)	*	*	0.32 (585626)	*	0.36 (0.24)

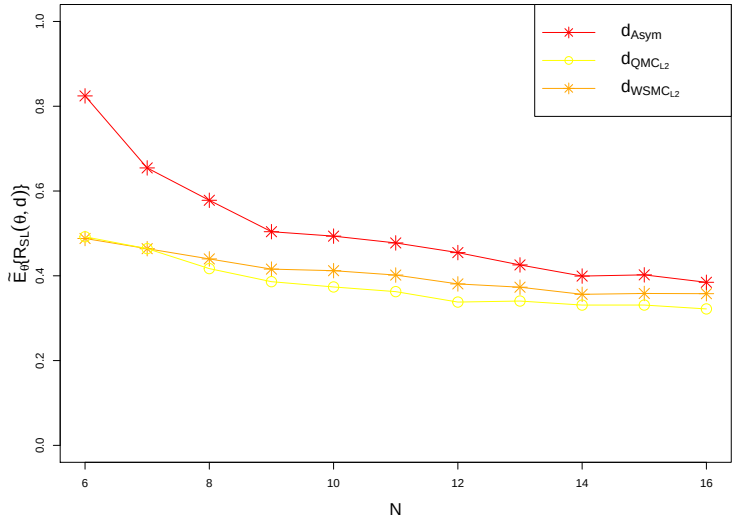

Figure 6.15: Plot of the Quad approximation to the prior expectation of the exact expected SL, where $\widetilde{E}_{\boldsymbol{\theta}}\{R_{SL}(\boldsymbol{\theta}, d)\}$ evaluated at different design found under different approaches. There are three different coloured shapes refer to the decision-theoretic optimal designs under the SL, and these are d_{Asym} (red star), $d_{QMC_{L2}}$ (yellow open circle) and $d_{WSMC_{L2}}$ (orange star).

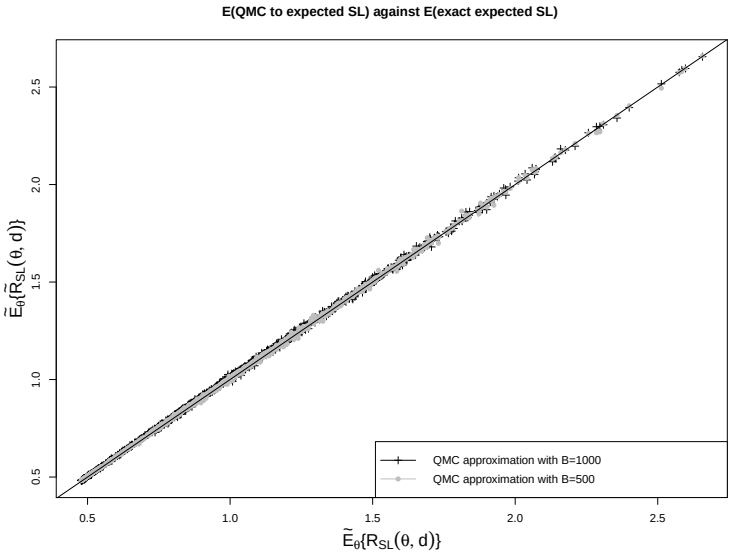

Figure 6.16: Plot of the Quad approximation to the prior expectation of QMC (with $B = 1,000$ and $B = 500$) approximation to the expected SL, $\widetilde{E}_{\boldsymbol{\theta}}\left\{\widetilde{R}_{SL}(\boldsymbol{\theta}, d)\right\}$, plotted against the Quad approximation to the prior expectation of the exact expected SL, $\widetilde{E}_{\boldsymbol{\theta}}\{R_{SL}(\boldsymbol{\theta}, d)\}$, for four factors logistic regression ($N = 6$).

6.4.5 Discussion

The pseudo-Bayesian decision-theoretic optimal designs for $k = 4$ factor logistic regression, under SE and SL, were found by minimising the objective function using the ACE algorithm. The Bayesian risk was approximated via Quad approximation. Under this example, the exact risk (the MSE and the expected SL) cannot be found for large N; therefore, we applied some approximations to the risk functions, and these were QMC with $B = 1,000$ and $B = 500$, QIMPS with $B = 1,000$, Quad, QMC 2 with $B = 1,000$ and Asym approximation. It was found that the approximate expectation of the QMC approximation (with $B = 1,000$) to the risk is the most computationally expensive approach amongst the others, but it produces potentially more accurate designs (in terms of closeness to the True design found under the approximate expectation of the exact risk).

Furthermore, the approximate expectations of the exact and Quad approximation to the risk are infeasible to be minimised in order to find optimal designs for large N, even with the use of high performance computing. This is because they require more CPU time and memory than the others; the number of responses generated from the binomial distribution, under these approaches, increases exponentially with the number of runs.

We considered the WSMC optimisation technique with the use of the QMC approximation (with $B = 500$) to aid in reducing the CPU time required to solve the optimisation problem in phase I of the ACE algorithm. The resulting design under this technique was compared with the designs found for moderate-sized runs. It was concluded that the approximate expectation of the QMC with $B = 500$ is the most accurate technique and can be used to find designs for moderate-sized runs, followed by WSMC with $B = 500$ and Asym approximation. The latter is simple to implement and does not require much computing time compared with others. However, it produced inaccurate designs for small experimental runs, as it is based on an approximation to the variance matrix of the asymptotic approximate distribution of MLEs. Hence, finding a design under this approach for small runs is not recommended.

6.5 Two factors Poisson regression model

For $k = 2$ factor Poisson regression, there are $p = 3$ unknown parameters. We assume an independent uniform prior distribution to each element of the unknown parameters, i.e.

$$\rho_0 \sim U[-1, 1] \quad \theta_1 \sim U[-2, 2] \quad \theta_2 \sim U[-1, 1]. \tag{6.3}$$

Each element of the design d is generated from $U[-1, 1]$. The prior expectation of the risk function (the MSE and the expected SL) under this example is analytically intractable and results from an $p = 3$-dimensional integral; hence, it is approximated via Quad approximation. Here, the MSE or the expected SL cannot be analytically calculated, even for small number of runs. Therefore, some approximations to the risk functions are considered, namely Asym, ST, QMC with $B = 1,000$ and QIMPS with $B = 1,000$, in addition to QMC 2 (with $B = 1,000$) approximation which it is considered to approximate the trace of the MSE.

Similar to $k = 4$ factor logistic regression model, the objective functions (the approximate expectations of the MSE and the expected SL) that we aim to minimise are coded using a low-level programming language (C++) and the optimal designs are found by applying the ACE algorithm. Furthermore, the algorithm is run on IRIDIS 4.

We compare the resulting designs found under the Quad approximation to the prior expectations of approximations to the risk function by evaluating the approximate expectation of the QMC (with $B = 50,000$) approximation to the risk function at the resulting designs (which we consider near exact evaluation of the approximate expectation of the MSE or the expected SL). In order to make the

comparison valid, the approximated designs are found under exactly the same implementation of ACE as those used to find the QMC design.

We check the validity of the approximate expectations of approximations to the risk function to ensure that they are asymptotically equivalent to the approximate expectation of the QMC (with $B = 50,000$) approximation to the risk function.

6.5.1 Pseudo-Bayesian decision-theoretic D-optimal designs

We find pseudo-Bayesian decision-theoretic D-optimal designs for a different number of runs, and then evaluate the $\widetilde{E}_{\boldsymbol{\theta}}\left\{\log|\widetilde{R}_{SE}(\boldsymbol{\theta},d)|\right\}$, where $R_{SE}(\boldsymbol{\theta},d)$ is approximated via QMC with $B = 50,000$, at all designs to compare their accuracies. In the first row of Figure 6.17, we plot this quantity $\widetilde{E}_{\boldsymbol{\theta}}\left\{\log|\widetilde{R}_{SE}(\boldsymbol{\theta},d)|\right\}$ that was evaluated at each resulting design against the size of N. We compare the accuracy of the approximated designs against QMC design with $B = 1,000$ ($d_{QMC_{L1}}$), which we consider it as the standard to be compared with against other approximated designs, as it has approximately the lowest expected MSE. It is noted that d_{ST} and d_{QIMPS} show inaccuracies, where the former is very far from other designs and needs a huge number of runs to get close to $d_{QMC_{L1}}$, whereas the latter is affected by the noise. It is also noted that d_D performs very well for all sizes of runs.

To check the validity of these approximations, we perform the same steps as those done previously with the logistic regression examples. Hence, we calculate the approximate expectations of the QMC (with $B = 50,000$ and $B = 1,000$), QIMPS and ST approximations to the MSE for $F = 1,000$ different designs, where the fth design is generated by perturbing the QMC design (which was found under the Quad approximation to the prior expectation of the QMC (with $B = 50,000$) approximation to the MSE).

The first row of Figure 6.18 shows the plots of the approximate expectations of QMC with $B = 1,000$ (left), QIMPS (middle) and ST (right) plotted against the Quad approximation to the prior expectation of the QMC (with $B = 50,000$) approximation to the MSE for $N = 6$ (i.e. the smallest number of runs considered). We can see that the approximate expectation of QMC approximation to the MSE appears very accurate, unlike the other approximations that can be inaccurate, especially for designs that are close to the minimum. Nevertheless, the ordering of the designs under these approximations is the same as that for the approximate expectation of the QMC approximation with $B = 50,000$. This indicates that these approximations are minimised for designs close to the design that minimises the approximate expectation of the QMC (with $B = 50,000$) approximation.

The CPU time required to find each design relative to the CPU time used for $d_{QMC_{L1}}$ is calculated in Table 6.10 (i.e. the CPU time for each design / the CPU time for $d_{QMC_{L1}}$). We can see that all designs are less computationally expensive to be found than $d_{QMC_{L1}}$. We also can see that d_D requires the smallest amount of CPU time to be found, as it is found in seconds. Therefore, it could be used instead of the other designs under this example because it exhibits good performance in terms of the expectation of the MSE, and it requires the smallest amount of CPU time.

Table 6.10: The relative computing time (the CPU time for each design / the CPU time for $d_{QMC_{L1}}$) under $k = 2$ factor Poisson regression for different optimal designs found by minimising $\widetilde{E}_{\boldsymbol{\theta}}\left\{\log|\widetilde{R}_{SE}(\boldsymbol{\theta}, d)|\right\}$, and the exact computing time of seconds is given for QMC with B=1000 that is considered as the reference method.

N	d_D	$d_{QMC_{L1}}$	d_{QIMPS}	d_{ST}
5	0.0004	11979	0.14	0.002
6	0.0004	15914	0.14	0.002
7	0.0004	19929	0.14	0.002
8	0.0003	24247	0.13	0.002
9	0.0003	29247	0.13	0.002
10	0.0003	33824	0.16	0.002
11	0.0003	38779	0.14	0.002
12	0.0003	45077	0.14	0.002
13	0.0003	51382	0.14	0.002
14	0.0002	58365	0.13	0.002
15	0.0002	65019	0.14	0.002
16	0.0002	73582	0.14	0.002
20	0.0002	109535	0.14	0.003
24	0.0002	153899	0.13	0.003
28	0.0001	206782	0.12	0.003

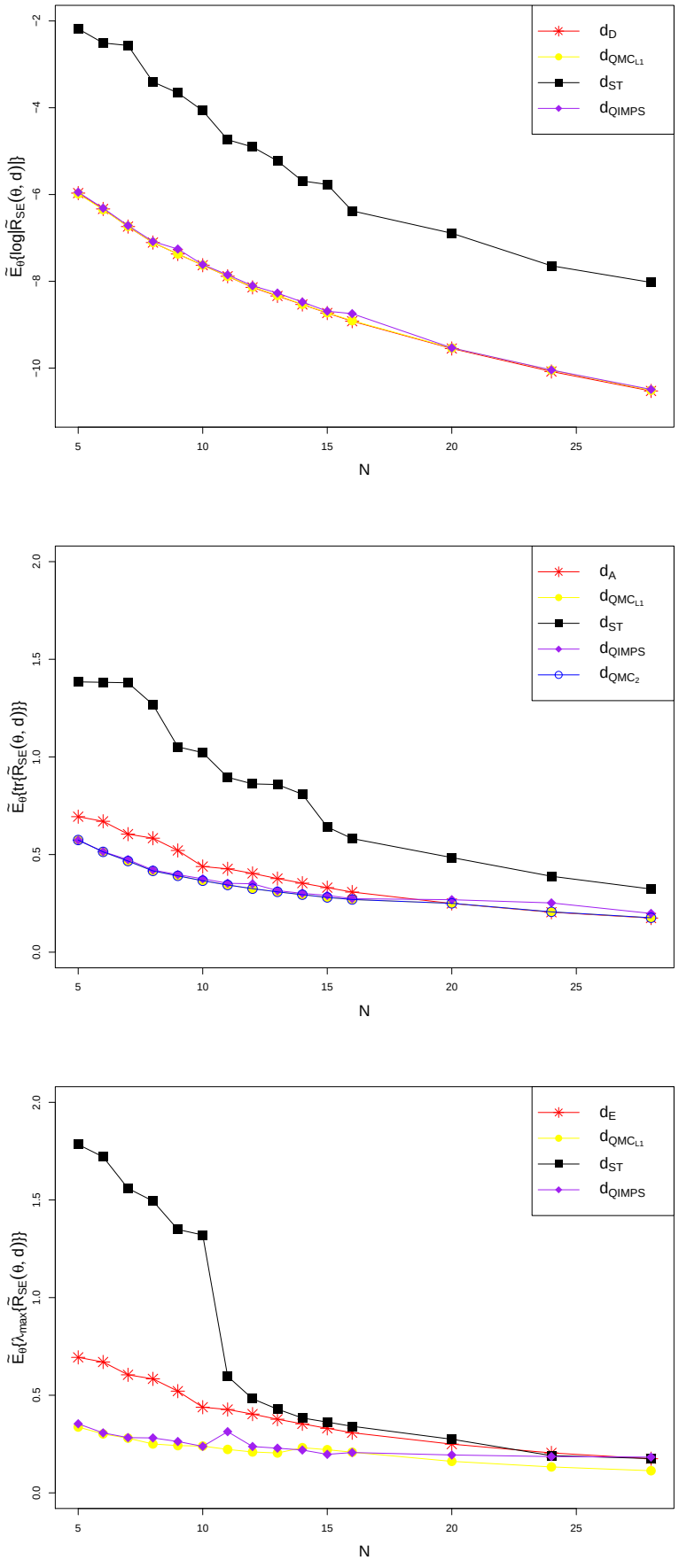

Figure 6.17: Plots of $\widetilde{E}_{\boldsymbol{\theta}}\left\{\log|\widetilde{R}_{SE}(\boldsymbol{\theta}, d)|\right\}$ (the first row), $\widetilde{E}_{\boldsymbol{\theta}}\left\{\mathrm{tr}\left\{\widetilde{R}_{SE}(\boldsymbol{\theta}, d)\right\}\right.$ (the second row) and $\widetilde{E}_{\boldsymbol{\theta}}\left\{\lambda_{max}\left\{\widetilde{R}_{SE}(\boldsymbol{\theta}, d)\right\}\right\}$ (the third row) evaluated at the resulting designs, and plotted against different sizes of N for $k = 2$ factor Poisson regression model.

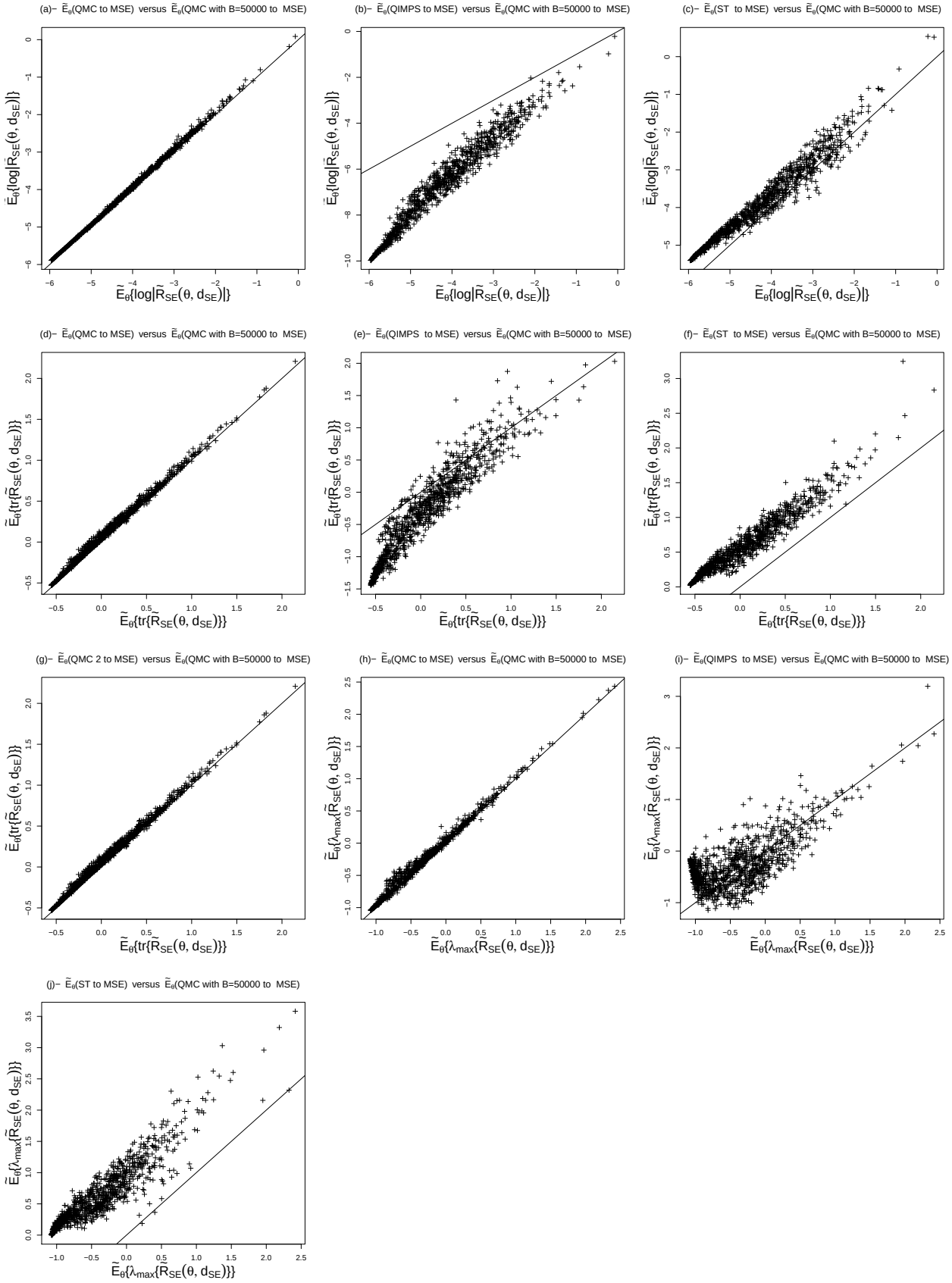

Figure 6.18: Plots of the approximate expectation of approximations to the MSE, plotted against the approximate expectation of the QMC approximation with $B = 50,000$ to the MSE for two factors Poisson regression model.

6.5.2 Pseudo-Bayesian decision-theoretic A-optimal designs

We find pseudo-Bayesian decision-theoretic A-optimal designs for different experimental runs. In the second row of Figure 6.17, we plot the $\widetilde{E}_{\boldsymbol{\theta}}\left\{\text{tr}\left\{\widetilde{R}_{SE}(\boldsymbol{\theta}, d)\right\}\right\}$ that was evaluated at each resulting design against different sizes of N, and compare the accuracy of the approximated designs against QMC design with $B = 1,000$. It is noted that d_{ST} and d_A improve as N gets larger, but the latter does not require a huge number of runs to get close to $d_{QMC_{L1}}$ which are required for the former.

The validity of the Quad approximation to the expectation of QMC, QIMPS, ST and QMC 2 approximation to the MSE is checked. The approximate expectations of these approximations are plotted against the approximate expectation of QMC approximation (with $B = 50,000$) to the MSE, in the second row of Figure 6.18 for the first three approximations, respectively, and in the third row of this figure (left) for the latter approximation.

Although the approximate expectations of the QIMPS and ST approximations to the MSE are minimised for design close to the design that minimises the approximate expectation of the QMC approximation with $B = 50,000$, these two approximations exhibit inaccuracies. However, the expectations of QMC with $B = 1,000$ and QMC 2 approximations exhibit high accuracies as they lie on the line that aids in the comparison, indicating that these approximations have the same accuracy as the approximate expectation of the QMC approximation (with $B = 50,000$) to the MSE.

Table 6.11 shows the RT for different designs found under the approximate expectation of different approximations to the MSE. Similar to Section 6.5.1, all designs do not require more CPU time to be found than $d_{QMC_{L1}}$. It is noted that d_{QMC2} requires CPU time to be found as approximately the same as $d_{QMC_{L1}}$.

Table 6.11: The relative computing time (the CPU time for each design / the CPU time for $d_{QMC_{L1}}$) under $k = 2$ factor Poisson regression model for different optimal designs found by minimising $\widetilde{E}_{\boldsymbol{\theta}}\left\{\text{tr}\left\{\widetilde{R}_{SE}(\boldsymbol{\theta}, d)\right\}\right\}$, and the exact computing time of seconds is given for QMC with B=1000 that is considered as the reference method.

N	d_A	$d_{QMC_{L1}}$	d_{QIMPS}	d_{ST}	d_{QMC2}
5	0.0006	12499	0.14	0.002	0.95
6	0.0005	16319	0.13	0.002	0.96
7	0.0007	21475	0.13	0.002	0.91
8	0.0005	25337	0.13	0.002	0.95
9	0.0004	31361	0.12	0.002	0.91
10	0.0003	35036	0.14	0.002	0.96
11	0.0003	42616	0.12	0.002	0.92
12	0.0003	45418	0.13	0.002	0.97
13	0.0003	55207	0.12	0.002	0.91
14	0.0002	57995	0.13	0.002	0.98
15	0.0002	64292	0.14	0.002	1
16	0.0002	72194	0.13	0.002	1
20	0.0002	105900	0.14	0.003	0.99
24	0.0002	148353	0.13	0.003	0.99
28	0.0001	200412	0.12	0.003	0.99

6.5.3 Pseudo-Bayesian decision-theoretic E-optimal designs

Pseudo-Bayesian decision-theoretic E-optimal designs are found for different sizes of runs. In the third row of Figure 6.17, the $\widetilde{E}_{\boldsymbol{\theta}}\left\{\lambda_{max}\left\{\widetilde{R}_{SE}(\boldsymbol{\theta}, d)\right\}\right\}$ was evaluated at the resulting designs, and plotted against various values of runs. Similar comments apply here as that for pseudo-Bayesian decision-theoretic A-optimal designs under this example.

The validity of the Quad approximation to the expectations of QMC, QIMPS and ST approximations to the MSE are checked, where the approximate expectations of these approximations are plotted against the approximate expectation of QMC approximations with $B = 50,000$ in the third row of Figure 6.18 for the first two approximations (middle and right plot, respectively), and in the fourth row of this figure for the latter approximation. Similar comments apply here as that for Section 6.5.1, where the expectations of QIMPS and ST approximations to the

MSE are minimised for a design close to the design that minimises the expectation of the QMC approximation with $B = 50,000$. These two approximations exhibit inaccuracies, especially the former exhibits noisy data. By contrast, the approximate expectation of QMC approximation to the MSE with $B = 1,000$ appears very accurate. This means that this approximation has the same accuracy as the approximate expectation of QMC approximation to the MSE with $B = 50,000$.

In Table 6.12, we calculate the RT for all designs found under this example using different approaches. Similar comments apply as that for pseudo-Bayesian decision-theoretic D-optimal designs under Poisson regression model.

Table 6.12: The relative computing time (the CPU time for each design / the CPU time for $d_{QMC_{L1}}$) under $k = 2$ factor Poisson regression model for different optimal designs found by minimising $\widetilde{E}_{\boldsymbol{\theta}}\left\{\lambda_{max}\left\{\widetilde{R}_{SE}(\boldsymbol{\theta}, d)\right\}\right\}$, and the exact computing time of seconds is given for QMC with B=1000 that is considered as the reference method.

N	d_E	$d_{QMC_{L1}}$	d_{QIMPS}	d_{ST}
5	0.0005	11906	0.15	0.002
6	0.0004	15924	0.14	0.002
7	0.0003	20066	0.13	0.002
8	0.0003	24041	0.13	0.002
9	0.0003	29029	0.13	0.002
10	0.0003	34153	0.14	0.002
11	0.0003	39285	0.13	0.002
12	0.0003	45090	0.13	0.002
13	0.0003	50759	0.13	0.002
14	0.0002	56817	0.13	0.002
15	0.0002	64236	0.14	0.002
16	0.0002	71650	0.14	0.002
20	0.0002	105743	0.14	0.003
24	0.0002	144952	0.13	0.003
28	0.0002	198232	0.13	0.003

6.5.4 Pseudo-Bayesian decision-theoretic optimal designs under the SL

We find pseudo-Bayesian decision-theoretic optimal designs under the sign loss for moderate-sized runs. Here, we consider the QMC and Asym approximations to

Table 6.13: The relative computing time (the CPU time for each design / the CPU time for $d_{QMC_{L1}}$) under $k = 2$ factor Poisson regression model for d_{Asym} and $d_{QMC_{L1}}$ found by minimising $\widetilde{E}_{\boldsymbol{\theta}}\left\{\widetilde{R}_{SL}(\boldsymbol{\theta}, d)\right\}$, and the exact computing time of seconds is given for QMC with B=1000 that is considered as the reference method.

N	d_{Asym}	$d_{QMC_{L1}}$
5	0.004	11839
6	0.003	15746
7	0.003	20239
8	0.003	25226
9	0.003	29701
10	0.003	34605
11	0.002	40099
12	0.002	45261
13	0.002	52196
14	0.002	59356
15	0.002	66048
16	0.002	73692
20	0.002	107340

the expected SL, and exclude the QIMPS due to the inaccuracies that exhibited this approximation.

In Figure 6.19, we plot $\widetilde{E}_{\boldsymbol{\theta}}\left\{\widetilde{R}_{SL}(\boldsymbol{\theta}, d)\right\}$ that evaluated at $d_{QMC_{L1}}$ and d_{Asym} against various values of N to compare the accuracy of these designs. Interestingly, d_{Asym} performs very well where they are very similar to $d_{QMC_{L1}}$. As a result, we could obtain a negligible difference in the expectation of the expected SL for values of runs over approximately seven by using d_{Asym}.

The validity of the Quad approximation to the expectation of QMC (with $B = 1000$) approximation to the expected SL is checked in Figure 6.20. This is to ensure that it is asymptotically equivalent to the Quad approximation to the expectation of QMC (with $B = 50,000$) approximation to the expected SL. Apparently, the expectation of the QMC approximation with $B = 1000$ appear very accurate.

We calculate the RT for each design in Table 6.13, and found that d_{Asym} is much less expensive than $d_{QMC_{L1}}$, as it is found in seconds for a range of runs under this example.

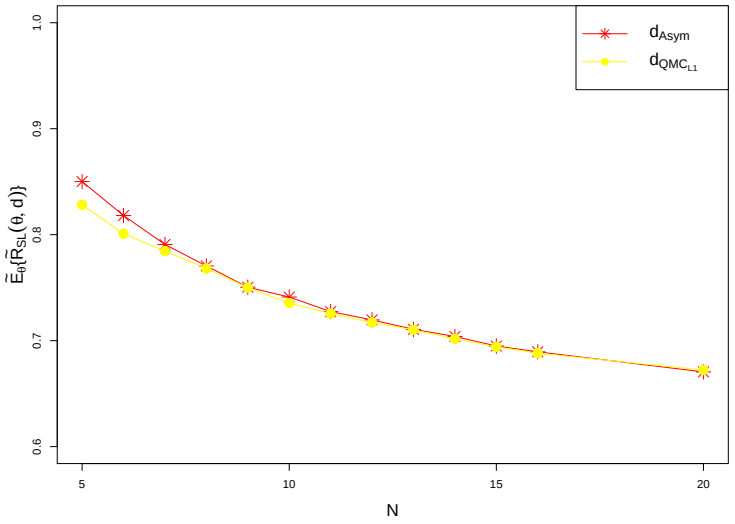

Figure 6.19: Plot of the Quad approximation to the prior expectation of the exact expected SL, where the $\widetilde{E}_{\boldsymbol{\theta}}\left\{\widetilde{R}_{SL}(\boldsymbol{\theta}, d)\right\}$ was evaluated at d_{Asym} (red star) and $d_{QMC_{L1}}$ (yellow open circle), plotted against different values of N.

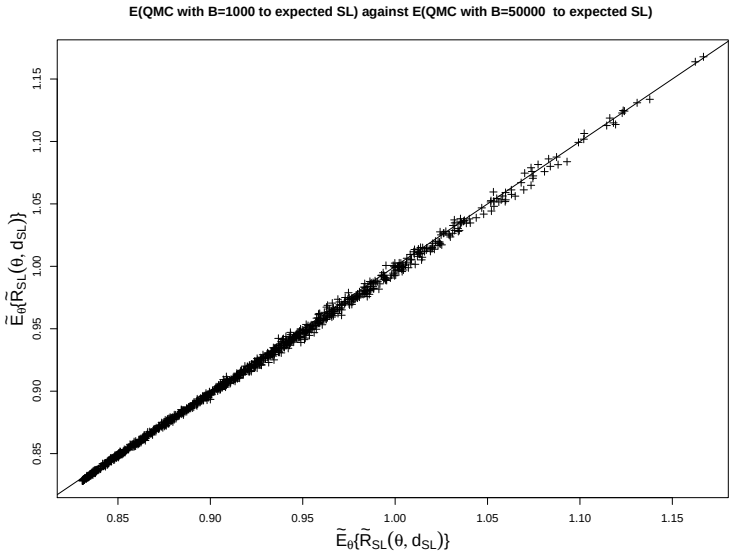

Figure 6.20: Plot of the Quad approximation to the prior expectation of QMC (with $B = 1000$) approximation to the expected SL, $\widetilde{E}_{\boldsymbol{\theta}}\left\{\widetilde{R}_{SL}(\boldsymbol{\theta}, d)\right\}$, plotted against the Quad approximation to the prior expectation of QMC (with $B = 50,000$) approximation to the expected SL, $\widetilde{E}_{\boldsymbol{\theta}}\left\{\widetilde{R}_{SL}(\boldsymbol{\theta}, d)\right\}$, for two factors Poisson regression ($N = 6$).

6.5.5 Discussion

Pseudo-Bayesian decision-theoretic optimal designs for $k = 2$ factor Poisson regression model under the SE and SL were found by minimising the approximate

expectation of the risk function (i.e the MSE or the expected SL based on the considered loss function) using the ACE algorithm. The Bayesian risk was approximated via Quad approximation. By contrast, the risk cannot be analytically calculated, even for a small number of runs. Therefore, some approximations to the risk function were proposed to allow us to find designs under this example, and these were the QMC with $B = 1,000$, $QIMPS$ with $B = 1,000$, ST and $Asym$ approximations.

The accuracy of the approximated designs found under the proposed approaches was investigated and compared by evaluating the Quad approximation to the expectation of QMC (with $B = 50,000$) approximation to the risk function at the resulting designs; we considered it a near-exact evaluation of the Quad approximation to the expectation of the risk. As a result, we found that the QMC design performs very well and has approximately the lowest expected risk; hence, we considered it as the standard for comparison against other approximated designs. It was also found that the asymptotic designs require much less CPU time than the other designs; however, they showed inaccuracies for small N.

The validity of the approximate expectation of approximations to the risk functions was investigated by perturbing the QMC design found under the Quad approximation to the expectation of QMC (with $B = 50,000$) approximation to the risk. It was found that the approximate expectations of the QIMPS and ST approaches can be inaccurate, although these approximations were minimised for a design near the design that minimised the expectation of the QMC (with $B = 50,000$) approximation to the risk.

Under pseudo-Bayesian decision-theoretic A-optimal designs, the approximate expectations of the QMC and QMC 2 approximations to the MSE were the most accurate technique noted for all values of runs. The validity of these approximations was also checked, and they exhibited high accuracy.

6.6 Summary

In this chapter, we found local and pseudo-Bayesian decision-theoretic optimal designs under SE loss for the simple logistic regression model. For the four factors logistic regression and two factors Poisson regression models, we found pseudo-Bayesian decision-theoretic optimal designs under SE and SL. The objective function under the local optimality was given by the risk function, and was approximated via the QMC, Asym, ST and Quad approximation. Under pseudo-Bayesian optimality, the objective function was the expected risk (either the expected MSE

or the expectation of the expected SL), where the prior expectation was approximated via the Quad rule, whereas the risk was approximated via the QIMPS in addition to those used under local optimality. The QIMPS approximation to the risk was used to reduce the intensive CPU time required for computing the QMC approximation.

For $k = 4$ factors logistic regression, finding the pseudo-Bayesian decision-theoretic optimal design was infeasible for a medium-size of experimental runs. In order to address this challenge, the warm-starting optimisation technique was proposed to be used in phase I of the ACE algorithm as well as QMC with a reduced sample (with $B = 500$). As a result, we found that the QMC approximation with $B = 500$ produces accurate designs, although it is computationally expensive. It was also found that designs under the warm-starting technique have better accuracy than the design found under asymptotic approximation.

In conclusion, the QMC approximation to the risk is the most accurate technique amongst the proposed approximations that were applied to find designs tailored to the experimental aim of parameter estimation. This has been concluded after investigating all considered examples in this chapter. Some examples of decision-theoretic optimal designs found in this chapter are provided in "Appendix D".

Chapter 7

Decision-theoretic Optimal Designs for Model Discrimination

7.1 Introduction

Often there are a number of competing statistical models (or likelihoods) for a set of responses and the experimental aim is to discriminate between these models. Suppose the set of models is denoted $\mathcal{M}$, and model $m \in \mathcal{M}$ has a probability distribution given by $f_m(\mathbf{Y}; \boldsymbol{\theta}_m, d)$ where $\boldsymbol{\theta}_m$ is the $p_m \times 1$ vector of unknown parameters for this model.

In this thesis, we consider an information criterion as the approach to select the best model. The information criterion aims to trade-off model complexity against model fit. The selected model is given by

$$\hat{m}(\mathbf{Y}) = \arg\min_{m \in \mathcal{M}} IC(m),$$

where $IC(m)$ denotes the value of the information criterion for model $m \in \mathcal{M}$. We consider the Akaike information criterion (AIC), formulated by Akaike (1974), given by

$$\text{AIC}(m) = -2l_m(\hat{\boldsymbol{\theta}}_m(\mathbf{Y}); \mathbf{Y}, d) + 2p_m. \tag{7.1}$$

The reason of using AIC over any other model selection approach is that it allows us to compare multiple non-nested models at the same time. The set of models we consider are those arising from variable selection. Suppose there are a total of k controllable factors under consideration. These k factors can either be included in the model or not. This means there are a total of $|\mathcal{M}| = 2^k$ different models.

In this chapter, we extend our investigation for finding decision-theoretic optimal designs for experiments involving GLMs to the experimental aim of model

discrimination. Finding such designs is accomplished by minimising a risk function defined as the expectation of an appropriate loss function $\lambda(\hat{m}(\mathbf{Y}), m|\boldsymbol{\theta}_m, d)$ reflecting the aim of discriminating a model from a set of competing statistical models. The expectation is taken with respect to the distribution of the data $\mathbf{Y}$ given a true model m and quantities of interest $\boldsymbol{\theta}_m$ for model m. The loss function $\lambda(\hat{m}(\mathbf{Y}), m|\boldsymbol{\theta}_m, d)$ specifies the loss of identifying the true model m with the selected model $\hat{m}(\mathbf{Y})$ using $\mathbf{Y}$ that obtained from d given the true model and its parameters $\boldsymbol{\theta}_m|m$. In particular, we aim to find pseudo-Bayesian decision-theoretic optimal designs. Details for the objective function that we aim to minimise for finding such designs is given in the next section.

7.2 The risk function under model discrimination

When the aim of an experiment is to discriminate between a number of competing statistical models, the risk function is defined as the expectation of a loss function, with respect to the distribution of $\mathbf{Y}$ given $\boldsymbol{\theta}_m$ and m. Mathematically,

$$\begin{aligned} R(d, \boldsymbol{\theta}_m, m) &= E_{\mathbf{Y}|\boldsymbol{\theta}_m, m}\left[\lambda(\hat{m}(\mathbf{Y}), m|\boldsymbol{\theta}_m, d)\right], \\ &= \int \lambda(\hat{m}(\mathbf{Y}), m|\boldsymbol{\theta}_m, d) f_m(\mathbf{Y}; \boldsymbol{\theta}_m, d) d\mathbf{Y}, \end{aligned} \tag{7.2}$$

where $\mathbf{Y}$ is a continuous variable, and the model parameters $\boldsymbol{\theta}_m$ are considered to be fixed. In the case of Y is a discrete variable, a sum is used instead of the integral. Here, the risk function $R(d, \boldsymbol{\theta}_m, m)$ depends on both the model parameters for model m and the model m. Hence, pseudo-Bayesian decision-theoretic optimal designs minimise the objective function

$$\begin{aligned} \Psi_{DT}(d) &= E_{\boldsymbol{\theta}_m, m}\left[R(d, \boldsymbol{\theta}_m, m)\right], \\ &= \sum_{m \in \mathcal{M}} \pi(m) E_{\boldsymbol{\theta}_m|m}\left[R(d, \boldsymbol{\theta}_m, m)\right], \\ &= \sum_{m \in \mathcal{M}} \pi(m) \int_{\boldsymbol{\Theta}_m} R(d, \boldsymbol{\theta}_m, m) \pi(\boldsymbol{\theta}_m) d\boldsymbol{\theta}_m, \end{aligned} \tag{7.3}$$

over the design space $\mathcal{X}$. The Bayesian risk in (7.3) is not available in closed form resulting from $|\mathcal{M}|$ p-dimensional integrals. Typically, it needs a numerical evaluation of the p-dimensional integrals. In addition, the risk function $R(d, \boldsymbol{\theta}_m, m)$ is not available in closed form and results from an N-dimensional integral. Hence, approximation methods to the Bayesian risk and risk function are required.

7.3 Methods for approximating the Bayesian risk and risk function

It is extremely complicated to find pseudo-Bayesian decision-theoretic optimal design when the model uncertainty exists as well as the objective function is analytically intractable, and this is due to the substantial computational challenges of minimising an analytically intractable objective function over a multi-dimensional design space. Therefore, we aim to find near exact pseudo-Bayesian decision-theoretic optimal designs. In this section, approximation methods to the Bayesian risk and risk function are described.

7.3.1 Approximating the Bayesian risk

Suppose there are q models, $q = 1, \ldots, 2^k$; hence, we use the deterministic quadrature approximation technique to the Bayesian risk, where the $|\mathcal{M}|$ p-dimensional integrals in (7.3) can be approximated using the radial-spherical integration rule with non-constant weights w_{mqh} and abscissae $\boldsymbol{\theta}_{mqh}$, i.e.

$$\begin{aligned}\Psi_{DT}(d) &= E_{\boldsymbol{\theta}_m,m}\left[R(d,\boldsymbol{\theta}_m,m)\right],\\ &= \sum_{q\in\mathcal{M}} \pi_{mq} E_{\boldsymbol{\theta}_{mq}|mq}\left[R(d,\boldsymbol{\theta}_{mq},mq)\right],\\ &= \sum_{q\in\mathcal{M}} \pi_{mq} \sum_{h=1}^{H} w_{mqh} R(d,\boldsymbol{\theta}_{mqh},mq),\end{aligned} \tag{7.4}$$

where π_{mq} is a prior model probability for the qth model.

7.3.2 Approximating the risk function

In this section, we approximate the risk function given in (7.2) via QMC, i.e.

$$R(d;\boldsymbol{\theta}_m,m) \approx \frac{1}{B}\sum_{b=1}^{B} \lambda(\hat{m}(\mathbf{Y}(\mathbf{u}_b)),m|\boldsymbol{\theta}_m,d), \tag{7.5}$$

where QMC approximation would sample $\{\mathbf{u}_b\}_{b=1}^{B}$, $\mathbf{u}_b = (u_1,\ldots,u_N)^T$, of size B randomly from a uniform distribution $U[0,1]$. The generated sample from $U[0,1]$ is then fixed. Following that, this sample and a design d are used to generate responses $\mathbf{Y}$ from the inverse cumulative distribution function of a given distribution $f_m(\mathbf{Y}(\mathbf{u});\boldsymbol{\theta}_m,d)$, for each given set of the true values of $\boldsymbol{\theta}_m$. Here, the QMC approximation to the risk is considered because it is the most accurate approximation to the risk found across examples in Chapter 6. The detailed steps of this approximation method are

1. Generate a sample $\mathbf{u}_1, \ldots, \mathbf{u}_B$ from the uniform distribution $U(0,1)$, and make it fixed.
2. Generate a model $m \subset \mathcal{M}$.
3. Specify true values for $\boldsymbol{\theta}_m$ under model m.
4. Generate a design $d = \{\mathbf{z}_1, \ldots, \mathbf{z}_N\}$ from a given design space.
5. Generate $\mathbf{Y}_1, \ldots, \mathbf{Y}_B$ responses from the inverse cumulative distribution function of a given distribution $f_m(\mathbf{Y}(\mathbf{u}); \boldsymbol{\theta}_m, d)$, given $\mathbf{u}_1, \ldots, \mathbf{u}_B$, $\boldsymbol{\theta}_m$, m and d.
6. For $b = 1, \ldots, B$, complete steps (i) and (ii), where each iteration forms the bth element of a vector of length B which consists of the selected models $\hat{m}(\mathbf{Y}(\mathbf{u}_b))$.
   (i) Fit all $|\mathcal{M}| = 2^k$ models and calculate the $AIC(m)$ in (7.1) for each model, given $\mathbf{Y}(\mathbf{u}_b)$.
   (ii) Calculate $\hat{m}(\mathbf{Y}(\mathbf{u}_b)) = \arg\min_{m \in \mathcal{M}} AIC(m)$.
7. Approximate $R(d; \boldsymbol{\theta}_m, m)$ via QMC given in (7.5).

Under the pseudo-Bayesian approach, the true values are the quadrature abscissae $\boldsymbol{\theta}_{mq1}, \ldots, \boldsymbol{\theta}_{mqH}$ from the quadrature rule and based on the prior distribution of $\boldsymbol{\theta}_{mq}$. Hence, we generate $B \times H$ responses from $f_{mq}(\mathbf{Y}(\mathbf{u}); \boldsymbol{\theta}_{mq}, d)$ (i.e. responses are generated for each given set of the true values for the parameters $\boldsymbol{\theta}_{mq}$). By substituting (7.5) in (7.4), we obtain the objective function

$$\Psi_{DT}(d) = \frac{1}{B} \sum_{q \in \mathcal{M}} \pi_{mq} \sum_{h=1}^{H} w_{mqh} \sum_{b=1}^{B} \lambda(\hat{m}(\mathbf{Y}(\mathbf{u}_b)), mq | \boldsymbol{\theta}_{mqh}, d), \tag{7.6}$$

that we aim to minimise in order to find the pseudo-Bayesian decision-theoretic optimal design.

Note that we use the ACE algorithm that is run on IRIDIS 4 to minimise the objective function given in (7.6) for finding the optimal designs, where the objective function is coded using C++. In this chapter, we find pseudo-Bayesian decision-theoretic optimal designs under two loss functions: the $0-1$ (Section 7.4.1) and FSL (Section 7.4.2), for $k = 2$ factor logistic regression model (Section 7.5).

7.4 Examples of loss functions

In this section, we give some examples of default loss functions tailored to the experimental aim of model discrimination.

7.4.1 $0-1$ loss

The $0-1$ loss can be expressed as one minus an indicator function of the equality of the selected model $\hat{m}(\mathbf{Y})$ and true model $m \in \mathcal{M}$. When the selected model is misidentified (i.e it is not equal to the true model), the loss is 1; otherwise, the loss is 0. Mathematically,

$$\lambda_{0-1}(\hat{m}(\mathbf{Y}), m|\boldsymbol{\theta}_m, d) = 1 - \mathbb{1}\left(\hat{m}(\mathbf{Y}) = m\right). \tag{7.7}$$

7.4.2 Factor-specific loss

The novel factor-specific loss (FSL) concerns the features of the model, not the model itself. The FSL is recommended to be used in screening experiments instead of $0-1$ loss, as the aim is to quantify the loss of identifying the experimental factors that have a significant influence on the response in an experiment. The FSL quantifies the sum of the losses as the selected model includes or does not include the experimental factors when it is compared with the true model. When the selected model includes a factor that does not exist in the true model, the loss is δ, where $0 < \delta \leq 2$. This reflects a type I error in statistical inference (rejecting the null hypothesis when it is true). When, however, the selected model does not include the factor that exists in the true model, the loss is 1, hence a Type II error (failing to reject the null hypothesis when it is false). The loss is 0 if the factor of the selected model exists in the true model.

To formalise, consider $z_1, \ldots, z_k$ are experimental factors under m and $\hat{z}_1, \ldots, \hat{z}_k$ are factors under $\hat{m}(\mathbf{Y})$. The FSL is denoted by $\lambda_{FSL}(\hat{z}_j, z_j)$, $j = 1, \ldots, k$, and given such that

$$\lambda_{FSL}(\hat{z}_j, z_j) = \sum_{j=1}^{k} \lambda_j(\hat{z}_j, z_j), \tag{7.8}$$

where the quantity $\lambda_j(\hat{z}_j, z_j)$ is presented in Table 7.1, and $\hat{z}_j$ can be expressed as

$$\hat{z}_j = \begin{cases} 1 & \text{if } \hat{m}(\mathbf{Y}) \text{ includes } j; \quad j = 1, \ldots, k, \\ 0 & \text{otherwise,} \end{cases}$$

and similarly for z_j. Experimenters can choose the value of δ, based on that value Type I and II errors differ in terms of the worse, i.e.

$$\begin{aligned} \delta \in [0, 1) \quad &\Rightarrow \quad \text{Type II error becomes worse,} \\ \delta = 1 \quad &\Rightarrow \quad \text{Type I error} = \text{Type II error,} \\ \delta > 1 \quad &\Rightarrow \quad \text{Type I error becomes worse.} \end{aligned}$$

Table 7.1: The loss of misidentifying factors under the selected model $\hat{m}(\mathbf{Y})$ compared with factors under the true model m, where 1 and 0 refer to the inclusion and exclusion of the factors under the specified models, respectively.

$\lambda_j(\hat{z}_j, z_j)$		m	
		0	1
$\hat{m}(\mathbf{Y})$	0	0	1
	1	δ_j	0

7.5 Two factors logistic regression model

Under $k = 2$ factor logistic regression, the total number of models is $|\mathcal{M}| = 2^2 = 4$, and these are

1. $m1 : \log\left(\frac{\pi_i}{1-\pi_i}\right) = \rho_0$.
2. $m2 : \log\left(\frac{\pi_i}{1-\pi_i}\right) = \rho_0 + \theta_1 z_{i1}$.
3. $m3 : \log\left(\frac{\pi_i}{1-\pi_i}\right) = \rho_0 + \theta_2 z_{i2}$.
4. $m4 : \log\left(\frac{\pi_i}{1-\pi_i}\right) = \rho_0 + \sum_{j=1}^{2} \theta_j z_{ij}$.

We assume for each element of unknown parameters an independent uniform prior distribution, i.e.

$$\rho_0 \sim U[-3, 3] \quad \theta_1 \sim U[4, 10] \quad \theta_2 \sim U[5, 11]. \tag{7.9}$$

Each element of the design d is assumed to be generated from $U[1, 1]$. Also, we assume 4 different prior model probabilities for m, i.e.

- $\pi_1(m) = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$.
- $\pi_2(m) = (\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{2})$.
- $\pi_3(m) = (\frac{1}{8}, \frac{1}{8}, \frac{1}{4}, \frac{1}{2})$.
- $\pi_4(m) = (\frac{1}{8}, \frac{1}{4}, \frac{1}{8}, \frac{1}{2})$.

To illustrate, consider the first prior model probabilities $\pi_1(m)$, each element of this prior refers to the probability assigned for the qth model, $q = 1, \ldots, 4$, respectively. Here, all the four models have equal probabilities.

The objective function that we aim to minimise in order to find pseudo-Bayesian decision-theoretic optimal designs, under the $0 - 1$ or FSL, is given by the Quad approximation to the prior expectation of the risk function, $\widetilde{E}_{\boldsymbol{\theta}_m, m}\{R(\boldsymbol{\theta}_m, m, d)\}$.

The risk function is either the expected $0-1$ or the expected FSL, based on the choice of the loss function. For $k = 2$ factors logistic regression model, the risk function can only be analytically calculated for small experiments (i.e. small values of N).

For large number of runs, it is extremely complicated to find pseudo-Bayesian decision-theoretic optimal design when the model uncertainty exists as well as the objective function is analytically intractable. Therefore, we apply the QMC approximation (with $B = 1000$ and $B = 500$) to the risk function $R(\boldsymbol{\theta}_m, m, d)$. We then compare $d_{QMC_{L1}}$ and $d_{QMC_{L2}}$, found under the approximate expectation of the QMC approximation to the risk with $B = 1000$ and $B = 500$, respectively, against d_T that was found under the approximate expectation of the exact risk. The comparison is in terms of their performance (the expected risk) and CPU time (similar to Chapter 6 we consider the RT for resulting designs, i.e. the CPU time for each design / the CPU time for d_T).

7.5.1 Pseudo-Bayesian decision-theoretic optimal designs under 0-1 loss

For each model probabilities $\pi(m)$ illustrated in Section 7.5, we find pseudo-Bayesian decision-theoretic optimal designs under the $0-1$ loss for $k = 2$ factor logistic regression for various values of N. In order to assess the performance of the resulting designs, we evaluate the approximate expectation of the exact expected $0-1$ loss, $\widetilde{E}_{\boldsymbol{\theta}_m,m}\{R_{0-1}(\boldsymbol{\theta}_m, m, d)\}$, at the resulting designs. The performance and RT for d_T, $d_{QMC_{L1}}$ and $d_{QMC_{L2}}$ are outlined in Table 7.2.

In Figure 7.1, similar to what we have done for $k = 4$ factor logistic regression model in Chapter 6, we check the validity of the approximate expectation of the QMC (with $B = 1000$ and $B = 500$) approximation to the expected 0-1 loss for each $\pi(m)$ (the smallest number of runs considered $N = 4$).

7.5.2 Pseudo-Bayesian decision-theoretic optimal designs under factor-specific loss

Similar to Section 7.5.1, pseudo-Bayesian decision-theoretic optimal designs, for $k = 2$ factor logistic regression model and various values of N, are found for each model probabilities, but here under the FSL. We set $\delta = 0.5$ to find such designs under this loss function, and evaluate the approximate expectation of the exact expected FSL loss, $\widetilde{E}_{\boldsymbol{\theta}_m,m}\{R_{FSL}(\boldsymbol{\theta}_m, m, d)\}$, at the resulting designs to assess the performance of these designs. We outline the performance and RT for d_T, $d_{QMC_{L1}}$ and $d_{QMC_{L2}}$ in Table 7.3. In Figure 7.2, the validity of the approximate expectation of the QMC (with $B = 1000$ and $B = 500$) approximation to the expected FSL is checked for each $\pi(m)$ (again the smallest number of runs considered $N = 4$).

Table 7.2: The Quad approximation to the prior expectation of the exact expected $0-1$ loss evaluated at d_T, $d_{QMC_{L1}}$ and $d_{QMC_{L2}}$. The relative computing time (the CPU time for each design / the CPU time for d_T) for these designs is given between (.), and the exact computing time of seconds is given for the exact risk that is considered as the reference method. The stars $(*)$ under each design mean that the corresponding design cannot be computed or requires highly expensive time to be found.

$\pi(m)$	N	d_T	$d_{QMC_{L1}}$	$d_{QMC_{L2}}$
	4	0.30 (967)	0.30 (60.66)	0.30 (30.53)
	5	0.12 (3246)	0.12 (35.30)	0.13 (17.59)
	6	0.09 (7629)	0.09 (19.25)	0.09 (9.43)
$\pi_1(m)$	7	0.07 (18894)	0.07 (10.35)	0.07 (5.32)
	8	0.07 (43517)	*	0.07 (2.81)
	9	0.07 (102020)	*	0.07 (1.46)
	10	0.07 (223801)	*	0.07 (0.80)
	4	0.53 (966)	0.53 (60.74)	0.53 (30.68)
	5	0.11 (3316)	0.11 (35.87)	0.11 (17.38)
	6	0.08 (7695)	0.09 (20)	0.09 (9.72)
$\pi_2(m)$	7	0.08 (19377)	0.07 (10.44)	0.08 (5.26)
	8	0.07 (43439)	*	0.07 (2.79)
	9	0.06 (102085)	*	0.06 (1.48)
	10	0.05 (224874)	*	0.06 (0.79)
	4	0.52 (961)	0.53 (60.88)	0.53 (30.66)
	5	0.10 (3240)	0.10 (35.97)	0.10 (17.82)
	6	0.08 (7872)	0.08 (19.03)	0.09 (8.99)
$\pi_3(m)$	7	0.07 (19613)	0.07 (10.23)	0.07 (5.16)
	8	0.06 (42862)	*	0.06 (2.85)
	9	0.05 (103638)	*	0.06 (1.47)
	10	0.05 (226855)	*	0.05 (0.79)
	4	0.53 (1615)	0.53 (36.35)	0.53 (18.14)
	5	0.11 (3398)	0.10 (33.83)	0.09 (17.32)
	6	0.08 (8221)	0.08 (18.66)	0.08 (9.40)
$\pi_4(m)$	7	0.07 (19406)	0.07 (10.09)	0.07 (5.31)
	8	0.06 (43794)	*	0.06 (2.78)
	9	0.05 (103416)	*	0.05 (1.47)
	10	0.05 (224277)	*	0.05 (0.78)

Table 7.3: The Quad approximation to the prior expectation of the exact expected FSL evaluated at d_T, $d_{QMC_{L1}}$ and $d_{QMC_{L2}}$. The relative computing time (the CPU time for each design / the CPU time for d_T) for these designs is given between (.), and the exact computing time of seconds is given for the exact risk that is considered as the reference method. The stars ($*$) under each design mean that the corresponding design cannot be computed or requires highly expensive time to be found.

$\pi(m)$	N	d_T	$d_{QMC_{L1}}$	$d_{QMC_{L2}}$
	4	0.26 (1089)	0.26 (67.72)	0.26(34.21)
	5	0.03 (3361)	0.03 (34.60)	0.03 (17.57)
	6	0.02 (7679)	0.02 (20.26)	0.02 (9.74)
$\pi_1(m)$	7	0.02 (18655)	0.01 (10.77)	0.02 (5.48)
	8	0.01 (43657)	*	0.01 (3.02)
	9	0.01 (103394)	*	0.01 (1.47)
	10	0.004 (226627)	*	0.004 (0.82)
	4	0.51 (1085)	0.51 (66.81)	0.51 (33.94)
	5	0.04 (3304)	0.04 (35.22)	0.04 (18.24)
	6	0.03 (8384)	0.04 (17.25)	0.03 (9.00)
$\pi_2(m)$	7	0.02 (20246)	0.02 (9.97)	0.02 (5.19)
	8	0.01 (43626)	*	0.02 (2.80)
	9	0.01 (104114)	*	0.01 (1.50)
	10	0.005 (227788)	*	0.01 (0.78)
	4	0.51 (1104)	0.51 (64.90)	0.51 (32.59)
	5	0.04 (3399)	0.04 (35.29)	0.03 (17.61)
	6	0.03 (8369)	0.03 (18.61)	0.02 (8.32)
$\pi_3(m)$	7	0.02 (19962)	0.03 (10.38)	0.02 (4.93)
	8	0.02 (45235)	*	0.02 (2.63)
	9	0.01 (105296)	*	0.01 (1.48)
	10	0.007 (229309)	*	0.01 (0.78)
	4	0.51 (1100)	0.51 (65.68)	0.51 (33.63)
	5	0.04 (3356)	0.04 (34.99)	0.03 (17.34)
	6	0.03 (8200)	0.04 (17.68)	0.03 (9.46)
$\pi_4(m)$	7	0.02 (19538)	0.03 (10.39)	0.03 (5.10)
	8	0.01 (44908)	*	0.02 (2.69)
	9	0.01 (104832)	*	0.01 (1.45)
	10	0.006 (227181)	*	0.01 (0.82)

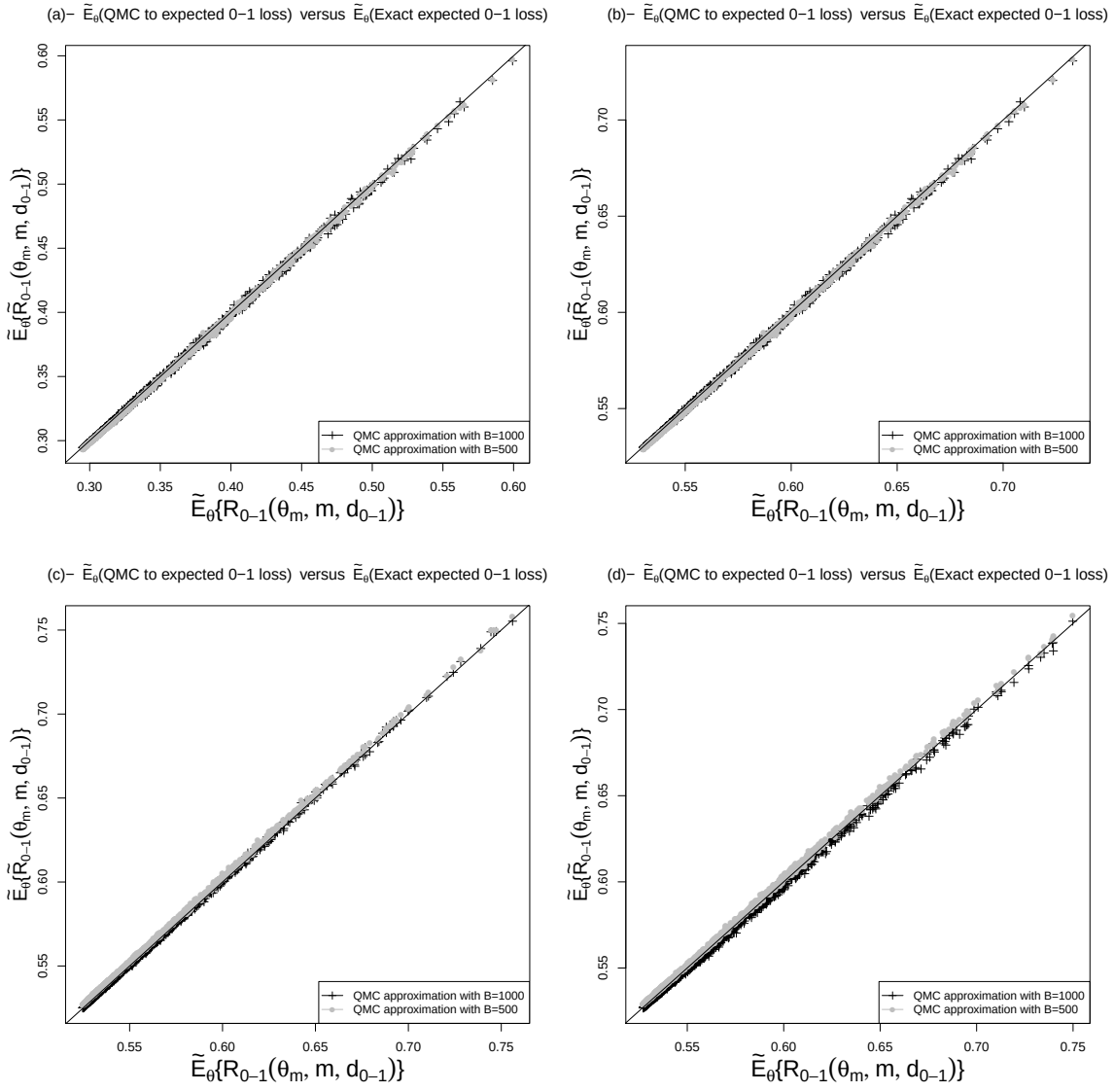

Figure 7.1: Plot of the Quad approximation to the prior expectation of QMC (with $B = 1,000$ and $B = 500$) approximation to the expected $0-1$ loss, plotted against the Quad approximation to the prior expectation of the exact expected $0-1$ loss for four different $\pi(m)$ under two factors logistic regression ($N = 4$).

7.5.3 Discussion

From Table 7.2 and Table 7.3, for all model probabilities, we can see that $d_{QMC_{L1}}$ and $d_{QMC_{L2}}$ have similar performance as d_T. It is also noted that as expected, the RT for $d_{QMC_{L1}}$ approximately doubles that for $d_{QMC_{L2}}$, which indicates that the CPU time required for finding the former is twice that required for the latter. For small values of N, the RT for both $d_{QMC_{L1}}$ and $d_{QMC_{L2}}$ indicates that these designs are computationally more expensive than d_T, whereas as the size of N increases the RTs decrease, indicating that d_T becomes computationally more expensive for large N. These results are similar to those found in Chapter 6 under

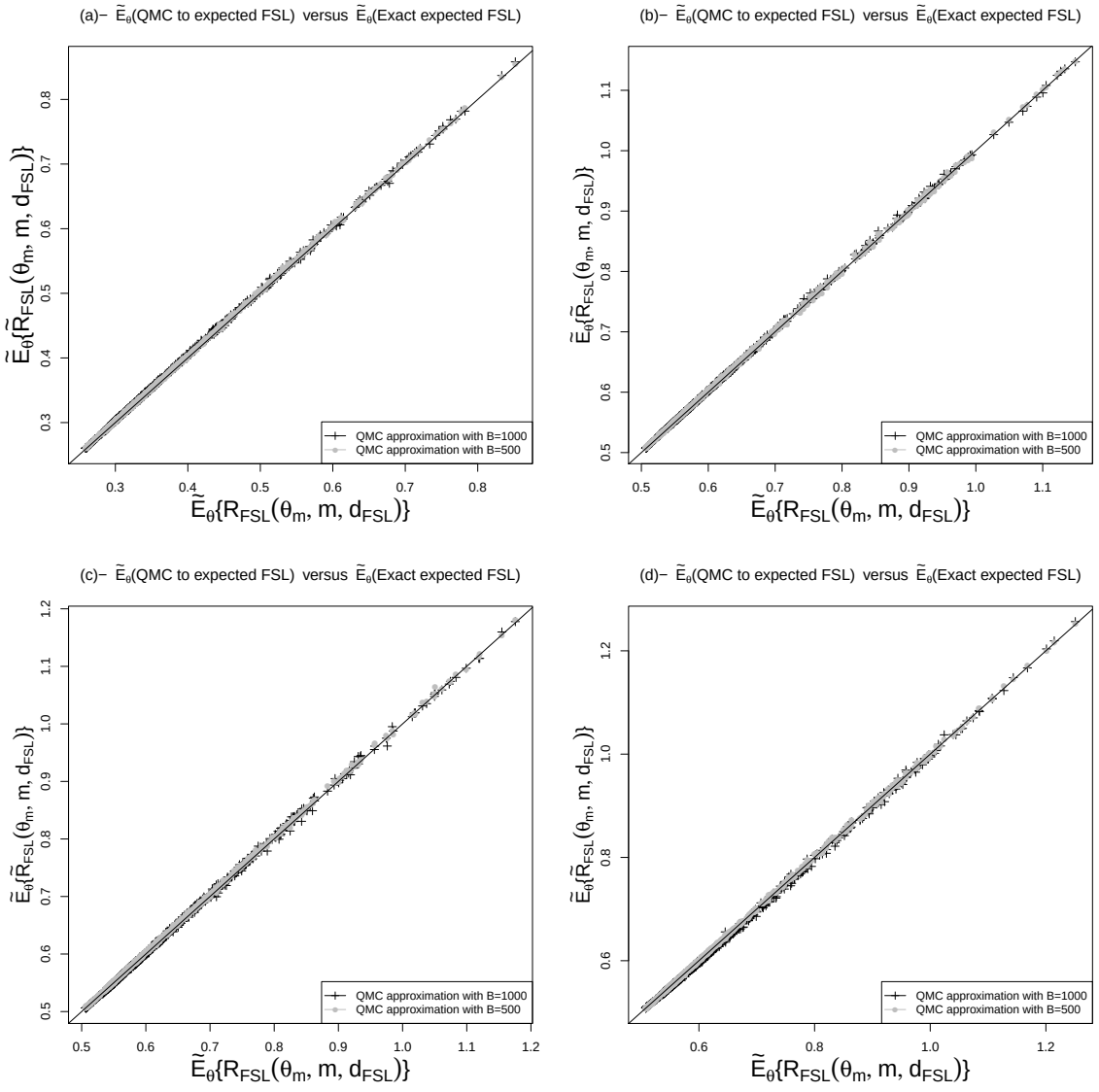

Figure 7.2: Plot of the Quad approximation to the prior expectation of the QMC (with $B = 1,000$ and $B = 500$) approximation to the expected FSL, plotted against the Quad approximation to the prior expectation of the exact expected FSL for four different $\pi(m)$ under two factors logistic regression ($N = 4$).

the experimental aim of parameter estimation.

According to Figure 7.1 and Figure 7.2, we can see that the approximate expectations of the QMC (with $B = 1,000$ and $B = 500$) appear very accurate in all four cases. As a result, the approximate expectation of QMC with $B = 500$ to the risk function can be used to find pseudo-Bayesian decision-theoretic designs for rather moderate-sized runs instead of the other approaches outlined in this chapter. This is due to the accuracy and the reasonable CPU time exhibited. It is predicted that compared with other designs, these designs require the smallest amount of time for a relatively large N.

7.6 Summary

In this chapter, we applied the Quad approximation to the Bayesian risk and QMC (with $B = 1,000$ and $B = 500$) to the risk function in order to find pseudo-Bayesian decision-theoretic designs for the $k = 2$ factor logistic regression model under the $0 - 1$ and FSL encapsulating the experimental aim of model discrimination. We found such designs in addition to the exact designs for rather moderate-sized runs, $4 \leq N \leq 10$. Then, the resulting designs were compared in terms of the performance and CPU time required for each design found for each N. It was found that the approximate expectation of the QMC (with $B = 500$) approximation to the risk can be used to find such optimal designs instead of other approaches, as their designs, $d_{QMC_{L2}}$, have performed very well for all values of experimental runs (in terms of closeness to the True designs d_T).

Chapter 8

Discussion and Future Work

8.1 Conclusion

In this thesis, we have introduced the decision-theoretic basis for frequentist designs applied for GLMs. It is an alternative approach to the standard approach in optimal experimental designs (alphabetical optimal designs), in which designs are found by minimising the risk function defined as the expectation of an appropriate loss function. Throughout the thesis, we have demonstrated the conceptual advantages of the decision-theoretic framework over alphabetical optimal designs in terms of its suitability for small sample sizes (it does not rely on large sample approximation to the variance matrix given by the inverse Fisher information matrix as the standard approach does) and its consideration of pre-stated experimental aims through the specification of loss functions.

The hurdles to finding such designs have been discussed and addressed in this research. The first problem was the analytical intractability of the risk function in addition to the high dimensionality exhibited. This was a challenging task in terms of finding exact frequentist decision-theoretic optimal designs; hence, methods for approximating the risk function were required. We have developed novel methods to approximate the risk function, including QMC, QIMPS and Quad. In addition, we have extended the ST approximation to the risk function proposed by (Pronzato and Pázman, 2013) for non-linear models to other types of models, including generalised linear models; see Chapter 4.

The second problem encountered was the dependance of the risk function on the unknown parameters we aim to estimate. We have applied the local and pseudo-Bayesian approaches to overcome this dependance on the parameters, where under the pseudo-Bayesian approach, the quadrature rule was applied to approximate the Bayesian risk. The third problem was the high dimensionality of the risk function

that we aim to minimise in order to find optimal designs, and this was addressed by the application of the methodology of the ACE algorithm for finding designs.

In this thesis, the developed approximation was investigated for a range of examples involving GLMs, and they have important applications in the design of experiments. Theses were the logistic regression and two factors Poisson regression models. Under logistic regression model, we considered three special cases of this model, which were simple logistic regression and the four and two factors logistic regression models. The first two examples have been considered in addition to the Poisson example for finding designs under the squared error and sign losses that are tailored for the experimental aim of parameter estimation; see Chapter 6. For the simple logistic regression model, we used the grid search method in addition to the modified quasi-Newton method to find decision-theoretic optimal designs.

We have extended our investigation to finding designs with the aim of estimating the unknown parameter to the experimental aim of model discrimination, and this has been investigated for the two factor logistic regression model under 0-1 and factor-specific losses; see Chapter 7.

The proposed approximation methods in this research have been compared with an asymptotic (Asym) approximation and the exact risk (where possible). The comparison between these approximations was conducted in terms of performance and computing time. Note that the objective functions that have been minimised under the pseudo-Bayesian approach have been coded with a low level programming language and used the packages **Rcpp** and **RcppArmadillo**. The algorithm was run on IRIDIS 4, the super computer facility at the University of Southampton, which has 2.6 GHz processors with 4GB memory.

Our findings across the examples investigated in this research showed that the QMC approximation is the most accurate approximation to the risk function, and can be used to find near-exact decision-theoretic optimal designs for various values of experimental runs. However, some approximations to the risk function, such as QIMPS, ST and Quad, exhibited inaccuracies for small values of N and required a considerable sample size to produce designs that are nearly exact the optimal ones. Furthermore, the Quad approximation to the risk required substantially more computing time (CPU time) and memory as the number of runs increased, and this was the case for finding the optimal design under the exact risk (where possible) in some examples under the pseudo-Bayesian approach.

One limitation of this research is that the objective function we aim to minimise in order to find the optimal design requires computationally intensive methods to obtain accurate results. However, this limitation has less importance with the use of high-performance computing technologies that continue to change rapidly.

8.2 Future work

8.2.1 Minimax decision-theoretic designs

This research could be extended to consider the minimax approach for finding decision-theoretic optimal designs. It is a powerful design selection criterion used to address the dependancy of the risk function on the values of the model parameters. It is a useful approach when we wish to avoid the worst-case scenario (i.e. the values of $\boldsymbol{\theta}$ that result in the maximum value of the risk function), where the maximum risk over the design space is minimised. Mathematically, the minimax decision-theoretic optimal design minimises

$$\Psi(d) = \max_{\boldsymbol{\theta} \subset \boldsymbol{\Theta}} R(\boldsymbol{\theta}; d). \tag{8.1}$$

When the parameter space $\boldsymbol{\Theta}$ is a singleton set, the resulting design is the same as the locally decision-theoretic optimal design (Chen et al., 2013).

Finding minimax decision-theoretic optimal designs is a substantially challenging task in terms of computational expense, as it requires much more computing time than that required for finding the pseudo-Bayesian optimal designs found in this research. This is because every time we need to evaluate the objective function given by (8.1) in a minimisation algorithm, we need to maximise the objective function first. To illustrate, if the minimisation of the objective function takes A evaluations and the maximisation of the objective function takes B evaluations, then finding the minimax design requires $A \times B$ evaluations of the objective function in (8.1). Hence, finding minimax decision-theoretic optimal designs in the case of a high dimensional design space is difficult. We need to develop a new algorithm that should be amenable to a high dimensional design space.

8.2.2 Other classes of models

We could extend this research to finding decision-theoretic optimal designs for other classes of models and use the approximation methods developed in this thesis. For example, we might consider non-linear models

$$Y = \eta(\mathbf{z}, \boldsymbol{\theta}) + \epsilon,$$

where ϵ is independent and identically distributed (i.i.d) and assumed to be normally distributed $N(0, \sigma^2)$. A compartmental model is an example of a non-linear model, which is the most commonly studied in the field of optimal experimental designs (Atkinson et al., 1993). It consists of $k = 1$ factor and $p = 3$ unknown parameters, and can be expressed as:

$$\eta(z_i, \boldsymbol{\theta}) = \theta_3 \left[\exp(-\theta_1 z_i) - \exp(-\theta_2 z_i)\right].$$

The model describes the transport of materials in an organism during a particular time z_i (e.g. $z_i \in [0, 24]$ hours), and the term $\eta(z_i, \boldsymbol{\theta})$ denotes the expected response. The response measures the concentration of the materials in the organism at that time z_i, where at $z_i = 0$ the expected response is zero. As the time z_i increases, the expected response increases up to the maximum, and then start gradually decreasing to zero as z_i gets larger. The parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)^T$ are all positive, where $\theta_2 > \theta_1$. The parameters θ_1 and θ_2 describe the flow of materials between compartments, while θ_3 is the initial concentration of the materials before they are absorbed into the organism.

Furthermore, in order of increasing complexity, we might extend this research to allow non-normally distributed error structure by considering a generalised non-linear model (GNM). It differs from GLMs in that it includes non-linear parameters. The measured response under GNMs consists of three components:

- *Distribution of the response*: it is a random component and from the exponential family of distributions.
- *Systematic predictor* $\eta(\mathbf{z}; \boldsymbol{\theta})$: it forms the systematic part of the model and involves functions of the k controllable factors which be non-linear in the p model parameters $\boldsymbol{\theta}$.
- *Link function*: the mean response $E(Y_i) = \mu_i$ is related to $\eta(\mathbf{z}_i; \boldsymbol{\theta})$ through the link function $g(.)$, i.e
$$g(\mu_i) = \eta(\mathbf{z}_i; \boldsymbol{\theta}).$$

An example of GNMs with application in second-harmonic generation experiments might be considered. It was studied by Biedermann and Woods (2011), for finding Bayesian D-optimal designs for the experimental aim of parameter estimation.

8.2.3 Methods to approximate the Bayesian risk

In this research, we used the quadrature rule to approximate the Bayesian risk for finding pseudo-Bayesian decision-theoretic optimal designs. For finding such designs, we could use the QMC (with $B = 500$) approximation to the Bayesian

risk, i.e

$$E_{\boldsymbol{\theta}}\left\{R(d;\boldsymbol{\theta})\right\} \approx \frac{1}{B}\sum_{b=1}^{B}\widetilde{R}(d;\boldsymbol{\theta}_b). \tag{8.2}$$

It would then sample $\{\boldsymbol{\theta}_b\}_{b=1}^{B}$ from the prior distribution of the parameters $\pi(\boldsymbol{\theta})$. The term $\widetilde{R}(d;\boldsymbol{\theta}_b)$ in (8.2) is approximated using different approximations that were developed in this research. Then, we will compare the resulting designs found under these approaches in terms of performance and computing time.

Appendix A

A.1 Introduction to Bayesian decision-theoretic designs

The optimal design helps to improve the statistical inference about the quantities of interest $\boldsymbol{\theta}$, and this can be made by appropriately selecting the values of the controllable factors, resulting in small variability of the estimator of interest. Specific information is typically available prior to experimentation about $\boldsymbol{\theta}$, and that motivates the experiment. This also enhances the role of Bayesian methods in the field of experimental design that have been widely studied and popular in that field. Bayesian decision-theoretic optimal designs accommodate the aim of the experiment, and this substantially enhances the role of Bayesian decision-theoretic optimal designs.

Before giving a detailed description of the Bayesian decision-theoretic framework and the essential concepts involved under this framework, basic definitions regarding Bayesian framework are provided. Bayesian methods are based on Bayes' theorem

$$\begin{aligned} f(\boldsymbol{\theta}|\mathbf{Y}, d) &= \frac{f(\mathbf{Y}|\boldsymbol{\theta}, d)\pi(\boldsymbol{\theta})}{f(\mathbf{Y})}, \\ &= \frac{f(\mathbf{Y}|\boldsymbol{\theta}, d)\pi(\boldsymbol{\theta})}{\int_{\Theta} f(\mathbf{Y}|\boldsymbol{\theta}, d)\pi(\boldsymbol{\theta})d\boldsymbol{\theta}}, \end{aligned} \tag{A.1}$$

where $\mathbf{Y}$ and $\boldsymbol{\theta}$ are considered to be random variables. The integration in the denominator of (A.1) is replaced by a summation in the case of a discrate-valued of $\boldsymbol{\theta}$. Since the denominator of (A.1) does not depend on $\boldsymbol{\theta}$, Bayes' theorem can be rewritten as

$$f(\boldsymbol{\theta}|\mathbf{Y}, d) \propto f(\mathbf{Y}|\boldsymbol{\theta}, d)\pi(\boldsymbol{\theta}).$$

The quantity $f(\mathbf{Y}|\boldsymbol{\theta}, d)$ is the joint pdf of $\mathbf{Y}$ and equivalent to the likelihood function $L(\boldsymbol{\theta}|\mathbf{Y}, d)$, while the quantity $\pi(\boldsymbol{\theta})$ is the pdf of the prior distribution of $\boldsymbol{\theta}$ that reflects our knowledge of $\boldsymbol{\theta}$ prior to observing the data from the experiment. The quantity $f(\boldsymbol{\theta}|\mathbf{Y}, d)$ is the pdf of the posterior distribution of $\boldsymbol{\theta}$ given the observed data of $\mathbf{Y}$, and reflects our updated knowledge of $\boldsymbol{\theta}$ by combining the

information from the data with the information from the prior distribution using Bayes' theorem.

A.2 Bayesian decision-theoretic framework

The Bayesian decision-theoretic optimal design is found by minimising the expectation of the loss function (risk function), over a design space $\mathcal{X}$

$$\begin{aligned} R(d;\boldsymbol{\theta}) &= E_{\boldsymbol{\theta},\mathbf{Y}|d}\left[\lambda(\hat{\boldsymbol{\theta}}(\mathbf{Y}),\boldsymbol{\theta},d)\right] && \text{(A.2)}\\ &= \int \lambda(\hat{\boldsymbol{\theta}}(\mathbf{Y}),\boldsymbol{\theta},d) f(\boldsymbol{\theta},\mathbf{Y}) d\boldsymbol{\theta} d\mathbf{Y} && \text{(A.3)}\\ &= \int \lambda(\hat{\boldsymbol{\theta}}(\mathbf{Y}),\boldsymbol{\theta},d) f(\boldsymbol{\theta}|\mathbf{Y},d) f(\mathbf{Y}) d\boldsymbol{\theta} d\mathbf{Y} && \text{(A.4)}\\ &= \int \lambda(\hat{\boldsymbol{\theta}}(\mathbf{Y}),\boldsymbol{\theta},d) f(\mathbf{Y}|\boldsymbol{\theta},d) \pi(\boldsymbol{\theta}) d\boldsymbol{\theta} d\mathbf{Y}, && \text{(A.5)} \end{aligned}$$

where $R(d;\boldsymbol{\theta})$ denotes the risk function, and the expectation in (A.3) is with respect to the joint distribution of $\mathbf{Y}$ and quantities $\boldsymbol{\theta}$. The equivalence of (A.4) and (A.5) comes from the application of Bayes theorem. The latter is more useful for computational purposes (Overstall et al., 2016), as it is straightforward in simulation. The risk function usually results from high-dimensional integrals and analytically intractable equations, leading to the need of approximation methods. The standard approximation method is Monte Carlo integration rule, where the integral in (A.3) is approximated by a weighted sum

$$R(d;\boldsymbol{\theta}) \approx \sum_{b=1}^{B} w_b \lambda(\hat{\boldsymbol{\theta}}(\mathbf{Y}_b),\boldsymbol{\theta}_b,d), \qquad \text{(A.6)}$$

with $w_b > 0$. It would then sample $\{\boldsymbol{\theta}_b,\mathbf{Y}_b\}_{b=1}^{B}$ from the joint distribution $f(\boldsymbol{\theta},\mathbf{Y};d)$, and set $w_b = 1/B$. It is a stochastic technique and easy to implement, but it is computationally expensive.

A.3 Computational methods for constructing optimal designs

The construction of optimal designs is an optimisation problem, and the selected criterion defines the objective function. Many computer algorithms have been developed to solve this problem. The most widely-used technique for numerically searching for exact optimal designs are exchange algorithms (Nguyen and Miller, 1992). There are several approaches of exchange algorithms for constructing a discrete D-optimal design, such as Fedorov's exchange algorithm and its modifications, sequential algorithms, non-sequential algorithms and the KL and BLKL exchange algorithms (Atkinson et al., 2007). Exchange algorithms can be classified into two main classes: point exchange algorithms (Fedorov, 1972; Johnson and

Nachtsheim, 1983) and coordinate exchange algorithms (Meyer and Nachtsheim, 1995). The former is performed by a systematic exchange of design points with points from a candidate set of all possible design points to improve the value of the objective function. The performance of the design after these exchanges is measured by the objective function, where exchanges that result in improvement are accepted, otherwise are rejected. In practice, the number of candidate points might be quite large. The drawback of using this technique is that the size of problems that can be solved by such approaches is bounded. This is because of that the computational intensity required grows exponentially with an increase of experimental factors, requiring a large candidate set.

The latter is performed by changing one element z_{ij}, or "coordinate", of a design d at a time with no need for a candidate set. The "coordinate" is the setting of an individual factor in a single run. Meyer and Nachtsheim (1995) described a cyclic coordinate exchange algorithm for constructing exact optimal design under D- and linear optimality, using a variant of the Gauss-Southwell cyclic coordinate-descent algorithm within the k-exchange algorithm. The approach gives a substantive reduction in computing time. However, we will describe approximate coordinate exchange algorithms (ACE) (Overstall and Woods, 2017) which is the most recent computer algorithm for constructing an exact optimal design, and tailored to the decision-theoretic framework.

A.3.1 Approximate coordinate exchange algorithm

The approximate coordinate exchange algorithm (ACE) is a recent approach for finding optimal designs for a variety of statistical models with different numbers of controllable factors, numbers of experimental runs and randomisation constraints. Overstall and Woods (2017) introduce this methodology in the area of Bayesian designs, and demonstrate it for different practical statistical models that have important practical problems such as problems in chemical and biological sciences. Bayesian design for many of these problems are not available in the literature. The ACE algorithm is used for finding optimal designs in the decision-theoretic framework, when we are able to simulate the values of the model parameters from the prior distribution of the parameters, and responses from the statistical model.

In order to present this approach, we recall some notation. The design d has q-dimensional space, where $q = Nk$. Each element of d is a coordinate z_{ij}, where $i = 1, \ldots, N$ and $j = 1, \ldots, k$. The ACE algorithm incorporates the coordinate exchange and point exchange algorithms, and it is abbreviated to ACE. This algorithm is employed in two phases. In phase I, the ACE algorithm is applied by setting the initial design d_0 as the current design d^C. For each coordinate (z_{ij}),

(I) m points are generated from the design space $\mathcal{X}$, $\varphi = \{z_{ij}^1, \ldots, z_{ij}^m\} \in \mathcal{X}$, using one dimensional computer experiment (e.g., space-filling design is most popular, see Lin and Tang (2015) for a recent review), and (II) each point is replaced with the current coordinate, and the other design coordinate are taken as fixed. This results in m designs for each coordinate. Following this, (III) evaluation of each design with the new coordinate is done by the objective function, i.e. the approximate expectation of the approximated risk function $\Psi(d)$. Then, (IV) a Gaussian process model is fitted to $\Psi(d)$ and minimised in order to propose a new value for the coordinate. The design that achieves improvement is accepted to be the current design with probability obtained from a Bayesian test of equality of the Monte Carlo approximation, for example, to $\Psi(d)$ for the original design and the design with the new suggested coordinate as

$$P_I^* = 1 - t_{2B-2}\left(-\frac{B\Psi(d^C) - B\Psi(d^{C*})}{\sqrt{2B\hat{v}_I}}\right), \tag{A.7}$$

where t_{2B-2} corresponds to the probability distribution function for a t- distribution with $2B-2$ degrees of freedom, and

$$\hat{v}_I = \frac{\sum_{s=1}^{B}\left(\lambda(\hat{\boldsymbol{\theta}}(\mathbf{Y}_s^*), \boldsymbol{\theta}_s^*, d^{C*}) - \Psi(d^{C*})\right)^2 + \sum_{s=1}^{B}\left(\lambda(\hat{\boldsymbol{\theta}}(\mathbf{Y}_s^C), \boldsymbol{\theta}_s^C, d^C) - \Psi(d^C)\right)^2}{2B-2}$$

The terms $\{\mathbf{Y}_s^*, \boldsymbol{\theta}_s^*\}_{s=1}^B$ and $\{\mathbf{Y}_s^C, \boldsymbol{\theta}_s^C\}_{s=1}^B$ are independent random samples generated from $f(\mathbf{Y}, \boldsymbol{\theta}; d^{C*})$ and $f(\mathbf{Y}, \boldsymbol{\theta}; d^C)$, respectively. This is in the case of a stochastic approximation to $\Psi(d)$. Otherwise, a deterministic approximation is performed, the design that achieves improvement sets to the current design, if $\Psi(d)$ evaluated at the design with the new coordinate is smaller than $\Psi(d)$ evaluated at the current design, $\Psi(d^{C*}) < \Psi(d^C)$. The optimal design is then found by repeating N_1 times this sequence of computer experiments.

In phase II, the design that is resulted from phase I is refined, where a cluster of similar points in a design produced from phase I is consolidated using the point exchange algorithm which requires a candidate set, formed from the design arising from phase I. First, the design point is added to the current design whose replication minimises $\Psi(d)$ and that resulted in $n+1$ design points. Then, the new design is found by deleting the point whose deletion results in lowest $\Psi(d)$, where the new design has n points with these two points swapped. Similar to the phase I, if a stochastic approximation is performed to $\Psi(d)$, the new design is accepted with probability obtained from a Bayesian test, otherwise (a deterministic approximation is applied) the new design is set to be the current if $\Psi(d^{C*}) < \Psi(d^C)$. In order to find the optimal design, these steps are iterated N_2 times. In the next section, the ACE algorithm is detailed, and it is also implemented in the `R` package `acebayes` (Overstall et al., 2017a).

A.4 The ACE algorithm

- Phase I algorithm:
  1. Choose an initial design $d_0 \in \mathcal{X}$, and let the current design d^C being equal to d_0, where $d^C = d_0$.
  2. For $i = 1, \ldots, N$ and $j = 1, \ldots, k$, carry out the following steps:
     (a) Let $d^C(z_{ij}^q)$ being equal to d^C with the ijth coordinate replaced by z_{ij}^q, where $z_{ij}^1, \ldots, z_{ij}^m$ are the points from a one-dimensional space filling design in $\mathcal{X}_{ij}$, the design space for the ijth element of d.
     (b) For $q = 1, \ldots, m$, evaluate $\Psi(d)$ at $d^C(z_{ij}^q)$, where $\Psi[d^C(z_{ij}^q)]$. Then, fit a Gaussian process emulator to $\left\{z_{ij}^q, \Psi[d^C(z_{ij}^q)]\right\}_{q=1}^m$, and set $\hat{\Psi}_{ij}(z)$ to be the resulting predictive mean.
     (c) Find
     $$z_{ij}^* = \arg\min_{z \in \mathcal{X}_{ij}} \hat{\Psi}_{ij}(z),$$
     and set $d^{C*} = d^C(z_{ij}^*)$.
     (d) For a stochastic approximation to $\Psi(d)$(e.g Monte Carlo), set $d^C = d^{C*}$ with probability p^* derived from a Bayesian hypothesis test, while for a deterministic approximation such as quadrature approximation, set $d^C = d^{C*}$ if $\Psi(d^{C*}) < \Psi(d^C)$.
  3. Iterate step 2 N_1 times.
- Phase II algorithm:
  1. Let the current design d^C being equal to the final design from phase I of the ACE algorithm.
  2. For $i = 1, \ldots, N$, set
     $$d_i^{(1)} = [(d^C)^T, (\boldsymbol{z}_i^C)^T]^T$$
     , where $(\boldsymbol{z}_i^C)^T$ is the ith row of d^C that form $d_i^{(1)}$ by augmenting d^C with a repeat of the ith run.
  3. Find the point $i^* = \arg\min_{i=1,\ldots,N} \Psi(d_i^{(1)})$ and set $d^{(2)} = d_{i^*}^{(1)}$.
  4. For $h = 1, \ldots, N+1$, set
     $$d_h^{(3)} = \left[(\boldsymbol{z}_1^{(2)})^T, \ldots, (\boldsymbol{z}_{h-1}^{(2)})^T, (\boldsymbol{z}_{h+1}^{(2)})^T, \ldots, (\boldsymbol{z}_{N+1}^{(2)})^T\right]^T,$$
     where $(\boldsymbol{z}_h^{(2)})^T$ is the hth row of $d^{(2)}$ that form $d_h^{(3)}$ by removing the hth run.
  5. Find the point $h^* = \arg\min_{h=1,\ldots,N+1} \Psi(d_h^{(3)})$ and let $d^{C*} = d_{h^*}^{(3)}$
  6. For a stochastic approximation to $\Psi(d)$, set $d^C = d^{C*}$ with probability p^* obtained from a Bayesian hypothesis test, while for a deterministic

approximation (e.g quadrature approximation), set $d^C = d^{C*}$ if $\Psi(d^{C*}) < \Psi(d^C)$.

7. Iterate steps 2 to 6 N_2 times.

For both phases, the decision of terminating the algorithm depends on the choice of N_1 and N_2; this choice varies based on the nature of the approximation to the objective function $\Psi(d)$. In the `R` package `acebayes`, the default of these N_1 and N_2 set to 20 and 100, respectively. This algorithm is similar to other searching algorithms for optimal designs in that it converges to the local optima (Overstall et al., 2016). In order to prevent this convergence, it is suggested to run the algorithm T times, where each run starting from a randomly different chosen initial design d_0 (Atkinson et al., 2007, Chapter 12). Hence, the optimal design is the design having the lowest expected risk $\Psi(d)$.

Appendix B

B.1 An investigation to the normality of the parameter estimators for four factors logistic regression

B.1.1 Four factors logistic regression model

Figure B.1 and Figure B.2 show the p-values of Shapiro-Wilk test for multivariate normality of $\hat{\boldsymbol{\theta}}(\mathbf{Y})$ and $\hat{\boldsymbol{\theta}}^*(\mathbf{Y})$, respectively, plotted against the experimental runs N, where $1000 \leq N \leq 10000$. It can be seen that the parameter estimates of $\hat{\boldsymbol{\theta}}(\mathbf{Y})$ and $\hat{\boldsymbol{\theta}}^*(\mathbf{Y})$ are not normally distributed for small experiments and require enormous experimental runs to be asymptotically normal.

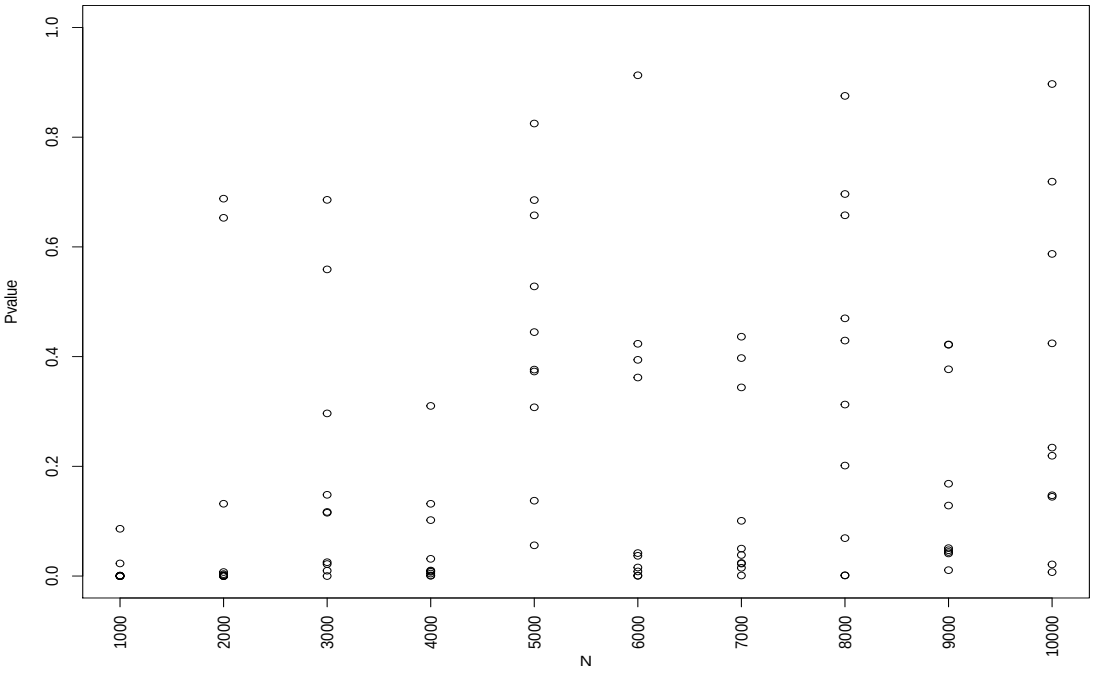

Figure B.1: Distribution of the P-values for a test of H_0: the MLEs are normally distributed for different sample sizes under four factors logistic regression model.

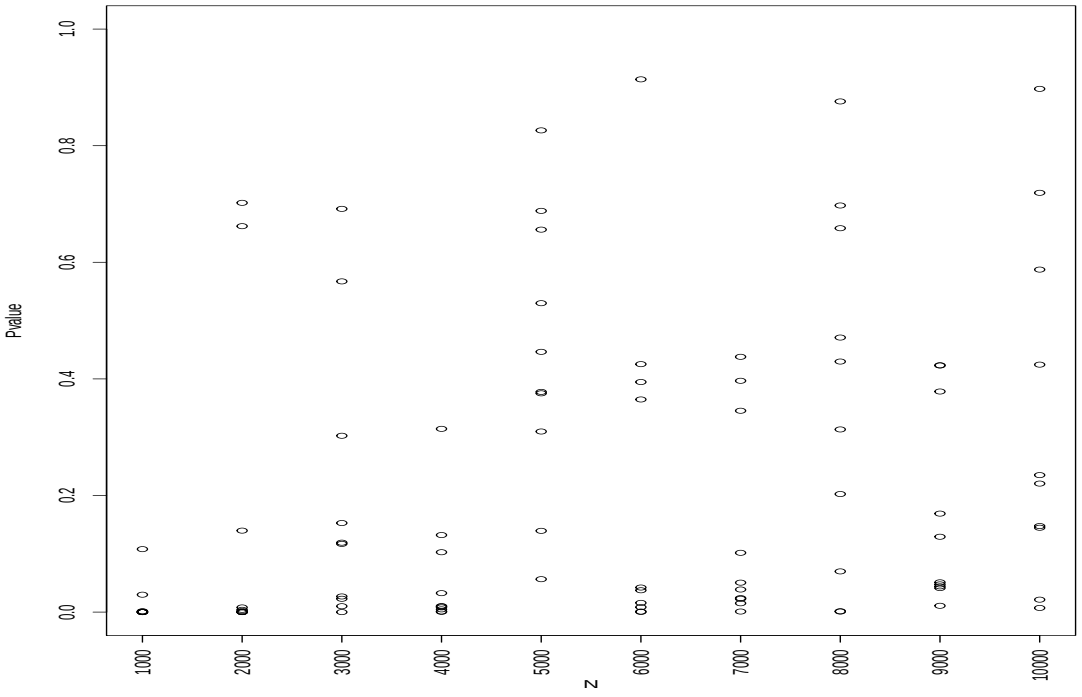

Figure B.2: Distribution of the P-values for a test of H_0: the MPLEs are normally distributed for different sample sizes under four factors logistic regression model.

Appendix C

C.1 Locally decision-theoretic D-optimal designs for simple logistic regression model

C.1.1 Investigating the estimates of ρ_0 and θ_1 under simple linear logistic regression model

The increase in the exact $\log|R_{SE}(\boldsymbol{\theta}, d)|$ for small n is investigated. It is found that this increase occurs when the responses $\mathbf{Y} = (0, 0)$ or $\mathbf{Y} = (n, n)$, where under these responses the slope parameter θ_1 is always zero. We calculate the estimates of ρ_0 and θ_1 under these responses, and plot them against a range of size n.

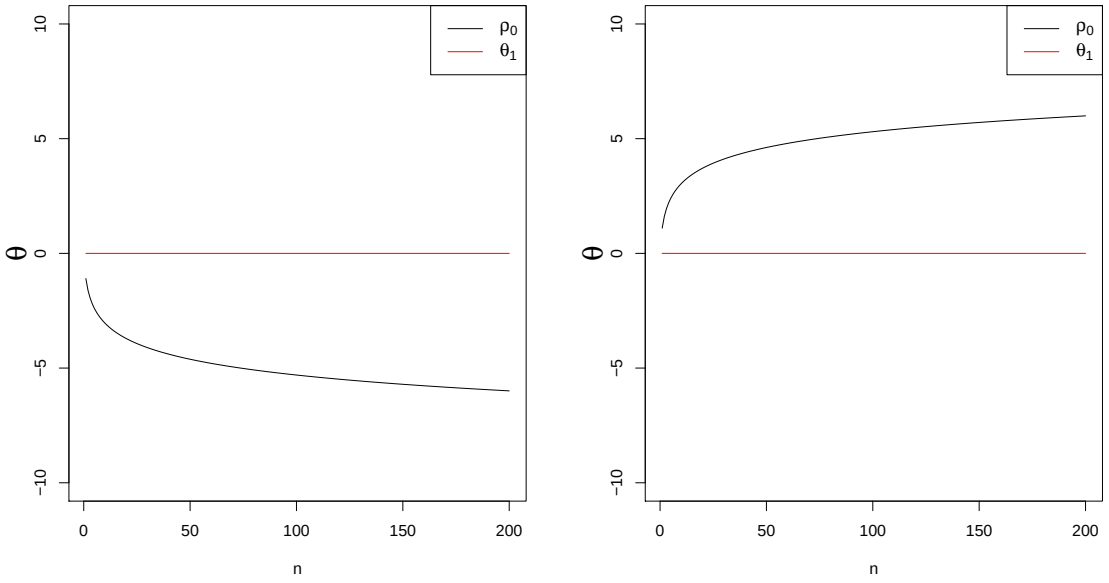

Figure C.1: Plots of the estimates of ρ_0 and θ_1 found under the exact locally decision-theoretic D-optimal design when $\mathbf{Y} = (0, 0)$ (left) and $\mathbf{Y} = (n, n)$ (right).

C.2 Locally and pseudo-Bayesian decision-theoretic A-optimal designs

C.2.1 Locally decision-theoretic A-optimal designs

We use the QMC 2 approximation to $\mathrm{tr}\{R_{SE}(\boldsymbol{\theta}, d)\}$ in addition to those used under the decision-theoretic D-optimality in order to find the approximate locally decision-theoretic A-optimal designs. We refer to the design that found under the QMC 2 approximation to $\mathrm{tr}\{R_{SE}(\boldsymbol{\theta}, d)\}$ as QMC 2 design. We find exact and approximate locally decision-theoretic A-optimal designs that minimise $\mathrm{tr}\{R_{SE}(\boldsymbol{\theta}, d)\}$ and $\mathrm{tr}\left\{\widetilde{R}_{SE}(\boldsymbol{\theta}, d)\right\}$, respectively, for various sizes of n. The first row of Figure C.2 shows plot of the exact $\mathrm{tr}\{R_{SE}(\boldsymbol{\theta}, d)\}$ that was evaluated at the resulting designs, and plotted against n. We can see that locally A-optimal design d_{A_*} found under the Asym approximation to $\mathrm{tr}\{R_{SE}(\boldsymbol{\theta}, d)\}$ is far from True design d_T, but it improves as n gets larger. It is also noted that Quad, QMC and QMC 2 designs are near True design unlike ST design, where the exact $\mathrm{tr}\{R_{SE}(\boldsymbol{\theta}, d)\}$ evaluated at each of these designs are very similar to $\mathrm{tr}\{R_{SE}(\boldsymbol{\theta}, d_T)\}$ for all sizes of n. This indicates the accuracies of their corresponding approximations. In Figure C.3, the relative A-efficiency is plotted against n. It is noted that A-optimal design is less efficient for small n compared with other designs. However, all designs found under this example have the same efficiency for large n.

The steep increase at $n = 15$ is investigated by plotting the contour plots of all approximations considered to $\mathrm{tr}\{R_{SE}(\boldsymbol{\theta}, d)\}$ in addition to the exact $\mathrm{tr}\{R_{SE}(\boldsymbol{\theta}, d)\}$ in Figure C.5. In this figure, six contour plots for six different approaches of finding locally decision-theoretic A-optimal designs are plotted. These are a contour plot of the exact $\mathrm{tr}\{R_{SE}(\boldsymbol{\theta}, d)\}$ (the first row on left), a contour plot of the approximated $\mathrm{tr}\left\{\widetilde{R}_{SE}(\boldsymbol{\theta}, d)\right\}$ via ST (the first row on right), Asym (the second row on left), Quad (the second row on right), QMC with size $B = 5000$ (the third row on left) and QMC 2 approximation with size $B = 5000$ (the third row on right). We can see that ST design is far from the True design. This is not the case for $n = 5$ and 100 , where d_{ST} design is close to d_T at these sizes of n (see Figure C.6 and C.7, respectively). This sudden change might cause the sharp increase of the exact $\mathrm{tr}\{R_{SE}(\boldsymbol{\theta}, d)\}$ when it is evaluated at d_{ST}.

The validity of the approximations needs to be checked. Figure C.8 shows plots of different approximations to the MSE, plotted against the exact MSE. We can see that all the approximations to the MSE appear very accurate.

Here, we do similar to what has been accomplished under locally decision-theoretic D- optimal designs for justifying the increase in the exact MSE. Figure C.9 shows

plots of $\mathrm{tr}\{R_{SE}(\boldsymbol{\theta}, d)\}$ (green line), $R_{SE}(\rho_0, d)$ (red line) and $R_{SE}(\theta_1, d)$ (black line) that were evaluated at d_{ST} and plotted against n. It is concluded that the increase occurred for small n in the exact $\mathrm{tr}\{R_{SE}(\boldsymbol{\theta}, d_{ST})\}$ is dominated by $R_{SE}(\rho_0, d_{ST})$, where the $\mathrm{tr}\{R_{SE}(\boldsymbol{\theta}, d)\}$ shows similar pattern to $R_{SE}(\rho_0, d_{ST})$ and different from $R_{SE}(\theta_1, d_{ST})$. The increase of $\mathrm{tr}\{R_{SE}(\boldsymbol{\theta}, d)\}$ is found as the responses $\mathbf{Y} = (0, 0)$ or $\mathbf{Y} = (n, n)$ that resulted in zero values for the estimates of θ_1 for all n. The estimates of ρ_0 and θ_1 are calculated for both $\mathbf{Y} = (0, 0)$ and $\mathbf{Y} = (n, n)$ for a range of n, and plotted against n (see Figure C.10). As a result, the unacceptable increase of $\mathrm{tr}\{R_{SE}(\boldsymbol{\theta}, d_{ST})\}$ is caused by the estimates of ρ_0 and θ_1 under the above responses.

The RT, under the trace of MSE, is calculated, where the QMC approximation is the most expensive approximation to the MSE, followed by QMC 2 and Quad, with $RT = 110.87, 38.98$ and 1.97, respectively. However, the ST (RT=0.42) and locally D-optimal designs (RT=0.11) found under the ST and asymptotic approximations to the MSE, respectively, require less executing time than the true design d_T.

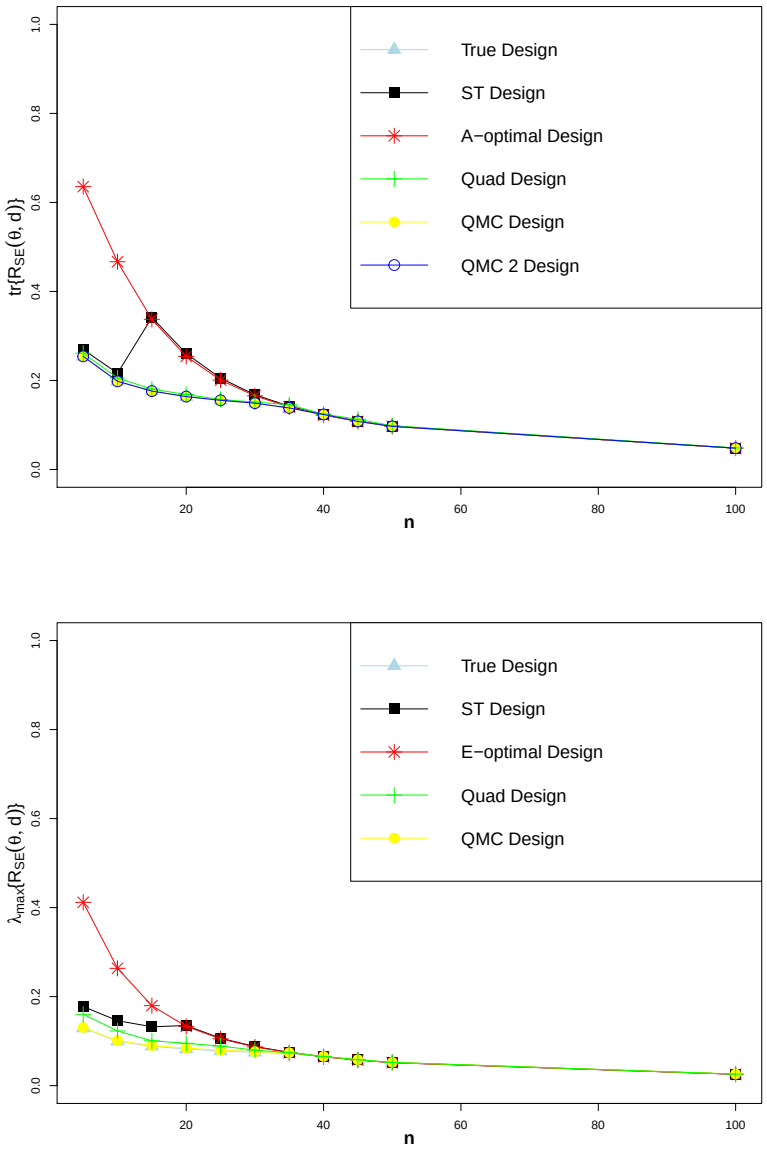

Figure C.2: Plots of the exact $\mathrm{tr}\{R_{SE}(\boldsymbol{\theta}, d)\}$ (the first row) and $\lambda_{max}\{R_{SE}(\boldsymbol{\theta}, d)\}$ (the second row) evaluated at the True and approximated designs, and plotted against different sizes of n for simple logistic regression model.

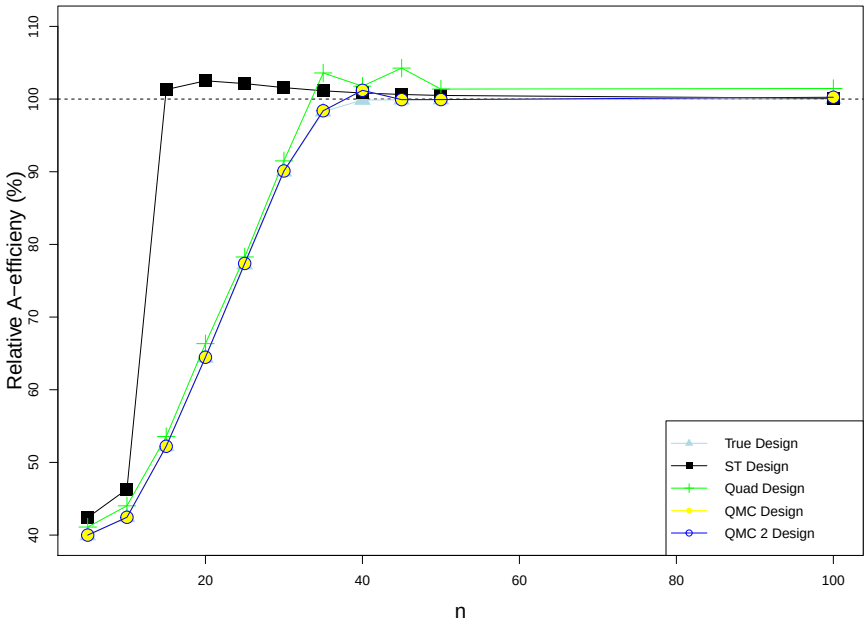

Figure C.3: Plots of the relative A-efficiency for the True and approximated designs, and plotted against different sizes of n for simple logistic regression model.

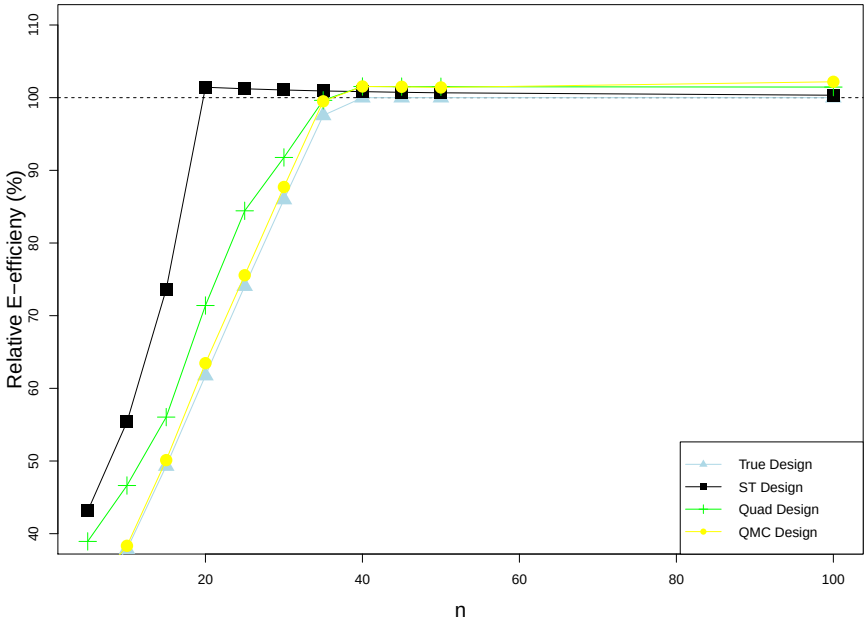

Figure C.4: Plots of the relative E-efficiency for the True and approximated designs, and plotted against different sizes of n for simple logistic regression model.

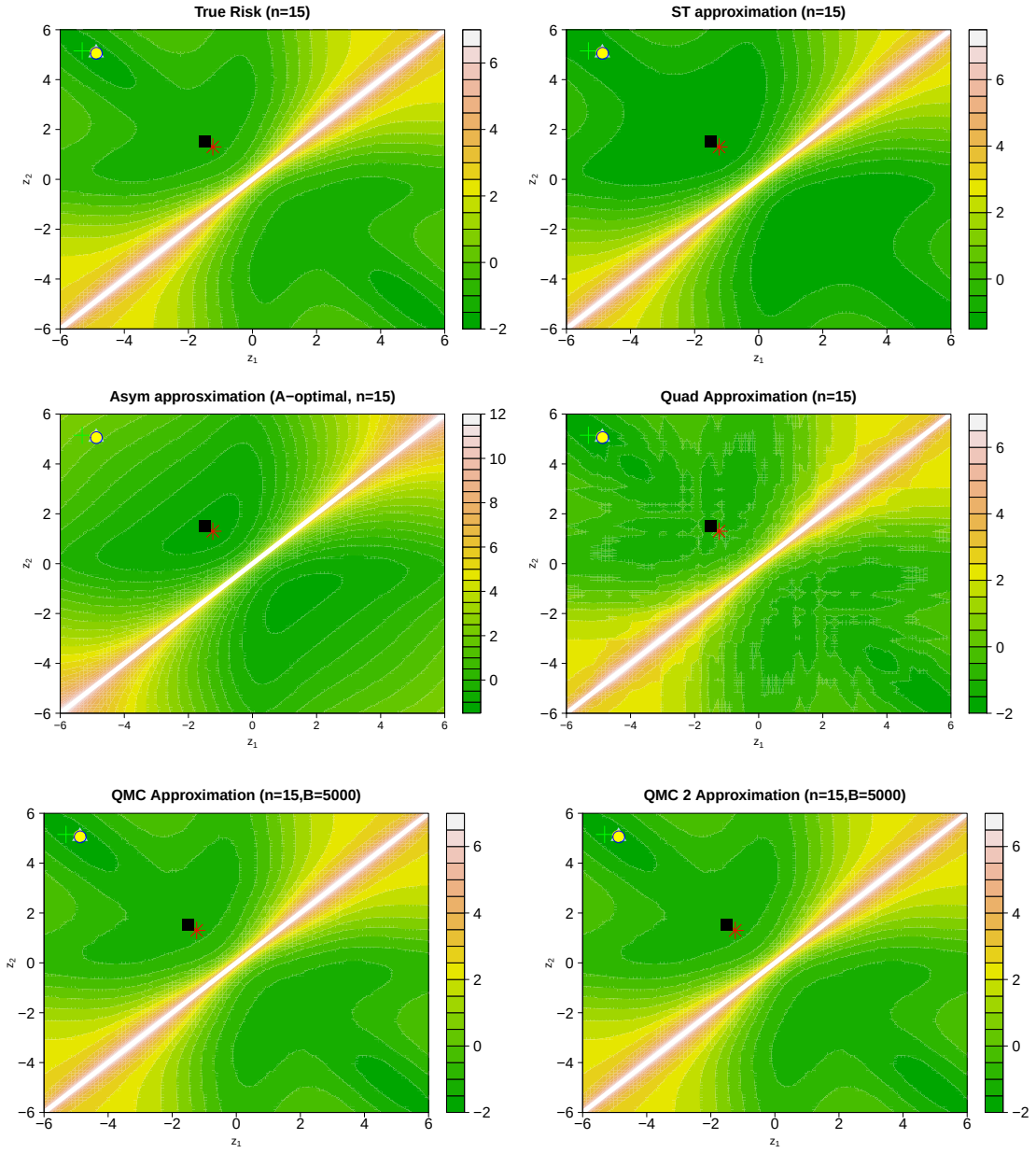

Figure C.5: Contour plots of tr $\{R_{SE}(\boldsymbol{\theta}, d)\}$ (the first row (left)) and the trace of the approximated mean squared error tr $\left\{\widetilde{R}_{SE}(\boldsymbol{\theta}, d)\right\}$ via ST (the first row (right)), Asym (the second row (left)), Quad (the second row(right)), QMC with size $B = 5,000$ (the third row (left)) and QMC 2 with size $B = 5,000$ (the third row (right)), under simple logistic regression model for a sample of size $n = 15$. A plot key for each contour is plotted which refers to the values of tr $\{R_{SE}(\boldsymbol{\theta}, d)\}$ and tr $\left\{\widetilde{R}_{SE}(\boldsymbol{\theta}, d)\right\}$ for the remaining contour plots. In Each plot, there are sex coloured-shapes denote the decision-theoretic A-optimal designs found under the true risk (light blue triangle), ST(black square), Asym (red star), Quad (green plus), QMC (yellow filled circle) and QMC 2 approximation (blue open circle).

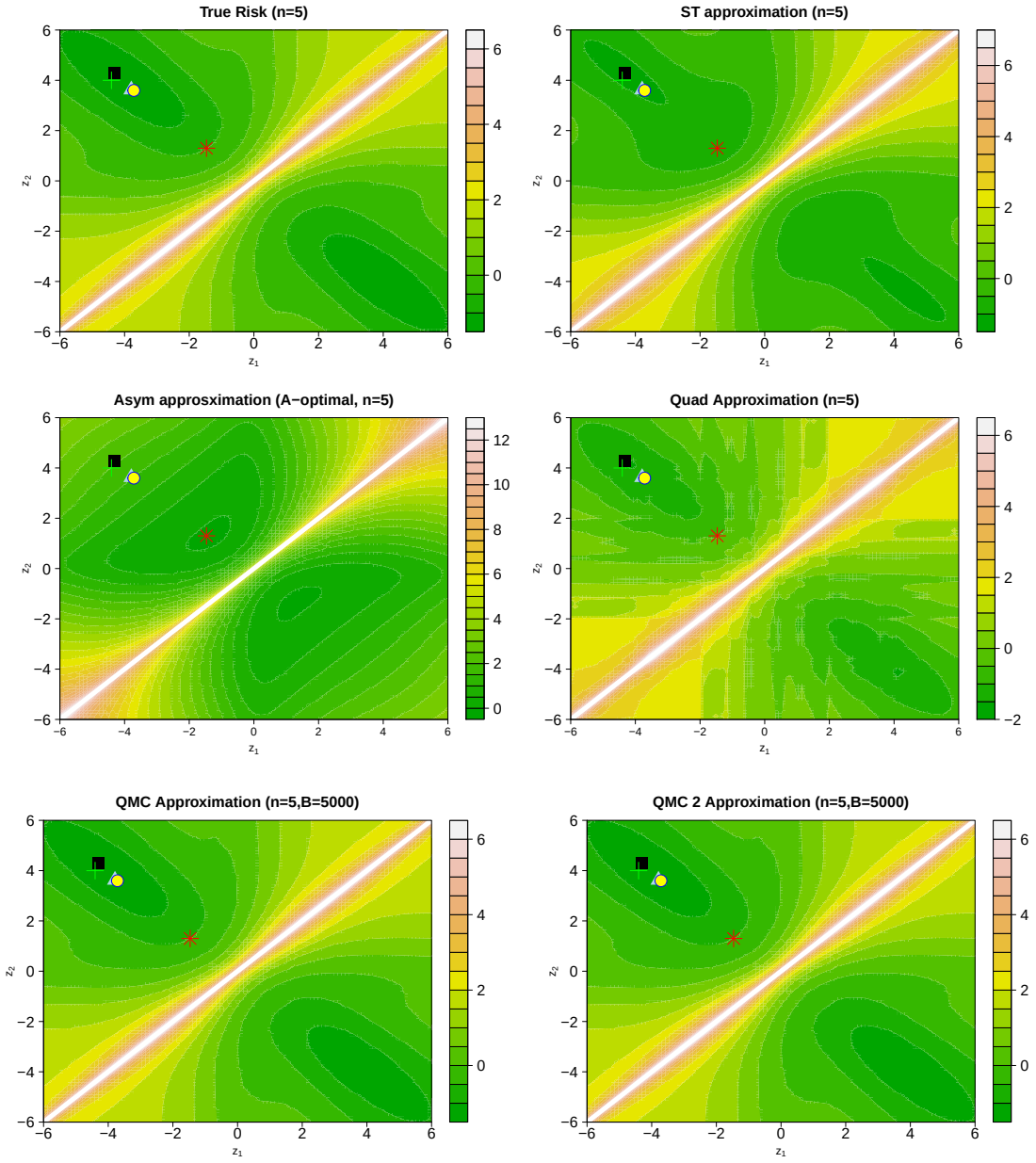

Figure C.6: Contour plots of the exact $\mathrm{tr}\{R_{SE}(\boldsymbol{\theta}, d)\}$ (the first row (left)) and approximated MSE $\mathrm{tr}\left\{\widetilde{R}_{SE}(\boldsymbol{\theta}, d)\right\}$ via ST (the first row (right)), Asym (the second row (left)), Quad (the second row(right)), QMC with size $B = 5,000$ (the third row (left)) and QMC 2 with size $B = 5,000$ (the third row (right), under simple logistic regression model for a sample of size $n = 5$. A plot key for each contour is plotted which refers to the values of $\mathrm{tr}\{R_{SE}(\boldsymbol{\theta}, d)\}$ and $\mathrm{tr}\left\{\widetilde{R}_{SE}(\boldsymbol{\theta}, d)\right\}$ for the remaining contour plots. In Each plot, there are six coloured-shapes denote the decision-theoretic A-optimal designs found under the exact risk (light blue triangle), ST(black square), Asym (red star), Quad (green plus), QMC (yellow filled circle) and QMC 2 approximation (blue open circle).

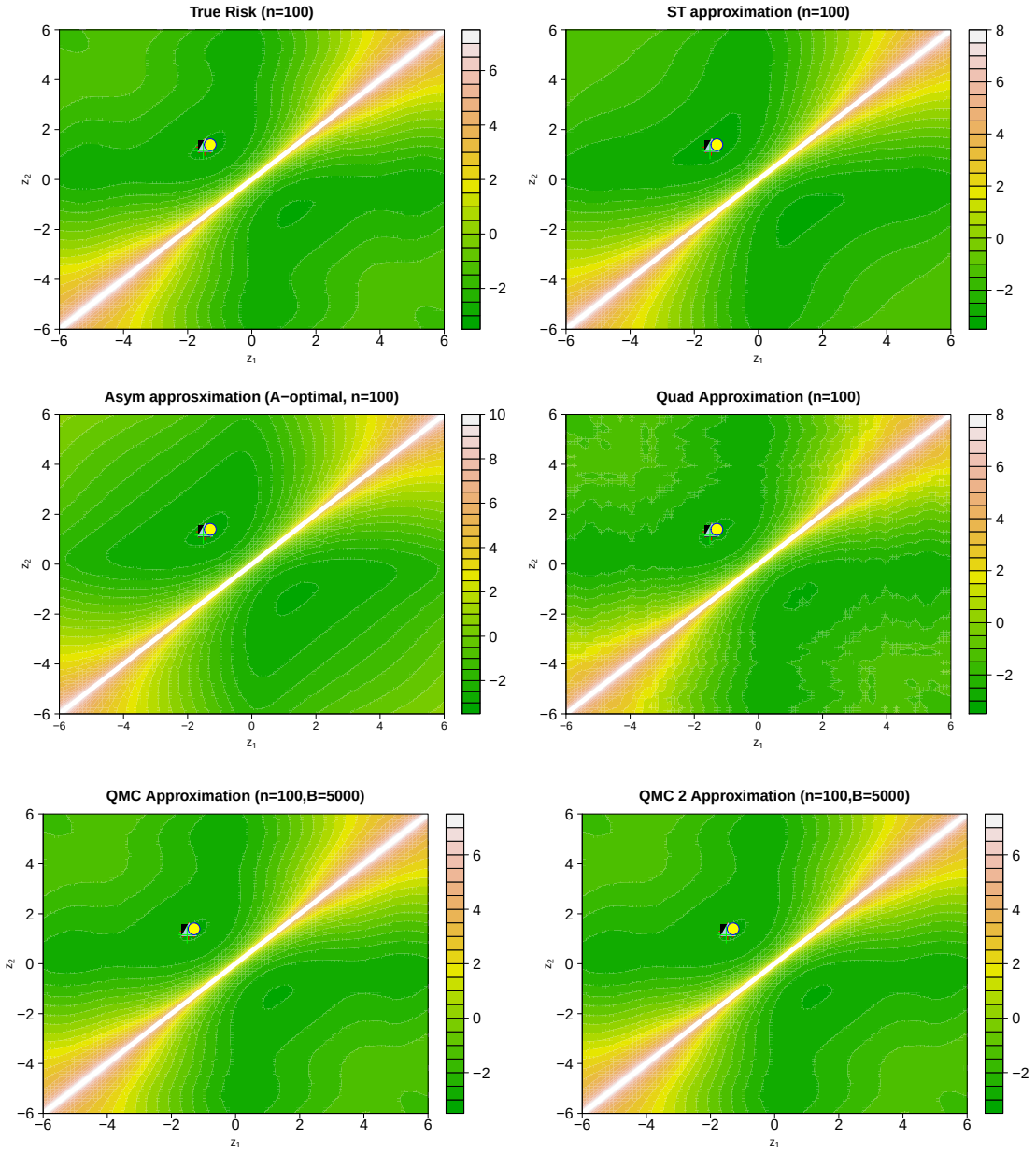

Figure C.7: Contour plots of the exact $\mathrm{tr}\{R_{SE}(\boldsymbol{\theta}, d)\}$ (the first row (left)) and approximated MSE $\mathrm{tr}\left\{\widetilde{R}_{SE}(\boldsymbol{\theta}, d)\right\}$ via ST (the first row (right)), Asym (the second row (left)), Quad (the second row(right)), QMC with size $B = 5,000$ (the third row (left)) and QMC 2 with size $B = 5,000$ (the third row (right), under simple logistic regression model for a sample of size $n = 100$. A plot key for each contour is plotted which refers to the values of $\mathrm{tr}\{R_{SE}(\boldsymbol{\theta}, d)\}$ and $\mathrm{tr}\left\{\widetilde{R}_{SE}(\boldsymbol{\theta}, d)\right\}$ for the remaining contour plots. In Each plot, there are six coloured-shapes denote the decision-theoretic A-optimal designs found under the exact risk (light blue triangle), ST(black square), Asym (red star), Quad (green plus), QMC (yellow filled circle) and QMC 2 approximation (blue open circle).

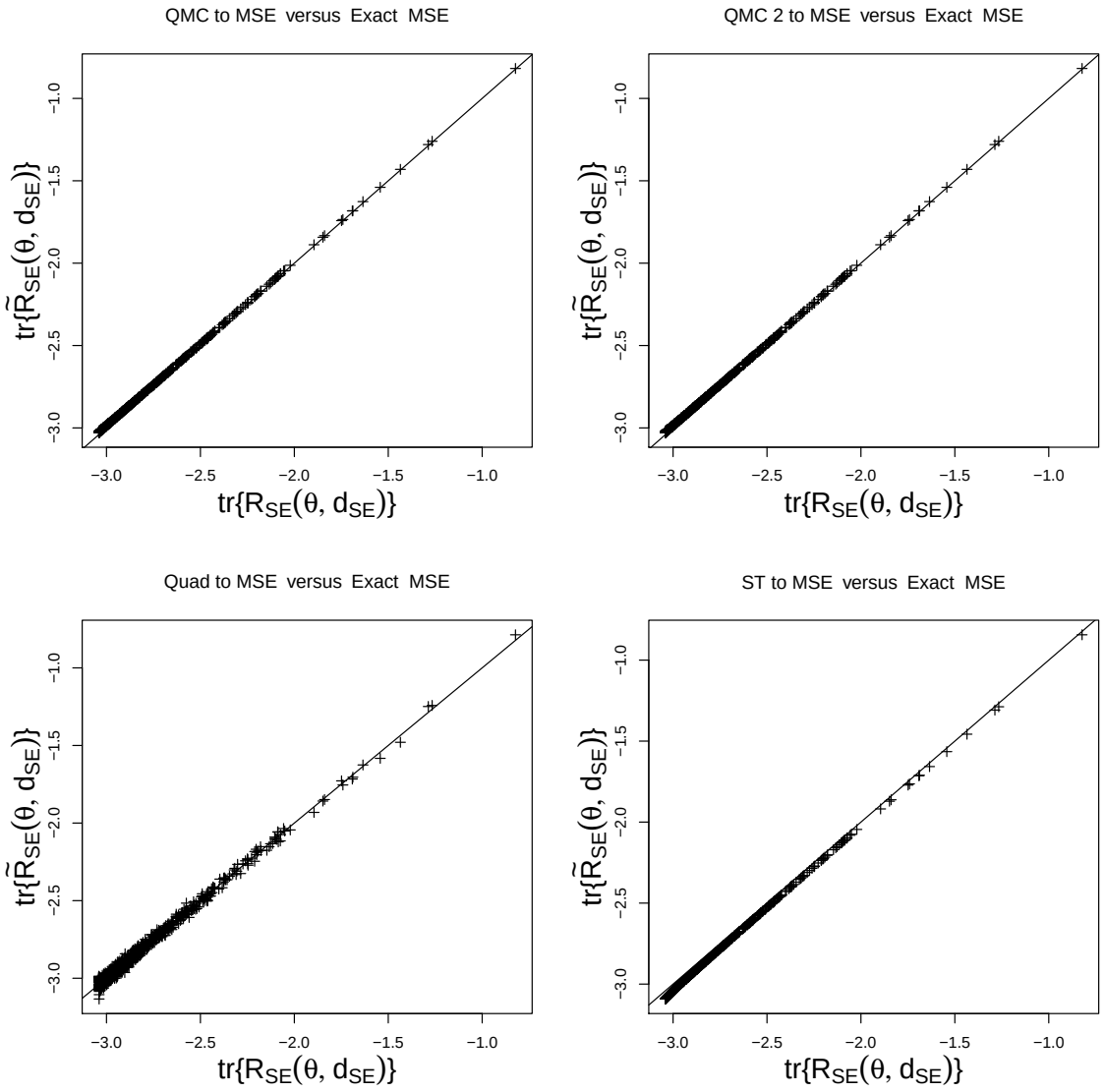

Figure C.8: Plots of approximations to the MSE $\operatorname{tr}\left\{\widetilde{R}_{SE}(\boldsymbol{\theta}, d)\right\}$ plotted against the exact MSE $\operatorname{tr}\{R_{SE}(\boldsymbol{\theta}, d)\}$ for simple logistic regression for $n = 100$.

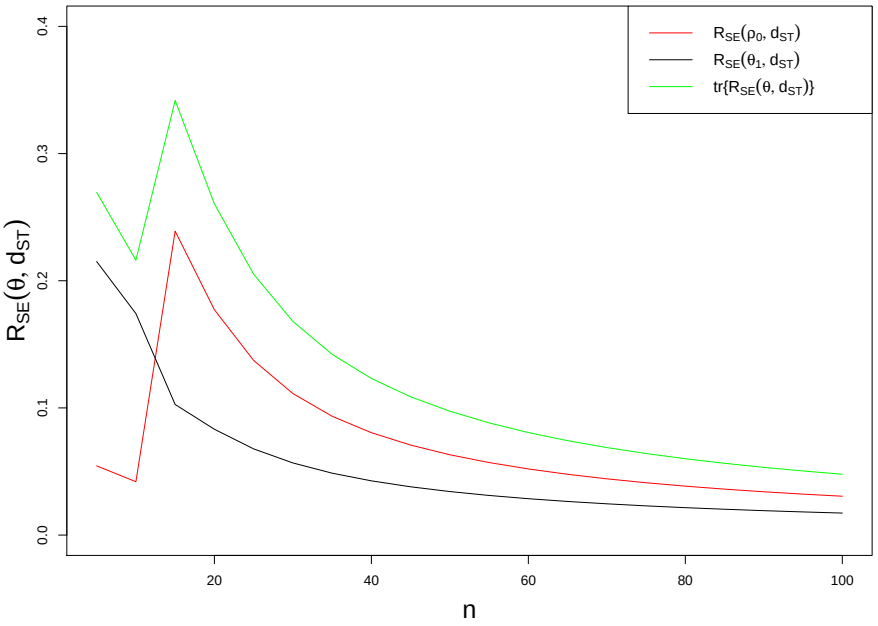

Figure C.9: Plots of the MSE for $R_{SE}(\rho_0, d_{ST})$, $R_{SE}(\theta_1, d_{ST})$ and the trace of $R_{SE}(\boldsymbol{\theta}, d_{ST})$ evaluated at d_{ST} found under the trace of the MSE.

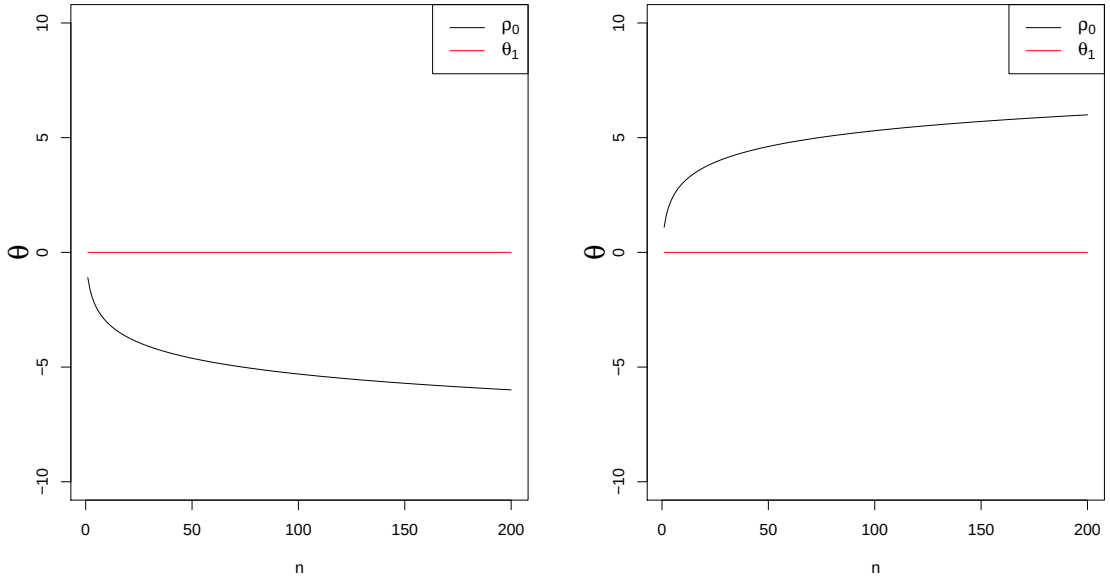

Figure C.10: Plots of the estimates of ρ_0 and θ_1 found under the approximate locally decision-theoretic A-optimal designs via ST when $\mathbf{Y} = (0, 0)$ (left) and $\mathbf{Y} = (n, n)$ (right).

C.2.2 Pseudo-Bayesian decision-theoretic A-optimal designs

We find exact and approximate pseudo-Bayesian decision-theoretic A-optimal designs by minimising $\widetilde{E}_{\boldsymbol{\theta}}\left\{\text{tr}\left\{R_{SE}(\boldsymbol{\theta}, d)\right\}\right\}$ and $\widetilde{E}_{\boldsymbol{\theta}}\left\{\text{tr}\left\{\widetilde{R}_{SE}(\boldsymbol{\theta}, d)\right\}\right\}$, respectively. In the first row of Figure C.11, we evaluate $\widetilde{E}_{\boldsymbol{\theta}}\left\{\text{tr}\left\{R_{SE}(\boldsymbol{\theta}, d)\right\}\right\}$ at the resulting designs under $\pi_1(\boldsymbol{\theta})$ (left) and $\pi_2(\boldsymbol{\theta})$ (right), for various values of n. We can see that all the approximate expectations of the used approximations to $\text{tr}\left\{R_{SE}(\boldsymbol{\theta}, d)\right\}$ produce designs near optimal design (close to True design) and have good performance (in terms of the approximate prior expectation of $\text{tr}\left\{R_{SE}(\boldsymbol{\theta}, d)\right\}$) even for small n. It also appears that we could obtain a negligible difference in $\widetilde{E}_{\boldsymbol{\theta}}\left\{\text{tr}\left\{R_{SE}(\boldsymbol{\theta}, d)\right\}\right\}$ for values of n over approximately thirty-five by using the pseudo-Bayesian A-optimal design.

In the Figure C.12, we investigate the performance of optimal designs under $\pi_2(\theta)$ if $\pi_1(\theta)$ is correct (a & c), and the performance of optimal designs under $\pi_1(\theta)$ if $\pi_2(\theta)$ is correct (b & d). We can see that the optimal designs are sensitive to the chosen prior, as their performance are different than what we observe in Figure C.12.

In Figure C.13, we check the validity of the approximate expectations of the used approximations to the MSE under $\pi_1(\boldsymbol{\theta})$ and $\pi_2(\boldsymbol{\theta})$, and we can see that all the approximations appear very accurate, except the approximate expectations of QIMPS and ST approximations to the MSE. However, based on the ordering of the designs in terms of the approximate expectation of the MSE for QIMPS and ST approximations, these approximations are minimised for designs close to the design that minimises the approximate expectation of the exact MSE.

The RT for each of these approximations is calculated, and found that the RT of the approximate expectation of the QMC approximation to the MSE is 32.78. This means that this approximation is approximately 33 times more expensive to be computed than the approximate expectation of the exact MSE. The approximate expectation of the QIMPS is the second expensive approximation with RT=29.36, followed by RT=23.20 for QMC 2. The approximate expectations of other approximations to the MSE require less CPU time than that for the exact MSE.

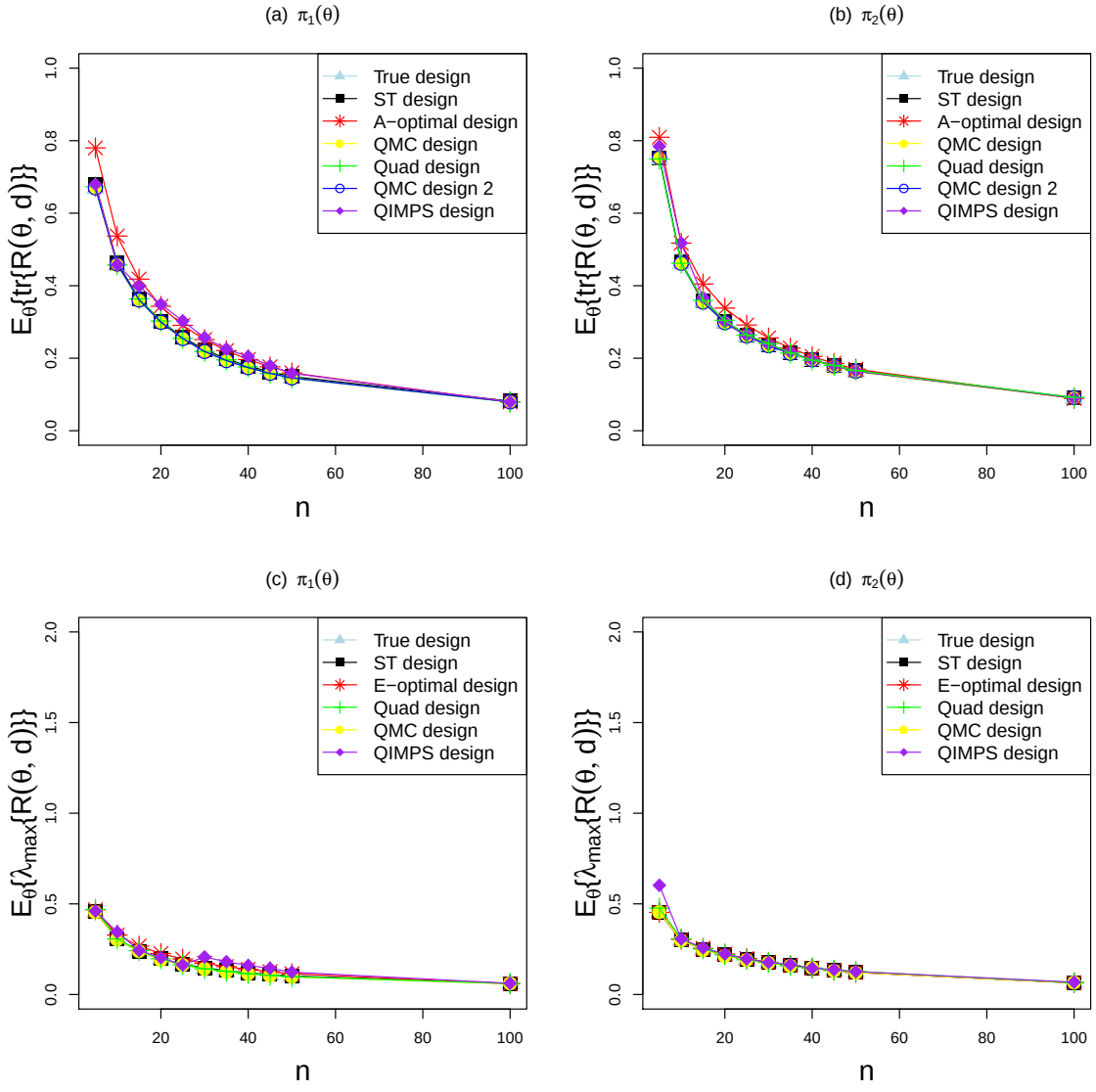

Figure C.11: Plots of the two approximated prior expectation ($\pi_1(\boldsymbol{\theta})$ (a) and (c), $\pi_2(\boldsymbol{\theta})$ (b) and (d)) of the exact $\mathrm{tr}\{R_{SE}(\boldsymbol{\theta}, d)\}$ (the first row) and $\lambda_{max}\{R_{SE}(\boldsymbol{\theta}, d)\}$ (the second row) evaluated at the resulting designs, and plotted against different size of binomial trials n for the simple logistic regression model.

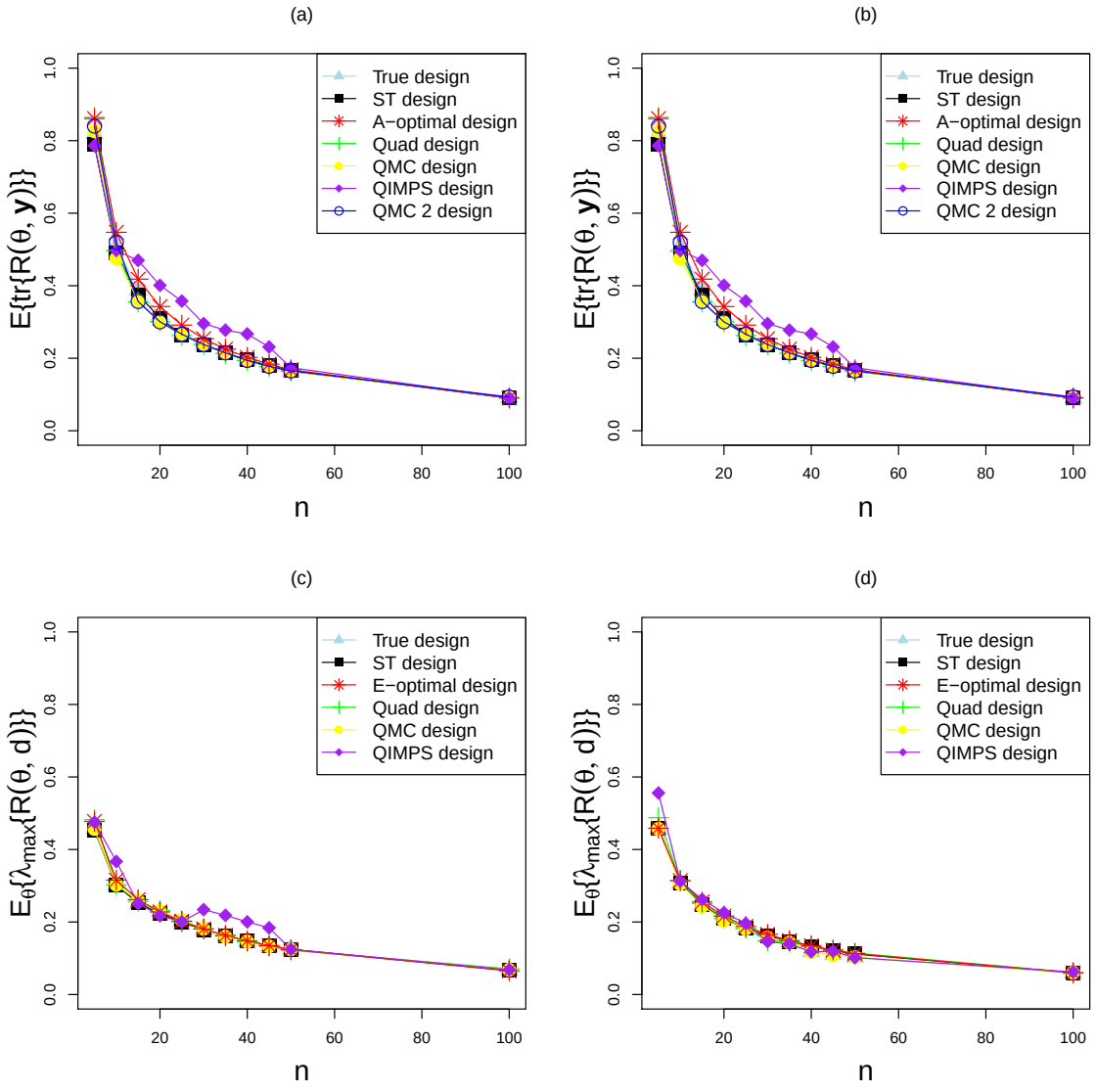

Figure C.12: Plots of the exact $\mathrm{tr}\{R_{SE}(\boldsymbol{\theta}, d)\}$ (the first row) and $\lambda_{max}\{R_{SE}(\boldsymbol{\theta}, d)\}$ (the second row) evaluated at optimal designs. These plots illustrates the misspecification of the prior distributions, where the performance of the optimal designs found under prior 1 are investigated under prior 2 (a & c) and vice versa (b & d).

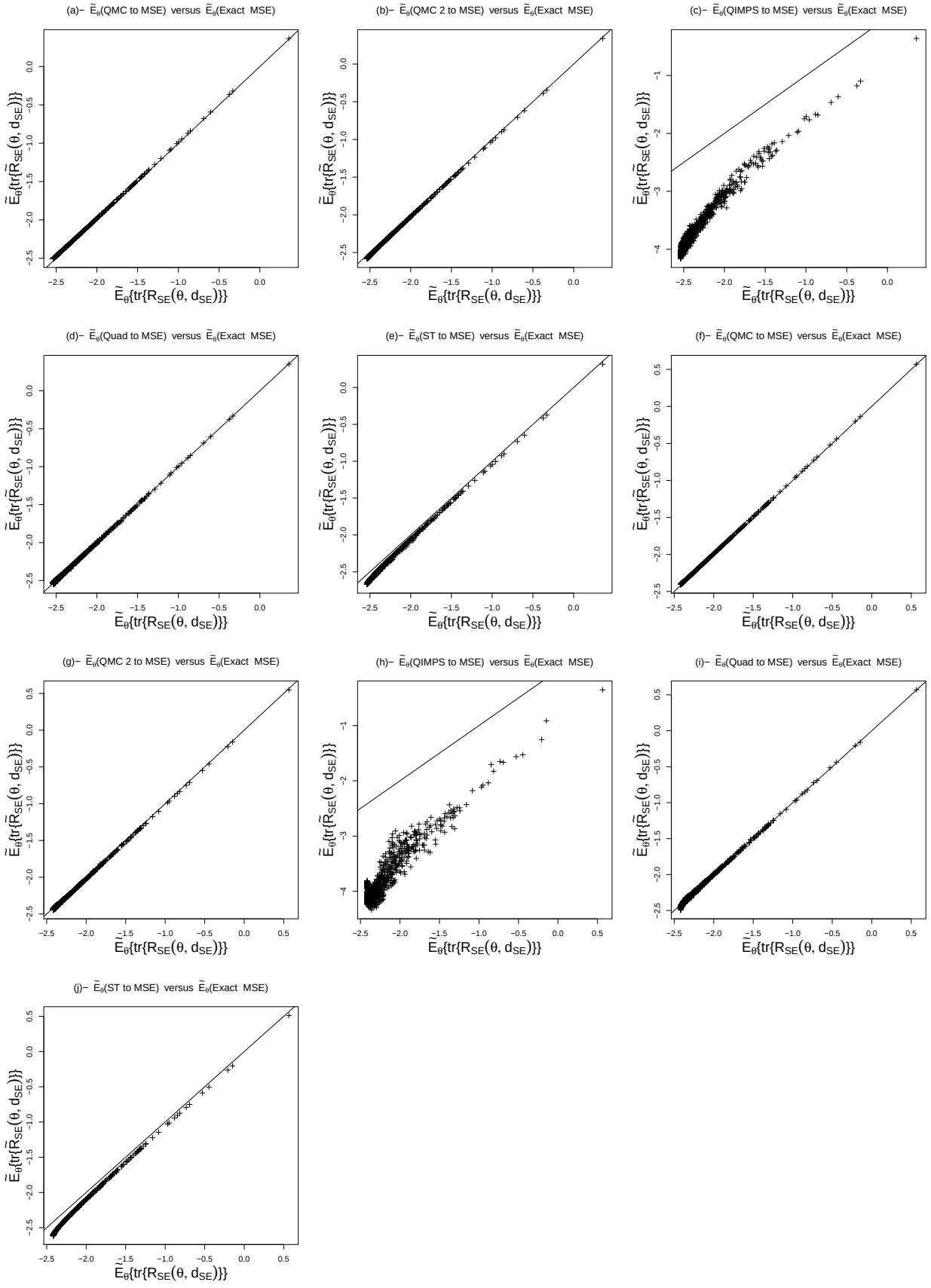

Figure C.13: Plots of the approximate expectation of approximations to the MSE, plotted against the approximate expectation of the exact MSE for the simple logistic regression with $n = 100$ under $\pi_1(\boldsymbol{\theta})$ and $\pi_2(\boldsymbol{\theta})$.

C.3 Locally and pseudo-Bayesian decision-theoretic E-optimal designs

C.3.1 Locally decision-theoretic E-optimal designs

Exact and approximate locally decision-theoretic E-optimal designs are found for different sizes of n by minimising $\lambda_{max}\{R_{SE}(\boldsymbol{\theta}, d)\}$ and $\lambda_{max}\left\{\widetilde{R}_{SE}(\boldsymbol{\theta}, d)\right\}$, respectively. We plot the exact $\lambda_{max}\{R_{SE}(\boldsymbol{\theta}, d)\}$ that was evaluated at the resulting designs against n in the second row of Figure C.2. We can see that QMC design is close to True design for all sizes of n, indicating the accuracy of the QMC approximation. However, it can be seen that the ST, Asym and Quad approximations show inaccurate results for small n since the exact $\lambda_{max}\{R_{SE}(\boldsymbol{\theta}, d)\}$ evaluated at the corresponding designs are far from the exact $\lambda_{max}\{R_{SE}(\boldsymbol{\theta}, d_T)\}$. Hence, poor designs under these approaches are produced. As the size of n increases, these approximations attain improvement where this can be seen for large n. It is noted that a slight increase occurs in the exact $\lambda_{max}\{R_{SE}(\boldsymbol{\theta}, d)\}$ evaluated at d_{ST} for $n = 20$, so this increase needs to be investigated. In Figure C.4, the relative E-efficiency is plotted against n. Similar comments apply as that for locally decision-theoretic A-optimal designs, see Figure C.3.

The increase occurred in the exact $\lambda_{max}\{R_{SE}(\boldsymbol{\theta}, d)\}$ is investigating by plotting the contour plots of all approximations considered to $\lambda_{max}\{R_{SE}(\boldsymbol{\theta}, d)\}$ in addition to the exact $\lambda_{max}\{R_{SE}(\boldsymbol{\theta}, d)\}$ for $n = 20$. Figure C.14 shows five contour plots of the exact and approximations to $\lambda_{max}\{R_{SE}(\boldsymbol{\theta}, d)\}$. From this figure (the first row on left), it is found that ST design is far from True design. It is also noted that the contour plot of the ST approximation (the first row on right) to $\lambda_{max}\left\{\widetilde{R}_{SE}(\boldsymbol{\theta}, d)\right\}$ has multi-modal (multiple local minima). Therefore, the modified quasi-Newton method ("L-BFGS-B") may fail to find the local optimal design. The other local minima is $d_{ST} = (-5.5, 5.5)^T$ that might be considered where $\lambda_{max}\{R_{SE}(\boldsymbol{\theta}, d_{ST})\} = 0.12$.

Figure C.15 shows plots of the approximations to the MSE against the exact MSE to check the validity of the used approximations to the MSE. It can be seen that the QMC, Quad and ST approximations are very accurate.

We calculate the RT for each approximation to $\lambda_{max}\{R_{SE}(\boldsymbol{\theta}, d)\}$, where the R.T for the QMC approximation is 46.82 indicating that the QMC approximation is more expensive to be computed than the exact MSE. Similarly, the Quad approximation to the MSE (RT=1.55) requires more executing time than the exact MSE. However, the ST and locally E-optimal designs require less executing time to be found than the True design, with RT= 0.57 and 0.39, respectively.

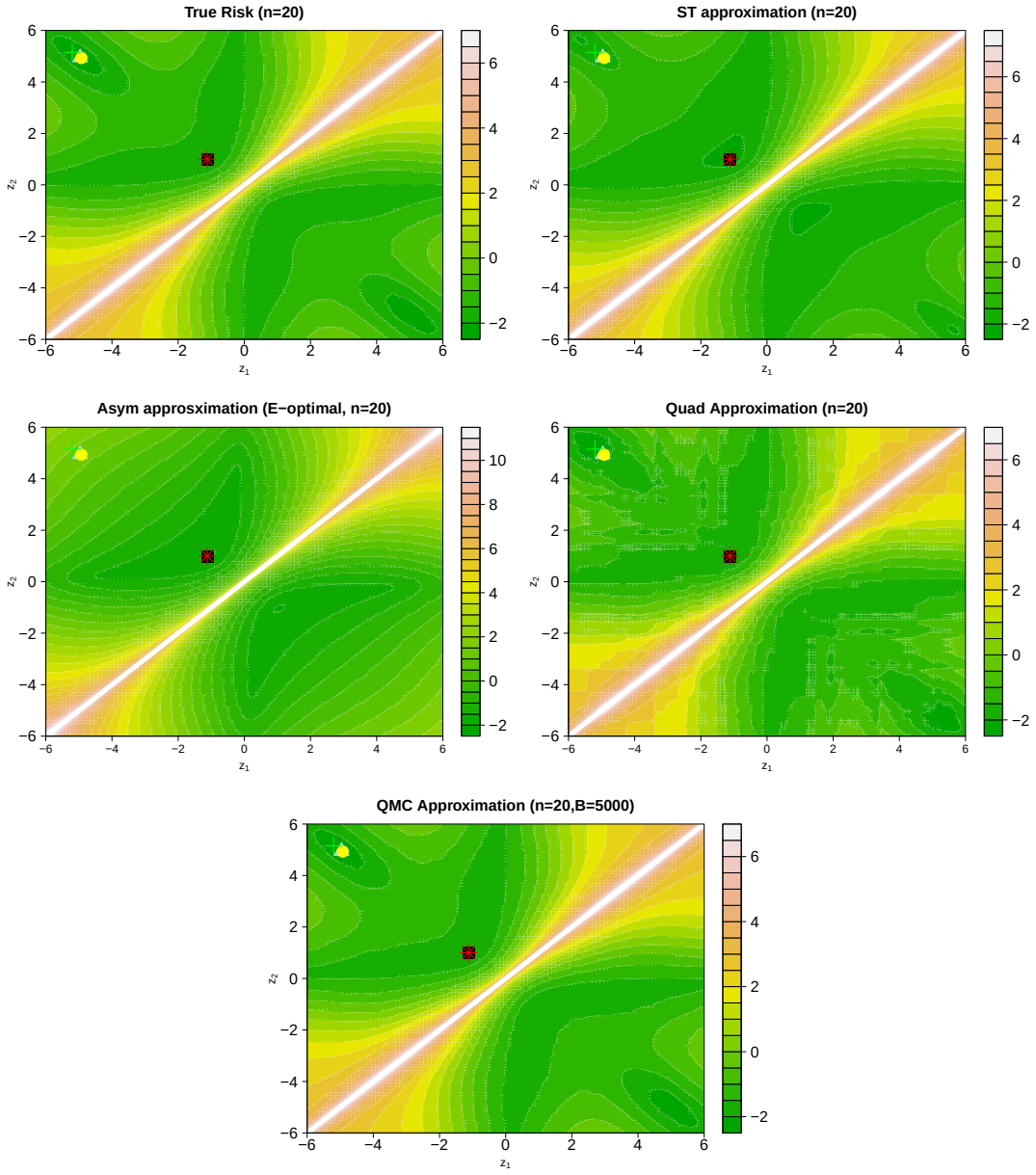

Figure C.14: Contour plots of $\lambda_{max}\{R_{SE}(\boldsymbol{\theta}, d)\}$ (the first row on left) and the maximum eigenvalue of the approximated mean squared error $\lambda_{max}\left\{\widetilde{R}_{SE}(\boldsymbol{\theta}, d)\right\}$ via ST (the first row on right), Asym (the second row on left), Quad (the second row on right) and QMC with size $B = 5,000$ (the third row), under the simple logistic regression model for a sample of size $n = 5$. A plot key for each contour is plotted which refers to the values of $\lambda_{max}\{R_{SE}(\boldsymbol{\theta}, d)\}$ and $\lambda_{max}\left\{\widetilde{R}_{SE}(\boldsymbol{\theta}, d)\right\}$ for the remaining contour plots. In Each plot, there are five coloured-shapes denote the decision-theoretic E-optimal designs found under the exact risk (light blue triangle), ST(black square), Asym (red star), Quad (green plus) and QMC approximation (yellow filled circle).

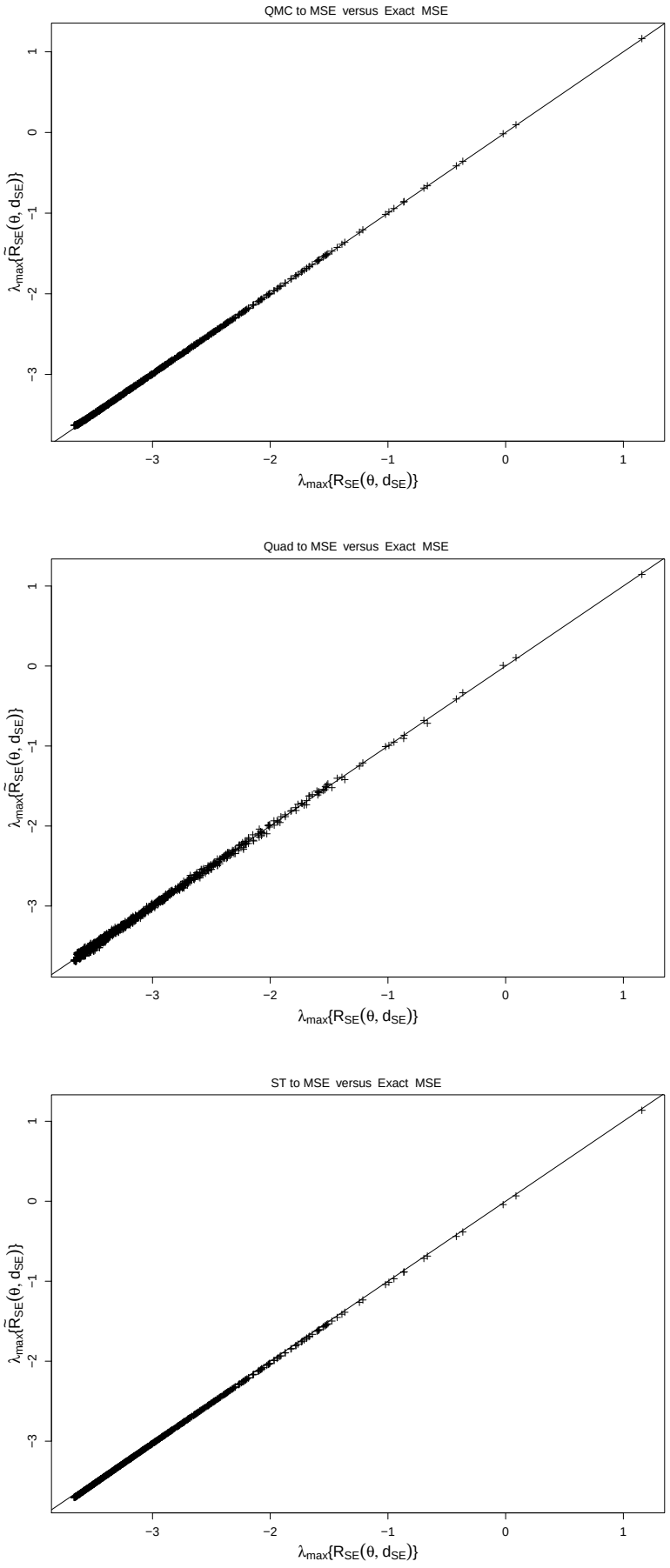

Figure C.15: Plots of the QMC (the first row), Quad (the second row) and ST (the third row) approximations to the MSE $\lambda_{max}\left\{\widetilde{R}_{SE}(\boldsymbol{\theta}, d)\right\}$ against the exact MSE $\lambda_{max}\left\{R_{SE}(\boldsymbol{\theta}, d)\right\}$ for simple logistic regression for $n = 100$.

C.3.2 Pseudo-Bayesian decision-theoretic E-optimal designs

We find exact and approximate pseudo-Bayesian decision-theoretic E-optimal designs that minimise $\widetilde{E}_{\boldsymbol{\theta}}\left\{\lambda_{max}\left\{R_{SE}(\boldsymbol{\theta}, d)\right\}\right\}$ and $\widetilde{E}_{\boldsymbol{\theta}}\left\{\lambda_{max}\left\{\widetilde{R}_{SE}(\boldsymbol{\theta}, d)\right\}\right\}$, respectively. In the second row of Figure C.11, we evaluate $\widetilde{E}_{\boldsymbol{\theta}}\left\{\lambda_{max}\left\{R_{SE}(\boldsymbol{\theta}, d)\right\}\right\}$ at the resulting designs, and found that all design perform well except QIMPS design under both $\pi_1(\boldsymbol{\theta})$ (left) and $\pi_2(\boldsymbol{\theta})$ (right). Surprisingly, the pseudo-Bayesian E-optimal design performs very well under both $\pi_1(\boldsymbol{\theta})$ and $\pi_2(\boldsymbol{\theta})$ for all values of n, where it has similar performance to the True design.

In Figure C.16, we plot the approximate expectation of the used approximations to the MSE against the approximate expectation of the exact MSE to check the validity of these approximations. All the approximations appear very accurate except the QIMPS that showed some inaccuracies. However, we can confirm that the approximate expectations of these approximations to the MSE are minimised for designs close to that minimises the approximate expectation of the exact MSE. This is based on the ordering of the designs, in terms of the approximate expectations of the MSE, that is the same as that for the approximate expectation of the exact MSE.

The RT for the approximate expectation of the QMC approximation to the MSE is 26.96, followed by RT= 22.34 for the QIMPS. This means that the approximate expectation of the QMC is approximately 27 times more expensive than the time required to compute the approximate expectation of the exact MSE. By contrast, other approximations require less executing time than that required for the approximate expectation of the exact MSE.

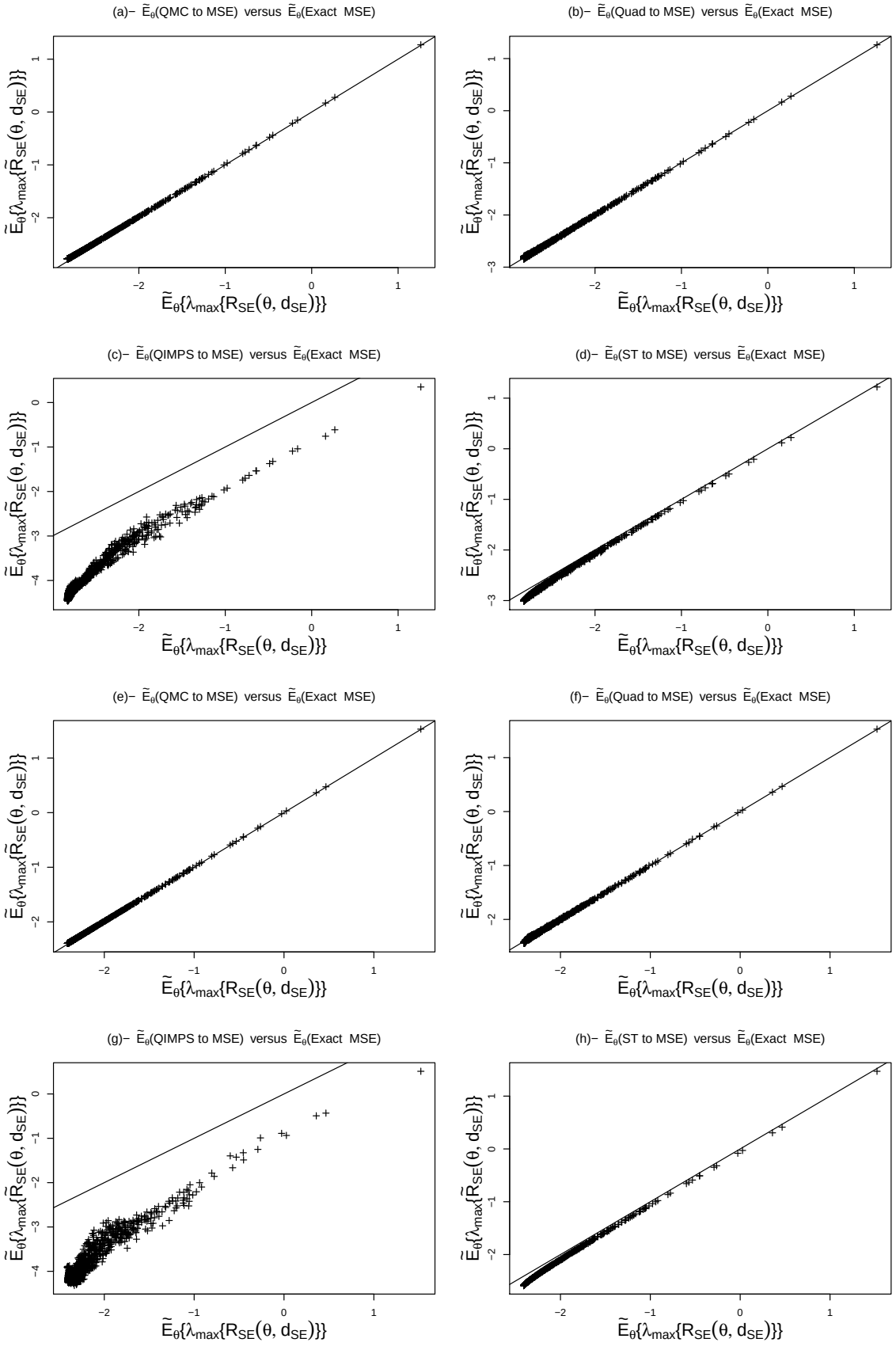

Figure C.16: Plots of the approximate expectation of approximations to the MSE, plotted against the approximate expectation of the exact MSE for $n = 100$ under $\pi_1(\boldsymbol{\theta})$ (the first two rows) and $\pi_2(\boldsymbol{\theta})$ (the last two rows).

Appendix D

D.1 Parameter estimation: four factors logistic regression model

D.1.1 Pseudo-Bayesian decision-theoretic D-optimal designs

Pseudo-Bayesian decision-theoretic D-optimal designs found under different methods for different experimental runs.

- True design found under the exact risk (d_T)
  - Design for $N = 6$

	k1	k2	k3	k4
1	-1.00	-0.87	1.00	-1.00
2	-0.79	-1.00	1.00	-1.00
3	-1.00	-1.00	1.00	1.00
4	-1.00	-1.00	0.38	-1.00
5	1.00	1.00	-1.00	1.00
6	1.00	1.00	-1.00	1.00

  - Design for $N = 9$

	k1	k2	k3	k4
1	1.00	1.00	-1.00	1
2	-1.00	-0.86	1.00	-1
3	-1.00	-1.00	1.00	1
4	-1.00	-1.00	1.00	1
5	-1.00	-1.00	1.00	1
6	-1.00	-1.00	0.39	-1
7	1.00	1.00	-1.00	1
8	-0.80	-1.00	1.00	-1
9	-1.00	-1.00	0.39	-1

- IMPS design found under the IMPS approximation to the risk (d_{IMPS})
  - Design for $N = 6$

	k1	k2	k3	k4
1	0.86	0.99	0.23	1.00
2	0.42	0.89	-0.27	-0.19
3	-1.00	-0.90	1.00	0.33
4	-0.81	-1.00	-0.21	-0.58
5	-1.00	-0.80	-0.08	0.06
6	-0.78	-0.82	0.98	-0.75

– Design for $N = 9$

	k1	k2	k3	k4
1	0.68	1.00	-0.84	-1.00
2	-0.87	-1.00	0.78	-1.00
3	-1.00	-1.00	0.34	-0.24
4	-1.00	-0.37	0.99	-0.09
5	-0.89	-0.44	0.92	-0.10
6	-0.93	-0.47	0.43	0.77
7	0.96	0.76	-1.00	-0.28
8	1.00	0.72	0.45	0.85
9	1.00	0.75	-0.37	0.89

– Design for $N = 16$

	k1	k2	k3	k4
1	-0.98	-0.57	0.25	-1.00
2	1.00	0.70	-1.00	-1.00
3	0.83	0.81	-0.98	-1.00
4	0.99	1.00	1.00	-0.70
5	-1.00	-1.00	0.95	0.40
6	1.00	1.00	1.00	-0.10
7	-0.64	-0.85	0.70	0.70
8	0.97	0.98	-0.44	-0.35
9	0.57	1.00	-0.62	1.00
10	-1.00	-0.70	-0.19	-1.00
11	-0.77	-1.00	-0.24	1.00
12	-1.00	-0.88	-0.46	-0.52
13	1.00	0.85	0.10	-0.61
14	1.00	0.65	-0.35	-1.00
15	-1.00	-0.62	0.05	-1.00
16	1.00	0.85	0.22	-1.00

- Pseudo-Bayesian D-optimal design found under the Asym approximation to the risk (d_D)
  – Design for $N = 6$

	k1	k2	k3	k4
1	-0.96	0.25	-1.00	-1.00
2	-0.79	0.85	0.93	0.48
3	-0.95	0.82	-0.57	1.00
4	0.85	-0.88	0.19	1.00
5	0.36	0.29	1.00	-1.00
6	0.30	-0.45	-1.00	-0.48

– Design for $N = 9$

	k1	k2	k3	k4
1	-1.00	1.00	1.00	1.00
2	1.00	-0.92	0.76	1.00
3	0.52	-0.56	-1.00	-1.00
4	1.00	-0.40	0.65	1.00
5	-0.12	0.87	1.00	-1.00
6	-1.00	0.29	-1.00	1.00
7	-0.75	0.87	1.00	-1.00
8	-0.12	-0.51	-1.00	-1.00
9	-0.99	0.84	-1.00	0.99

- QMC design found under the QMC approximation with $B = 1,000$ to the risk (d_{QMCL1})
  - Design for $N = 6$

	k1	k2	k3	k4
1	-0.93	-0.84	0.21	0.26
2	0.91	1.00	-1.00	0.77
3	-1.00	-0.95	0.49	-0.39
4	-1.00	-0.91	-0.12	0.58
5	-1.00	-0.96	0.34	0.19
6	0.93	0.74	-0.18	0.42

  - Design for $N = 8$

	k1	k2	k3	k4
1	-1.00	-0.90	-0.40	0.22
2	1.00	0.98	-0.52	0.51
3	-1.00	-1.00	0.13	0.17
4	-0.85	-1.00	0.47	0.39
5	1.00	1.00	-0.20	1.00
6	1.00	0.81	0.09	0.46
7	-1.00	-1.00	0.04	-0.25
8	0.94	1.00	-0.91	-1.00

- QMC design found under the QMC approximation with $B = 500$ to the risk (d_{QMCL2})
  - Design for $N = 6$

	k1	k2	k3	k4
1	-1.00	-0.75	0.59	-0.99
2	-0.97	-1.00	1.00	0.42
4	1.00	1.00	-0.90	0.67
5	-0.97	-1.00	1.00	0.95
6	-1.00	-1.00	1.00	0.75

  - Design for $N = 9$

	k1	k2	k3	k4
1	-0.96	-1.00	0.86	-0.20
2	1.00	1.00	-0.99	-1.00
3	1.00	1.00	-1.00	-0.69
4	1.00	1.00	-1.00	-0.93
5	0.97	0.94	-1.00	-0.11
6	1.00	1.00	-0.79	-0.49
7	1.00	0.73	-0.31	0.21
8	-0.96	-0.80	0.24	0.08
9	1.00	0.79	-0.55	0.29

– Design for $N = 16$

	k1	k2	k3	k4
1	1.00	1.00	0.22	1.00
2	-0.83	-1.00	0.98	-0.77
3	-1.00	-1.00	1.00	-1.00
4	-0.95	-0.92	-0.10	-1.00
5	1.00	1.00	-0.49	-0.90
6	1.00	1.00	-0.67	0.43
7	1.00	1.00	-1.00	1.00
8	-1.00	-0.89	0.70	-1.00
9	-1.00	-1.00	1.00	0.44
10	1.00	0.91	-0.52	0.17
11	-1.00	-0.92	0.83	1.00
12	-1.00	-1.00	0.23	1.00
13	0.99	0.93	-0.79	1.00
14	-1.00	-1.00	0.09	1.00
15	-1.00	-1.00	1.00	-0.73
16	-0.99	-0.91	0.78	-0.81

- Quad design found under the Quad approximation to the risk (d_{Quad})
  - Design for $N = 6$

	k1	k2	k3	k4
1	-1.00	-1.00	1.00	-0.15
2	1.00	1.00	-0.90	0.29
3	-1.00	-1.00	0.70	0.45
4	-1.00	-1.00	1.00	-0.01
5	-0.99	-1.00	1.00	-0.78
6	-1.00	-1.00	1.00	-0.70

  - Design for $N = 10$

	k1	k2	k3	k4
1	1.00	0.84	-1.00	0.80
2	-1.00	-1.00	0.94	-0.15
3	1.00	1.00	-1.00	0.88
4	1.00	1.00	-0.55	1.00
5	0.96	1.00	-0.99	-1.00
6	1.00	1.00	-0.90	0.31
7	-1.00	-1.00	1.00	0.46
8	1.00	0.97	-1.00	0.98
9	0.86	0.84	-1.00	0.19
10	1.00	0.97	-1.00	0.80

- ST design found under the ST approximation to the risk (d_{ST})
  - Design for $N = 6$

	k1	k2	k3	k4
1	-1.00	-0.85	1.00	-0.89
2	-1.00	-1.00	0.61	-1.00
3	1.00	1.00	-1.00	1.00
4	-1.00	-1.00	0.97	0.10
5	-0.82	-1.00	1.00	-0.83
6	1.00	1.00	-1.00	1.00

- WSMC design found under WSMC approximation to the risk with $B = 1,000$ (d_{WSMCL1})
  - Design for $N = 6$

	k1	k2	k3	k4
1	1	1.00	-0.20	1.00
2	1	1.00	-0.99	0.35
3	1	0.98	-1.00	0.82
4	-1	-0.97	0.71	0.47
5	-1	-1.00	1.00	-0.76
6	-1	-1.00	1.00	0.07

  - Design for $N = 13$

	k1	k2	k3	k4
1	-1.00	-1.00	-0.26	0.96
2	1.00	1.00	-0.75	-1.00
3	-0.98	-1.00	0.78	-1.00
4	-1.00	-0.89	0.51	-0.87
5	0.94	1.00	-0.81	-0.11
6	-1.00	-1.00	1.00	-0.35
7	-1.00	-1.00	1.00	-0.71
8	0.97	1.00	-0.82	1.00
9	1.00	1.00	-1.00	0.74
10	-1.00	-1.00	1.00	-1.00
11	-1.00	-0.96	1.00	1.00
12	1.00	0.99	-0.79	-0.28
13	-0.98	-1.00	1.00	-0.85

- WSMC design found under WSMC approximation to the risk with $B = 500$ (d_{WSMCL2})
  - Design for $N = 6$

	k1	k2	k3	k4
1	-0.95	-1.00	0.98	-0.65
2	-1.00	-1.00	1.00	-0.17
3	1.00	1.00	-0.84	1.00
4	1.00	0.81	-0.18	-0.09
5	-1.00	-1.00	0.39	-1.00
6	-1.00	-1.00	1.00	0.80

  - Design for $N = 16$

	k1	k2	k3	k4
1	-0.66	1.00	1.00	1
2	0.43	-0.72	-1.00	-1
3	-1.00	0.42	-1.00	1
4	-0.45	1.00	1.00	-1
5	-1.00	0.49	-1.00	-1
6	-1.00	0.99	1.00	-1
7	-0.40	-0.48	-1.00	1
8	-0.01	-0.70	-1.00	-1
9	1.00	-0.71	1.00	-1
10	-1.00	0.93	-1.00	-1
11	0.77	-1.00	-1.00	1
12	-0.83	0.92	-0.48	1
13	1.00	-0.35	1.00	1
14	0.64	-0.54	-1.00	-1
15	-0.40	0.33	1.00	1
16	1.00	-0.25	1.00	-1

D.1.2 Pseudo-Bayesian decision-theoretic A-optimal designs

Pseudo-Bayesian decision-theoretic A-optimal designs found under different methods for different experimental runs.

- True design found under the exact risk (d_T)
  - Design for $N = 6$

	k1	k2	k3	k4
1	0.27	-0.16	1.00	-0.97
2	0.62	0.06	-0.54	0.44
3	-0.95	0.65	0.37	-1.00
4	0.56	0.08	-1.00	0.90
5	0.20	0.32	-0.59	0.37
6	-0.56	0.91	1.00	0.81

  - Design for $N = 9$

	k1	k2	k3	k4
1	0.55	0.08	-0.36	-0.445
2	0.42	0.21	-0.34	-0.54
3	-0.28	0.88	0.91	-0.91
4	0.72	-0.72	0.56	1.00
5	0.94	-0.41	-1.00	-0.99
6	-1.00	0.43	-0.20	1.00
7	1.00	-0.47	-0.15	-0.01
8	0.63	0.07	-0.29	-0.52
9	-0.19	0.39	-0.57	-0.12

- IMPS design found under the IMPS approximation to the risk (d_{IMPS})
  - Design for $N = 6$

	k1	k2	k3	k4
1	0.70	-1.00	-1.00	0.02
2	1.00	-0.83	0.51	1.00
3	-0.37	-0.29	-0.70	-0.01
4	1.00	-0.27	1.00	-1.00
5	-0.20	0.41	-0.33	0.86
6	-1.00	0.48	-0.78	-1.00

  - Design for $N = 9$

	k1	k2	k3	k4
1	1.00	-0.67	1.00	-0.26
2	0.23	-0.31	-0.96	0.19
3	0.68	0.15	1.00	-0.82
4	0.39	-0.69	-0.99	-0.72
5	0.42	-0.53	0.37	0.64
6	0.01	0.01	-0.0001	1.00
7	-0.54	0.46	0.84	0.44
8	-0.64	0.81	0.37	-1.00
9	0.14	0.48	1.00	0.06

  - Design for $N = 16$

	k1	k2	k3	k4
1	-0.36	0.49	0.02	0.53
2	-0.98	0.33	-0.84	-1.00
3	0.57	-0.72	0.15	0.72
4	-0.83	0.56	0.25	1.00
5	-0.42	0.56	0.85	-1.00
6	-1.00	1.00	0.19	-0.23
7	0.42	-0.28	-0.87	-0.39
8	0.68	-1.00	0.029	0.48
9	0.52	0.12	0.87	-1.00
10	-0.97	0.45	-0.09	0.60
11	0.92	-0.79	0.77	-0.64
12	0.68	-0.74	-0.65	-0.54
13	-1.00	0.86	-0.68	0.78
14	0.51	-0.05	0.42	1.00
15	0.74	-1.00	-1.00	0.08
16	0.78	-0.44	0.74	1.00

- Pseudo-Bayesian A-optimal design found under the Asym approximation to the risk (d_A)
  - Design for $N = 6$

	k1	k2	k3	k4
1	0.34	-0.21	0.19	-0.91
2	0.49	-0.59	-0.07	0.41
3	-0.07	0.38	0.81	0.20
4	0.53	-0.08	0.04	-0.05
5	-0.53	0.42	-0.43	-0.02
6	-0.49	0.10	0.08	-0.12

  - Design for $N = 9$

	k1	k2	k3	k4
1	0.27	-0.65	-0.20	0.96
2	0.29	-0.16	0.97	-0.19
3	0.60	0.06	0.94	-0.23
4	0.33	0.14	-0.11	-0.24
5	-0.69	0.33	-0.04	0.004
6	0.49	-0.37	-0.37	1.00
7	0.02	-0.49	-0.47	-0.62
8	0.25	-0.28	-0.67	-0.68
9	-0.57	0.70	-0.12	0.06

- QMC design found under the QMC approximation to the risk with $B = 1,000$ (d_{QMCL1})
  - Design for $N = 6$

	k1	k2	k3	k4
1	-0.09	0.39	0.28	-1.00
2	-0.01	-0.88	1.00	1.00
3	-0.73	0.74	-0.95	-0.26
4	-1.00	0.58	1.00	0.63
5	-0.35	-0.36	0.55	0.48
6	-0.55	-0.18	0.48	0.61

– Design for $N = 7$

	k1	k2	k3	k4
1	-1.00	0.22	-1.00	-0.87
2	-0.71	-0.03	0.18	0.31
3	0.65	-0.80	-0.90	0.13
4	1.00	-0.52	0.51	-1.00
5	-0.54	-0.10	0.25	0.25
6	-0.28	-0.19	0.43	0.22
7	-1.00	0.40	1.00	0.96

- QMC design found under the QMC approximation to the risk with $B = 500$ (d_{QMCL2})
  – Design for $N = 6$

	k1	k2	k3	k4
1	-0.85	0.80	0.31	1.00
2	1.00	-0.47	-1.00	1.00
3	-0.12	0.13	0.87	-0.69
4	-0.77	0.28	-0.42	-0.96
5	0.30	0.05	-0.58	0.47
6	0.64	-0.17	-0.52	0.26

– Design for $N = 9$

	k1	k2	k3	k4
1	1.00	0.24	1.00	-0.94
2	-1.00	0.47	-0.19	0.19
3	0.45	-0.49	-0.98	-0.86
4	0.63	-0.09	0.20	0.21
5	0.01	0.62	0.28	-0.44
6	-0.41	0.60	0.19	-0.15
7	0.46	0.36	0.31	-0.41
8	-0.15	-0.25	-0.13	1.00
9	0.61	0.15	0.19	-0.32

– Design for $N = 16$

	k1	k2	k3	k4
1	-0.54	0.33	-0.52	0.85
2	-0.49	0.14	-0.76	-0.60
3	0.70	-0.07	0.81	-0.15
4	-0.57	0.52	0.74	0.08
5	-0.47	0.79	0.58	0.06
6	0.63	-0.63	-0.66	0.20
7	0.29	0.29	0.11	-0.04
8	0.44	-0.79	-0.40	0.23
9	0.55	-0.40	0.86	0.13
10	-0.26	-0.19	-0.13	-0.81
11	0.64	-0.24	0.47	0.64
12	-0.09	0.37	-0.35	0.48
13	-0.64	0.06	-0.41	0.84
14	-0.02	-0.32	0.37	0.14
15	0.04	-0.03	0.04	-1.00
16	0.12	0.19	-0.22	-0.88

- Quad design found under the Quad approximation to the risk (d_{Quad})
  - Design for $N = 6$

	k1	k2	k3	k4
1	-0.28	-0.27	0.54	-0.33
2	0.27	-0.25	-1.00	1.00
3	-0.36	-0.21	0.57	-0.34
4	-0.30	-0.29	1.00	-1.00
5	-1.00	0.91	-0.45	1.00
6	0.35	-0.28	0.99	1.00

  - Design for $N = 10$

	k1	k2	k3	k4
1	0.81	-0.45	0.17	-0.20
2	-0.22	0.24	-0.32	0.42
3	0.31	-0.66	-0.20	-0.71
4	0.15	0.36	-0.04	0.65
5	-0.68	-0.01	-0.75	0.04
6	0.26	0.51	0.61	-0.07
7	-0.52	0.68	0.65	-1.00
8	0.66	0.38	1.00	1.00
9	0.34	0.44	0.37	0.34
10	0.85	-0.13	0.26	0.33

- WSMC design found under the WSMC approximation to the risk with $B = 1,000$ (d_{WSMCL1})
  - Design for $N = 6$

	k1	k2	k3	k4
[1,] -0.12	-0.39	0.25	-0.27	
-0.63	0.45	-0.96	-0.20	
-0.65	0.71	-0.34	1.00	
0.55	-1.00	-0.92	0.79	
0.47	-0.92	1.00	-0.94	
-0.38	0.003	0.50	-0.25	

– Design for $N = 13$

	k1	k2	k3	k4
1	-0.09	0.19	-0.28	-0.86
2	0.84	-0.17	0.74	-0.34
3	0.66	-0.47	-0.17	0.62
4	-0.29	0.70	0.46	0.61
5	-0.24	0.03	-1.00	0.13
6	-0.50	0.50	0.49	0.65
7	0.77	-0.50	0.89	-0.42
8	-0.58	0.66	0.44	-0.30
9	0.01	0.39	-0.13	-0.27
10	-0.18	-0.17	-0.27	-0.84
11	-0.35	-0.30	-0.93	0.16
12	0.37	-0.57	0.09	0.66
13	-0.52	0.18	0.40	0.22

- WSMC design found under the WSMC approximation to the risk with $B = 500$ (d_{WSMCL2})
  – Design for $N = 6$

	k1	k2	k3	k4
1	-0.08	-0.40	0.30	-0.27
2	-0.67	0.50	-0.95	0.004
3	-0.62	0.71	-0.20	1.00
4	0.50	-0.93	-0.52	0.38
5	0.60	-1.00	1.00	-0.90
6	-0.40	0.03	0.42	-0.30

  – Design for $N = 16$

	k1	k2	k3	k4
1	-0.65	0.74	-0.02	-0.47
2	0.96	-0.16	0.85	-0.56
3	-0.17	-0.12	-0.29	0.94
4	0.39	-0.67	0.18	-0.50
5	0.26	0.27	0.67	0.63
6	0.66	-0.38	-0.71	-0.35
7	-0.80	0.59	0.06	-0.49
8	0.34	-0.57	-1.00	-0.38
9	0.37	-0.19	0.96	0.60
10	-0.06	0.08	-0.53	1.00
11	0.02	-0.65	-0.86	-0.39
12	-0.23	-0.27	-0.01	0.72
13	-0.33	0.70	-0.11	-0.46
14	-0.62	0.24	0.22	-0.35
15	0.46	-0.23	0.78	-0.61
16	0.68	-0.62	-0.17	-0.51

D.1.3 Pseudo-Bayesian decision-theoretic E-optimal designs

Pseudo-Bayesian decision-theoretic E-optimal designs found under different methods for different experimental runs.

- True design found under the exact risk (d_T)
  - Design for $N = 6$

	k1	k2	k3	k4
1	-0.03	-0.60	0.27	-0.03
2	-0.47	0.51	-1.00	-0.23
3	0.24	-0.92	0.95	0.17
4	0.31	-0.71	-1.00	-0.28
5	-0.24	-0.34	0.38	0.14
6	0.31	-0.62	-0.76	-0.89

  - Design for $N = 9$

	k1	k2	k3	k4
1	0.47	-0.87	1.00	0.93
2	-0.52	0.45	0.53	0.13
3	0.20	-0.38	-0.06	-0.38
4	0.05	-0.49	0.40	0.14
5	-0.49	0.38	-0.60	0.26
6	-0.31	-0.14	0.03	0.52
7	-0.23	-0.23	0.57	0.41
8	-0.56	0.85	0.09	-0.28
9	0.02	0.28	-0.57	-0.83

- IMPS design found under the IMPS approximation to the risk (d_{IMPS})
  - Design for $N = 6$

	k1	k2	k3	k4
1	-0.89	0.88	-0.75	-0.75
2	0.10	0.29	0.80	-0.87
3	0.84	-1.00	-0.62	-0.14
4	0.20	-0.68	1.00	0.87
5	-0.13	-0.19	0.23	-0.23
6	-0.93	0.60	0.16	-0.28

– Design for $N = 9$

	k1	k2	k3	k4
1	-0.66	0.95	0.78	0.64
2	-0.38	-0.09	-0.98	0.95
3	-0.20	0.17	-0.50	1.00
4	-0.69	0.04	-1.00	-0.55
5	-0.16	0.16	0.80	-0.35
6	-0.54	0.68	-0.49	-0.15
7	-0.75	0.54	-0.83	-0.88
8	0.98	-0.63	0.62	-0.80
9	0.21	-0.36	0.26	0.83

– Design for $N = 16$

	k1	k2	k3	k4
1	-0.39	0.16	-1.00	0.68
2	1.00	-0.93	-0.75	-1.00
3	-0.37	0.13	0.60	-0.14
4	1.00	-0.32	0.53	0.91
5	-0.41	0.31	0.25	-1.00
6	0.60	-0.41	0.98	0.30
7	-0.43	1.00	0.70	0.58
8	-1.00	0.92	0.30	0.19
9	-0.69	1.00	0.48	-0.67
10	0.69	-1.00	-0.81	0.84
11	-0.47	0.28	0.53	0.001
12	-0.52	0.05	-0.70	0.80
13	-0.41	0.70	0.82	0.58
14	-0.51	-0.29	-0.63	1.00
15	-0.34	0.31	-0.71	-0.10
16	0.97	-0.47	0.85	0.56

- Pseudo-Bayesian E-optimal design found under the Asym approximation to the risk (d_D)
  – Design for $N = 6$

	k1	k2	k3	k4
1	-0.08	0.08	-0.57	0.12
2	0.04	-0.23	0.31	0.40
3	-0.64	0.24	-0.14	-0.09
4	-0.14	0.46	0.18	-0.39
5	0.21	-0.37	-0.08	-0.38
6	0.34	0.05	0.16	0.26

– Design for $N = 9$

	k1	k2	k3	k4
1	0.53	-0.13	0.30	0.04
2	-0.63	0.12	-0.40	-0.27
3	0.27	-0.45	0.23	0.06
4	0.04	0.02	-0.04	0.66
5	-0.36	0.64	0.20	-0.14
6	0.15	0.30	-0.09	0.23
7	-0.40	0.22	0.44	-0.06
8	-0.13	-0.25	0.07	0.49
9	-0.09	0.14	-0.49	-0.14

- QMC design found under the QMC approximation with $B = 1,000$ to the risk (d_{QMCL1})
  - Design for $N = 6$

	k1	k2	k3	k4
1	-0.10	-0.56	0.32	0.06
2	-0.43	0.36	-1.00	-0.18
3	0.24	-0.92	1.00	0.12
4	0.43	-0.91	-1.00	-0.07
5	-0.22	-0.41	0.37	0.08
6	0.32	-0.81	-0.43	-0.70

  - Design for $N = 7$

	k1	k2	k3	k4
1	1.00	-0.26	-1.00	-1.00
2	0.09	0.32	1.00	-0.39
3	0.18	0.15	-0.62	-0.26
4	0.77	0.02	-0.28	-0.59
5	-0.11	0.01	0.73	0.63
6	0.50	-0.04	-0.15	-0.05
7	-0.67	0.24	0.03	0.59

- QMC design found under the QMC approximation with $B = 500$ to the risk (d_{QMCL2})
  - Design for $N = 6$

	k1	k2	k3	k4
1	-0.08	-0.52	0.33	-0.13
2	-0.59	0.60	-1.00	-0.18
3	0.24	-0.91	1.00	0.12
4	0.41	-0.69	-1.00	-0.94
5	-0.27	-0.33	0.47	0.10
6	0.11	-0.52	0.95	-0.94

  - Design for $N = 9$

	k1	k2	k3	k4
1	0.21	-0.67	0.47	-0.29
2	-0.07	0.01	-0.99	0.29
3	0.14	-0.07	-0.23	0.89
4	0.02	-0.47	0.39	-0.39
5	-0.20	-0.05	0.57	-0.36
6	0.36	-0.69	0.36	0.23
7	-0.66	0.85	-0.02	0.08
8	0.68	-0.98	1.00	-0.85
9	0.13	-0.55	-0.24	-0.27

– Design for $N = 16$

	k1	k2	k3	k4
1	0.17	0.29	-0.09	-0.30
2	0.58	-0.40	0.30	0.10
3	-0.18	-0.12	-0.45	0.21
4	-0.34	0.46	0.41	0.17
5	0.07	0.10	0.03	-0.42
6	-0.42	0.22	0.53	0.09
7	-0.11	0.57	0.33	0.21
8	-0.15	-0.33	0.05	-0.25
9	-0.41	-0.20	-0.37	0.19
10	-0.10	0.36	-0.56	0.08
11	-0.83	0.31	-0.38	-0.32
12	0.36	-0.52	0.36	0.06
13	-0.61	0.52	-0.50	-0.31
14	0.15	-0.31	-0.07	-0.47
15	-0.14	0.05	-0.63	0.17
16	0.84	-0.34	0.12	0.05

- Quad design found under the Quad approximation to the risk (d_{Quad})
  – Design for $N = 6$

	k1	k2	k3	k4
1	-0.09	-0.45	0.28	-0.09
2	-0.83	0.73	-0.31	1.00
3	-0.63	0.49	-1.00	-0.19
4	0.30	-0.77	0.98	-0.39
5	-0.33	-0.13	0.45	-0.15
6	0.36	-0.84	-1.00	0.72

  – Design for $N = 10$

	k1	k2	k3	k4
1	0.19	-0.64	0.07	0.02
2	-0.25	0.29	0.74	0.35
3	0.16	-0.41	-0.46	0.38
4	-0.21	0.61	0.29	0.71
5	0.06	-0.58	0.27	-0.21
6	-0.08	0.08	-0.22	0.44
7	-0.28	-0.13	0.21	-0.35
8	0.09	0.15	-0.72	0.88
9	-0.53	0.53	-0.21	-0.13
10	0.60	-1.00	0.80	-0.94

- WSMC design found under WSMC approximation to the risk with $B = 1,000$ (d_{WSMCL1})
  - Design for $N = 6$

	k1	k2	k3	k4
1	0.79	-0.25	-0.24	0.30
2	0.49	-0.18	0.57	0.94
3	-0.32	0.34	-0.59	0.14
4	1.00	-0.42	-0.94	1.00
5	0.60	-0.45	0.58	-0.28
6	-0.42	0.11	0.36	-0.17

  - Design for $N = 13$

	k1	k2	k3	k4
1	0.48	-0.19	-0.34	0.33
2	-0.08	-0.34	-0.06	0.40
3	0.62	-0.15	0.40	-0.07
4	-0.02	-0.03	-0.58	-0.37
5	-0.03	0.08	0.08	0.45
6	0.21	0.24	-0.12	-0.05
7	-0.10	-0.35	-0.52	-0.33
8	-0.09	0.14	0.25	-0.26
9	0.30	-0.41	-0.35	0.27
10	-0.27	-0.13	0.21	-0.08
11	-0.66	0.43	0.03	0.09
12	-0.44	0.64	-0.05	0.08
13	0.64	-0.53	0.53	-0.04

- WSMC design found under WSMC approximation to the risk with $B = 500$ (d_{WSMCL2})
  - Design for $N = 6$

	k1	k2	k3	k4
1	0.41	-0.18	0.59	-0.07
2	0.32	-0.61	-0.09	-0.03
3	-0.27	1.00	-0.14	0.91
4	0.23	0.35	0.21	0.20
5	0.19	0.11	-0.26	0.41
6	0.19	-0.14	0.29	-0.46

– Design for $N = 16$

	k1	k2	k3	k4
1	-0.58	0.68	-0.11	-0.24
2	0.28	-0.24	0.53	0.16
3	-0.13	-0.19	-0.03	0.51
4	0.41	-0.61	0.15	-0.26
5	0.08	0.37	0.004	-0.13
6	0.57	-0.17	-0.27	-0.10
7	-0.73	0.44	-0.003	-0.28
8	0.26	-0.31	-0.62	-0.10
9	-0.04	0.004	0.01	-0.14
10	0.004	0.05	-0.04	0.46
11	0.11	-0.55	-0.60	-0.12
12	-0.15	-0.37	0.07	-0.05
13	-0.29	-0.15	0.09	-0.05
14	0.15	0.21	-0.18	0.36
15	0.54	-0.06	0.43	0.18
16	0.54	-0.33	0.13	-0.30

D.1.4 Pseudo-Bayesian decision-theoretic optimal designs under the SL

Pseudo-Bayesian decision-theoretic optimal designs under the SL found under different methods for different experimental runs.

- True design found under the exact risk (d_T)
  - Design for $N = 6$

	k1	k2	k3	k4
1	0.02	-0.01	-0.49	1.00
2	0.58	1.00	0.87	-0.39
3	0.64	-0.26	0.76	-0.40
4	-0.66	0.07	-0.58	-1.00
5	-0.98	-0.31	0.18	0.06
6	0.49	0.69	-1.00	0.24

- Pseudo-Bayesian optimal design found under the Asym approximation to the risk (d_{Asym})
  - Design for $N = 6$

	k1	k2	k3	k4
1	-0.19	-0.11	-0.20	-0.99
2	-0.30	-0.14	-0.24	0.84
3	-0.53	0.59	-0.08	0.49
4	0.42	-0.28	-0.17	0.77
5	0.15	0.03	-0.26	-0.86
6	0.18	0.09	0.91	-0.04

  - Design for $N = 16$

	k1	k2	k3	k4
1	0.41	-0.68	0.30	0.99
2	-0.76	0.36	-0.04	-1.00
3	0.42	0.26	0.66	-0.99
4	-0.63	0.37	-1.00	1.00
5	-0.62	-0.13	-0.93	1.00
6	-0.50	0.31	1.00	0.88
7	0.23	-0.40	-0.88	-1.00
8	0.49	-0.35	-0.57	1.00
9	0.37	0.01	0.88	-1.00
10	-0.26	0.51	0.94	1.00
11	0.33	-0.27	1.00	-1.00
12	0.01	-0.56	-0.70	-0.95
13	-0.66	0.61	-0.41	-1.00
14	0.30	0.29	0.69	0.93
15	0.23	-0.09	-1.00	-0.80
16	0.46	-0.39	0.17	1.00

- QMC design found under the QMC approximation with $B = 1,000$ to the risk (d_{QMCL1})
  - Design for $N = 6$

	k1	k2	k3	k4
1	0.82	-0.71	-0.56	1.00
2	-0.67	-0.55	0.07	-0.28
3	0.91	0.16	1.00	-0.82
4	-0.88	0.79	0.29	1.00
5	0.60	0.02	-1.00	-0.25
6	0.13	-0.09	0.12	-1.00

- QMC design found under the QMC approximation with $B = 500$ to the risk (d_{QMCL2})
  - Design for $N = 6$

	k1	k2	k3	k4
1	-0.29	-1.00	-0.02	1.00
2	-0.63	-0.42	0.64	0.20
3	-0.49	0.43	0.38	1.00
4	-0.80	0.07	-0.98	-1.00
5	0.17	0.03	-0.81	-1.00
6	0.20	-0.03	0.67	-1.00

– Design for $N = 16$

	k1	k2	k3	k4
1	1.00	-0.46	0.48	1.00
2	-0.40	0.46	-0.60	1.00
3	-0.33	0.42	-0.05	-1.00
4	1.00	0.09	-0.14	-0.54
5	0.17	-0.10	0.18	-0.34
6	0.77	-0.99	0.74	1.00
7	-0.86	-0.15	0.69	-0.28
8	-0.87	0.78	-0.48	1.00
9	-0.80	0.38	-0.02	1.00
10	-0.24	-0.82	-1.00	-0.16
11	0.32	0.50	-0.62	0.31
12	0.55	-0.17	0.08	-1.00
13	-0.06	-0.51	0.72	-0.91
14	-0.55	0.34	-0.12	-1.00
15	-0.40	0.08	0.08	1.00
16	0.06	-0.10	0.44	-1.00

- Quad design found under the Quad approximation to the risk (d_{Quad})
  – Design for $N = 6$

	k1	k2	k3	k4
1	0.81	-0.43	0.72	-1.00
2	-0.68	-1.00	-1.00	-0.66
3	-0.93	-0.76	-0.23	-0.65
4	-0.78	0.26	-0.75	-0.84
5	-0.57	-0.71	0.39	-0.44
6	0.36	-0.29	-0.50	1.00

  – Design for $N = 9$

	k1	k2	k3	k4
1	0.06	-1.00	0.74	0.04
2	-0.37	0.02	-0.06	-1.00
3	0.65	0.39	-1.00	0.06
4	-0.78	0.77	0.02	1.00
5	1.00	0.75	0.77	-0.15
6	0.54	-0.31	0.25	1.00
7	0.86	-0.13	-0.07	1.00
8	0.43	-0.25	0.26	-1.00
9	0.04	0.18	-0.16	-1.00

- WSMC design found under WSMC approximation to the risk with $B = 1,000$ (d_{WSMCL1})
  – Design for $N = 6$

	k1	k2	k3	k4
1	-1.00	-0.01	-0.96	-0.91
2	-0.82	0.16	0.85	-0.45
3	-0.57	0.43	-0.25	1.00
4	-0.75	-0.42	0.16	0.37
5	-0.15	0.45	0.06	-0.94
6	0.36	-0.48	0.18	-0.48

– Design for $N = 13$

	k1	k2	k3	k4
1	0.54	0.04	-0.70	0.08
2	-0.34	-0.07	0.89	-0.69
3	-0.11	-0.14	0.97	1.00
4	0.02	0.43	-0.52	-1.00
5	-0.83	0.52	-1.00	1.00
6	0.58	-0.31	0.95	-0.75
7	-0.49	0.16	0.06	-1.00
8	-0.03	0.49	0.62	1.00
9	0.23	-0.74	-0.97	-0.85
10	0.65	-0.07	0.60	-1.00
11	0.61	-0.22	0.09	1.00
12	-0.37	0.34	-0.09	-1.00
13	-0.83	0.07	-0.74	1.00

- WSMC design found under WSMC approximation to the risk with $B = 500$ (d_{WSMCL2})
  – Design for $N = 6$

	k1	k2	k3	k4
1	0.76	-0.51	-0.09	-1.00
2	-0.93	-0.38	-0.46	0.03
3	0.32	0.44	-0.23	0.18
4	0.47	0.79	0.95	0.85
5	-0.55	0.28	-0.12	-1.00
6	0.10	-0.14	-0.26	1.00

  – Design for $N = 16$

	k1	k2	k3	k4
1	0.25	-0.61	-0.39	-1.00
2	-1.00	0.36	-0.08	0.61
3	-0.68	0.46	-0.60	1.00
4	0.28	-0.14	0.97	1.00
5	0.40	0.32	0.90	-1.00
6	0.52	-0.64	-1.00	1.00
7	0.29	-0.34	1.00	0.68
8	0.15	-0.75	-0.59	1.00
9	-0.11	-0.32	-0.46	-1.00
10	0.16	0.03	1.00	-1.00
11	-0.42	0.05	-0.64	-1.00
12	-0.50	0.78	-0.34	0.84
13	0.49	0.13	0.81	1.00
14	-0.25	0.49	-0.33	-1.00
15	0.33	-0.23	-1.00	0.88
16	0.58	-0.31	0.80	-0.93

D.2 Parameter estimation: two factors Poisson regression model

D.2.1 Pseudo-Bayesian decision-theoretic D-optimal designs

Pseudo-Bayesian decision-theoretic D-optimal designs found under different methods for different experimental runs.

- IMPS design found under the IMPS approximation to the risk (d_{IMPS})
  - Design for $N = 5$

	k1	k2
1	1.00	1.00
2	-1.00	-0.98
3	1.00	-1.00
4	-1.00	1.00
5	1.00	0.93

  - Design for $N = 28$

	k1	k2
1	1.00	1.00
2	1.00	1.00
3	-1.00	-1.00
4	-1.00	1.00
5	-1.00	-1.00
6	1.00	-1.00
7	1.00	1.00
8	1.00	1.00
9	1.00	-1.00
10	1.00	1.00
11	-1.00	1.00
12	1.00	1.00
13	-1.00	-1.00
14	-1.00	-0.92
15	-1.00	0.65
16	-1.00	1.00
17	1.00	-1.00
18	1.00	-0.97
19	1.00	-0.97
20	-1.00	-1.00
21	-1.00	-1.00
22	1.00	-0.99
23	1.00	0.98
24	1.00	-1.00
25	-1.00	1.00
26	-1.00	1.00
27	-1.00	1.00
28	-1.00	-1.00

- Pseudo-Bayesian D-optimal design found under the Asym approximation to the risk (d_D)
  - Design for $N = 5$

	k1	k2
1	1.00	-1.00
2	1.00	1.00
3	-1.00	-1.00
4	-1.00	1.00
5	-1.00	1.00

- QMC design found under the QMC approximation with $B = 1,000$ to the risk (d_{QMCL1})
  - Design for $N = 5$

	k1	k2
1	-1.00	-1.00
2	1.00	1.00
3	-1.00	1.00
4	1.00	1.00
5	1.00	-1.00

– Design for $N = 28$

	k1	k2
1	1.00	1.00
2	-1.00	-1.00
3	1.00	1.00
4	-1.00	1.00
5	1.00	-1.00
6	1.00	-1.00
7	-1.00	-1.00
8	-1.00	-1.00
9	-0.99	-1.00
10	-1.00	1.00
11	-1.00	-1.00
12	1.00	-1.00
13	1.00	-1.00
14	-1.00	1.00
15	1.00	1.00
16	-1.00	1.00
17	1.00	-1.00
18	1.00	-1.00
19	-1.00	-1.00
20	-1.00	1.00
21	1.00	1.00
22	-1.00	-1.00
23	-1.00	-1.00
24	1.00	1.00
25	-1.00	1.00
26	1.00	-1.00
27	-1.00	1.00
28	1.00	1.00

- ST design found under the Quad approximation to the risk (d_{ST})
  - Design for $N = 5$

	k1	k2
1	-1.00	-1.00
2	1.00	1.00
3	-0.28	1.00
4	1.00	-1.00
5	-1.00	1.00

  - Design for $N = 28$

	k1	k2
1	-1.00	1.00
2	1.00	-1.00
3	-0.32	1.00
4	-1.00	1.00
5	1.00	1.00
6	0.77	1.00
7	1.00	-1.00
8	1.00	1.00
9	-1.00	1.00
10	0.81	-1.00
11	-1.00	-1.00
12	1.00	1.00
13	1.00	1.00
14	1.00	-0.95
15	-1.00	-1.00
16	1.00	-0.62
17	0.01	1.00
18	1.00	1.00
19	-1.00	-1.00
20	1.00	-1.00
21	-1.00	1.00
22	1.00	-1.00
23	-1.00	-1.00
24	-1.00	1.00
25	1.00	1.00
26	-1.00	-1.00
27	-1.00	-1.00
28	-0.98	-1.00

D.2.2 Pseudo-Bayesian decision-theoretic A-optimal designs

Pseudo-Bayesian decision-theoretic A-optimal designs found under different methods for different experimental runs.

- IMPS design found under the IMPS approximation to the risk (d_{IMPS})
  - Design for $N = 5$

	k1	k2
1	1.00	1.00
2	-1.00	-1.00
3	1.00	-1.00
4	-1.00	1.00
5	1.00	1.00

  - Design for $N = 28$

	k1	k2
1	1.00	1.00
2	1.00	-0.96
3	-1.00	-1.00
4	-1.00	1.00
5	-0.99	-1.00
6	1.00	-1.00
7	-1.00	-0.95
8	-0.86	-1.00
9	1.00	1.00
10	-1.00	-0.95
11	-1.00	-1.00
12	1.00	-1.00
13	1.00	-1.00
14	-1.00	1.00
15	0.35	0.44
16	-0.95	1.00
17	1.00	-1.00
18	1.00	0.88
19	1.00	-0.98
20	1.00	0.98
21	0.33	1.00
22	-1.00	1.00
23	0.94	-0.97
24	-1.00	-1.00
25	1.00	1.00
26	-1.00	0.88
27	-1.00	1.00
28	0.20	-1.00

- Pseudo-Bayesian A-optimal design found under the Asym approximation to the risk (d_A)
  - Design for $N = 5$

	k1	k2
1	1.00	1.00
2	1.00	-1.00
3	-1.00	1.00
4	-1.00	-1.00
5	-0.003	0.27

- QMC design found under the QMC approximation with $B = 1,000$ to the risk (d_{QMCL1})
  - Design for $N = 5$

	k1	k2
1	-1.00	-1.00
2	-1.00	1.00
3	1.00	-1.00
4	-1.00	1.00
5	1.00	1.00

– Design for $N = 28$

	k1	k2
1	-1.00	1.00
2	1.00	1.00
3	-1.00	-1.00
4	1.00	-1.00
5	-0.08	1.00
6	-1.00	1.00
7	1.00	1.00
8	1.00	-1.00
9	0.03	-0.54
10	1.00	-1.00
11	1.00	-1.00
12	0.15	1.00
13	-0.07	1.00
14	0.03	-1.00
15	-1.00	-1.00
16	0.05	1.00
17	1.00	1.00
18	-1.00	1.00
19	-0.05	1.00
20	0.13	-1.00
21	-1.00	-1.00
22	0.12	-1.00
23	0.09	-0.95
24	-1.00	-1.00
25	1.00	-1.00
26	1.00	1.00
27	-1.00	-1.00
28	-1.00	1.00

- IMPS design found under the IMPS approximation to the risk (d_{IMPS})
  – Design for $N = 5$

	k1	k2
1	1.00	-1.00
2	-0.96	1.00
3	-1.00	-1.00
4	-1.00	1.00
5	0.98	0.93

- ST design found under the Quad approximation to the risk (d_{ST})
  – Design for $N = 5$

	k1	k2
1	1.00	-1.00
2	1.00	1.00
3	0.08	0.03
4	-1.00	1.00
5	-1.00	-1.00

– Design for $N = 28$

	k1	k2
1	1.00	1.00
2	1.00	1.00
3	-1.00	-1.00
4	-1.00	1.00
5	1.00	-1.00
6	-1.00	-1.00
7	1.00	-1.00
8	0.18	-0.17
9	-0.17	0.12
10	1.00	-1.00
11	-0.23	1.00
12	1.00	-1.00
13	-1.00	1.00
14	-1.00	-1.00
15	-0.43	-1.00
16	1.00	-0.78
17	-0.03	1.00
18	1.00	1.00
19	-1.00	1.00
20	1.00	1.00
21	-0.13	-1.00
22	0.19	1.00
23	-0.09	-1.00
24	-1.00	1.00
25	0.93	1.00
26	0.15	1.00
27	-1.00	-1.00
28	-1.00	1.00

D.2.3 Pseudo-Bayesian decision-theoretic E-optimal designs

Pseudo-Bayesian decision-theoretic E-optimal designs found under different methods for different experimental runs.

- IMPS design found under the IMPS approximation to the risk (d_{IMPS})
  - Design for $N = 5$

	k1	k2
1	0.99	1.00
2	-1.00	0.95
3	-0.95	-0.99
4	-1.00	1.00
5	0.92	-0.99

  - Design for $N = 28$

	k1	k2
1	0.21	-0.83
2	-1.00	-1.00
3	1.00	-0.95
4	-1.00	-0.99
5	0.84	1.00
6	-0.95	1.00
7	-1.00	-1.00
8	-0.96	-1.00
9	0.73	-1.00
10	0.97	-1.00
11	-0.96	-1.00
12	0.95	1.00
13	-1.00	-1.00
14	0.79	1.00
15	-1.00	1.00
16	-1.00	0.95
17	1.00	1.00
18	1.00	-1.00
19	-0.10	1.00
20	1.00	-1.00
21	-1.00	1.00
22	1.00	-0.86
23	1.00	1.00
24	-0.73	-1.00
25	1.00	1.00
26	-1.00	0.99
27	0.95	1.00
28	-0.64	1.00

- Pseudo-Bayesian E-optimal design found under the Asym approximation to the risk (d_E)
  - Design for $N = 5$

	k1	k2
1	1.00	-1.00
2	-1.00	-1.00
3	1.00	1.00
4	-1.00	1.00
5	0.02	-0.25

- QMC design found under the QMC approximation with $B = 1,000$ to the risk (d_{QMCL1})
  - Design for $N = 5$

	k1	k2
1	0.94	0.96
2	-1.00	-1.00
3	-1.00	-1.00
4	-1.00	0.98
5	0.99	-0.99

– Design for $N = 28$

	k1	k2
1	-1.00	-1.00
2	-1.00	1.00
3	-0.09	-1.00
4	1.00	1.00
5	0.13	-0.20
6	-1.00	1.00
7	1.00	-1.00
8	1.00	1.00
9	-0.003	0.31
10	0.10	-0.08
11	1.00	1.00
12	-1.00	1.00
13	0.04	0.29
14	0.05	0.16
15	-1.00	-1.00
16	-0.09	-1.00
17	1.00	1.00
18	-1.00	1.00
19	0.03	-0.21
20	-0.04	1.00
21	-1.00	-1.00
22	0.12	0.11
23	0.03	1.00
24	1.00	-1.00
25	1.00	-1.00
26	1.00	-1.00
27	0.10	-1.00
28	-1.00	-1.00

- ST design found under the ST approximation to the risk (d_{ST})
  - Design for $N = 5$

	k1	k2
1	1.00	-1.00
2	-0.99	-1.00
3	-0.05	-0.16
4	-1.00	1.00
5	0.98	1.00

  - Design for $N = 28$

	k1	k2
1	-0.11	1.00
2	1.00	1.00
3	1.00	-1.00
4	-1.00	1.00
5	-1.00	1.00
6	0.15	-0.61
7	1.00	1.00
8	-0.48	-0.05
9	0.23	-0.30
10	-1.00	-1.00
11	-1.00	-1.00
12	-1.00	1.00
13	1.00	1.00
14	-0.22	0.57
15	0.03	-1.00
16	0.47	0.78
17	0.06	1.00
18	1.00	-1.00
19	-1.00	1.00
20	-1.00	-1.00
21	0.12	-0.36
22	1.00	1.00
23	-0.07	-0.99
24	1.00	-1.00
25	1.00	-1.00
26	0.12	0.75
27	-0.04	-1.00
28	-1.00	0.27

D.2.4 Pseudo-Bayesian decision-theoretic optimal designs under the SL

Pseudo-Bayesian decision-theoretic optimal designs found under different methods for different experimental runs.

- Pseudo-Bayesian optimal design found under the Asym approximation to the risk (d_{Asym})
  - Design for $N = 5$

	k1	k2
1	1.00	1.00
2	-1.00	1.00
3	-1.00	-1.00
4	1.00	-1.00
5	-0.12	1.00

  - Design for $N = 28$

	k1	k2
1	1.00	1.00
2	-1.00	-1.00
3	-1.00	1.00
4	-1.00	-1.00
5	1.00	-1.00
6	-1.00	-1.00
7	-1.00	1.00
8	1.00	1.00
9	1.00	-1.00
10	-1.00	1.00
11	-1.00	-1.00
12	1.00	1.00
13	-1.00	1.00
14	1.00	-1.00
15	1.00	-1.00
16	1.00	1.00
17	-1.00	-1.00
18	-1.00	-1.00
19	-1.00	1.00
20	1.00	-1.00
21	-1.00	1.00
22	1.00	1.00
23	1.00	-1.00
24	-0.18	1.00
25	0.12	-1.00
26	0.12	-1.00
27	-0.18	1.00
28	-0.18	1.00

- QMC design found under the QMC approximation with $B = 1,000$ to the risk (d_{QMCL1})
  - Design for $N = 5$

	k1	k2
1	-1.00	1.00
2	0.99	-1.00
3	1.00	1.00
4	-1.00	-1.00
5	1.00	-1.00

  - Design for $N = 28$

	k1	k2
1	1.00	-1.00
2	-1.00	1.00
3	1.00	-1.00
4	-1.00	1.00
5	-1.00	-1.00
6	1.00	1.00
7	-1.00	1.00
8	1.00	1.00
9	0.01	-1.00
10	0.98	1.00
11	0.93	-1.00
12	-1.00	-1.00
13	-0.04	0.87
14	-0.35	1.00
15	1.00	1.00
16	-0.22	0.77
17	-0.42	-1.00
18	-1.00	-0.95
19	-0.88	-0.75
20	1.00	-1.00

D.3 Model discrimination: two factors logistic regression

D.3.1 Pseudo-Bayesian decision-theoretic optimal designs under the $0-1$ loss

Pseudo-Bayesian decision-theoretic optimal designs found under different methods for different experimental runs.

- True designs found under the exact risk (d_T)
  - Design for $N = 6$ and $\pi_1(\boldsymbol{\theta})$

	k1	k2
1	0.24	-1.00
2	1.00	1.00
3	0.99	-0.14
4	-1.00	-1.00
5	-1.00	0.17
6	-0.19	0.94

  - Design for $N = 6$ and $\pi_2(\boldsymbol{\theta})$

	k1	k2
1	-0.15	1.00
2	1.00	-0.08
3	1.00	0.95
4	0.10	-0.96
5	-1.00	-1.00
6	-1.00	0.15

– Design for $N = 6$ and $\pi_3(\boldsymbol{\theta})$

	k1	k2
1	0.91	-1.00
2	1.00	-0.16
3	-0.99	0.06
4	0.66	0.93
5	-0.23	1.00
6	0.07	-1.00

– Design for $N = 6$ and $\pi_4(\boldsymbol{\theta})$

	k1	k2
1	0.18	-0.95 2
-1.00	-0.52	
3	-0.17	0.98
4	1.00	-0.13
5	-0.97	0.11
6	0.99	-0.76

- QMC design found under the QMC approximation with $B = 1,000$ to the risk (d_{QMCL1})
  – Design for $N = 7$ and $\pi_1(\boldsymbol{\theta})$

	k1	k2
1	-0.17	1.00
2	-0.01	1.00
3	0.18	-1.00
4	1.00	-0.04
5	-1.00	-1.00
6	1.00	-0.08
7	-1.00	0.17

– Design for $N = 7$ and $\pi_2(\boldsymbol{\theta})$

	k1	k2
1	-0.15	1.00
2	-0.03	1.00
3	-1.00	-0.52883279
4	1.00	-0.11765439
5	1.00	-1.00000000
6	-1.00	0.08985813
7	0.10	-1.00

– Design for $N = 7$ and $\pi_3(\boldsymbol{\theta})$

	k1	k2
1	-1.00	0.04
2	-0.08	0.95
3	0.15	-0.99
4	-1.00	0.48
5	0.94	-0.11
6	1.00	1.00
7	1.00	-0.13

– Design for $N = 7$ and $\pi_4(\boldsymbol{\theta})$

	k1	k2
1	0.92	-1.00
2	-1.00	0.13
3	-0.33	1.00
4	0.12	-1.00
5	1.00	0.08
6	0.98	0.02
7	1.00	1.00

- QMC design found under the QMC approximation with $B = 500$ to the risk (d_{QMCL2})
  – Design for $N = 10$ and $\pi_1(\boldsymbol{\theta})$

	k1	k2
1	-1.00	0.14
2	0.31	-1.00
3	1.00	-0.09
4	-0.08	1.00
5	1.00	0.90
6	-0.91	-0.74
7	1.00	-0.53
8	-0.92	1.00
9	0.25	-1.00
10	-1.00	0.22

  – Design for $N = 10$ and $\pi_2(\boldsymbol{\theta})$

	k1	k2
1	-1.00	0.001
2	0.15	-0.92
3	1.00	-0.09
4,	0.24	-0.89
5	-0.25	1.00
6	-1.00	0.34
7	1.00	-0.21
8	-0.87	-0.63
9	0.85	0.97
10	-0.30	0.99

  – Design for $N = 10$ and $\pi_3(\boldsymbol{\theta})$

	k1	k2
1	0.96	-0.09
2	-1.00	0.25
3	0.26	-1.00
4	0.27	-0.96
5	1.00	-0.43
6	0.78	0.90
7	-1.00	0.13
8	-0.91	-0.85
9	-0.33	1.00
10	-0.14	1.00

– Design for $N = 10$ and $\pi_4(\boldsymbol{\theta})$

	k1	k2
1	0.10	-0.97
2	-1.00	0.19
3	0.92	-0.13
4	-1.00	0.23
5	-0.21	1.00
6	-0.91	-1.00
7	1.00	-0.07
8	-0.45	1.00
9	0.31	-1.00
10	0.95	0.64

D.3.2 Pseudo-Bayesian decision-theoretic optimal designs under the FSL

Pseudo-Bayesian decision-theoretic optimal designs found under different methods for different experimental runs.

- True designs found under the exact risk (d_T)
  - Design for $N = 6$ and $\pi_1(\boldsymbol{\theta})$

	k1	k2
1	-0.26	-0.92
2	-0.18	0.91
3	0.83	0.34
4	-0.98	0.10
5	0.94	0.11
6	-0.32	-0.74

  - Design for $N = 6$ and $\pi_2(\boldsymbol{\theta})$

	k1	k2
1	-0.001	-1.00
2	-0.86	-0.83
3	-0.14	1.00
4	1.00	-0.06
5	-1.00	0.08
6	0.05	0.04

– Design for $N = 6$ and $\pi_3(\boldsymbol{\theta})$

	k1	k2
1	0.01	-1.00
2	-0.85	-0.81
3	-0.15	1.00
4	1.00	0.003
5	-1.00	0.13
6	-0.01	0.14

– Design for $N = 6$ and $\pi_4(\boldsymbol{\theta})$

	k1	k2
1	-0.43	-0.92
2	-1.00	-0.89
3	-0.15	1.00
4	1.00	-0.03
5	-1.00	0.14
6	-0.20	-1.00

- QMC design found under the QMC approximation with $B = 1,000$ to the risk (d_{QMCL1})
  – Design for $N = 7$ and $\pi_1(\boldsymbol{\theta})$

	k1	k2
1	-1.00	0.06
2	-0.02	1.00
3	0.15	-1.00
4	0.85	1.00
5	-0.93	0.68
6	1.00	0.07
7	0.99	0.06

– Design for $N = 7$ and $\pi_2(\boldsymbol{\theta})$

	k1	k2
1	-0.09	-1.00
2	0.75	-0.91
3	-0.23	0.93
4	1.00	0.08
5	-0.97	0.01
6	0.27	0.94
7	0.93	0.11

– Design for $N = 7$ and $\pi_3(\boldsymbol{\theta})$

	k1	k2
1	0.04	-0.61
2	0.35	1.00
3	-0.51	1.00
4	1.00	0.02
5	0.46	0.87
6	0.18	-1.00
7	-1.00	0.11

– Design for $N = 7$ and $\pi_4(\boldsymbol{\theta})$

	k1	k2
1	-1.00	-0.94
2	0.88	0.53
3	0.07	0.93
4	0.87	0.36
5	-0.97	0.06
6	0.19	-1.00
7	0.96	0.11

- QMC design found under the QMC approximation with $B = 500$ to the risk (d_{QMCL2})
  – Design for $N = 10$ and $\pi_1(\boldsymbol{\theta})$

	k1	k2
1	-1.00	-0.25
2	0.25	-1.00
3	-1.00	0.14
4	-0.92	0.09
5	0.98	-0.41
6	0.05	1.00
7	-0.70	-1.00
8	-0.70	1.00
9	-0.89	-0.56
10	1.00	0.15

– Design for $N = 10$ and $\pi_2(\boldsymbol{\theta})$

	k1	k2
1	1.00	0.04
2	1.00	0.54
3	1.00	-0.37
4	-1.00	0.07
5	-0.24	1.00
6	-0.16	-0.93
7	-1.00	0.05
8	0.20	-1.00
9	-0.26	1.00
10	-1.00	-0.68

– Design for $N = 10$ and $\pi_3(\boldsymbol{\theta})$

	k1	k2
1	1.00	-0.05
2	0.98	0.69
3	1.00	-0.35
4	-1.00	0.09
5	-0.18	1.00
6	-0.22	-0.97
7	-1.00	0.03
8	0.32	-1.00
9	-0.27	1.00
10	-0.81	-0.81

– Design for $N = 10$ and $\pi_4(\boldsymbol{\theta})$

	k1	k2
1	-1.00	-0.08
2	-0.03	-1.00
3	0.69	-1.00
4	-0.97	0.61
5	1.00	0.15
6	0.09	0.94
7	1.00	-0.14
8	0.92	0.88
9	-0.03	1.00
10	0.93	0.45

Bibliography

Abdelbasit, K. M. and Plackett, R. (1983). Experimental design for binary data. *Journal of the American Statistical Association*, 78(381):90–98.

Akaike, H. (1974). A new look at the statistical model identification. In *Selected Papers of Hirotugu Akaike*, pages 215–222. Springer.

Atkinson, A. (1996). The usefulness of optimum experimental designs. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 59–76.

Atkinson, A., Demetrio, C., and Zocchi, S. (1995). Optimum dose levels when males and females differ in response. *Applied Statistics*, pages 213–226.

Atkinson, A. C., Chaloner, K., Herzberg, A. M., and Juritz, J. (1993). Optimum experimental designs for properties of a compartmental model. *Biometrics*, pages 325–337.

Atkinson, A. C., Donev, A. N., and Tobias, R. D. (2007). *Optimum experimental designs, with SAS*, volume 34. Oxford University Press Oxford.

Atkinson, A. C., Fedorov, V. V., Herzberg, A. M., and Zhang, R. (2012). Optimal experimental design for generalized regression models. *JSPI*, 144:81–91.

Biedermann, S. and Woods, D. C. (2011). Optimal designs for generalized non-linear models with application to second-harmonic generation experiments. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 60(2):281–299.

Brimkulov, U., Krug, G., and Savanov, V. (1980). Numerical construction of exact experimental designs when the measurements are correlated. *Zavodskaya Laboratoria (Industrial Laboratory)*, 36:435–442.

Casella, G. and Berger, R. L. (2002). *Statistical inference*, volume 2. Duxbury Pacific Grove, CA.

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

Chaloner, K. and Verdinelli, I. (1995). Bayesian experimental design: A review. *Statistical Science*, pages 273–304.

Chen, R.-B., Chang, S.-P., Wang, W., Tung, H.-C., and Wong, W. K. (2013). Optimal minimax designs via particle swarm optimization methods. *Preprints of the Isaac Newton Institute*.

Chernoff, H. (1953). Locally optimal designs for estimating parameters. *The Annals of Mathematical Statistics*, pages 586–602.

Dean, A. M. and Voss, D. (2008). *Design and analysis of experiments*. Springer, New York.

Dette, H., Biedermann, S., and Zhu, W. (2005). Geometric construction of optimal designs for dose-responsemodels with two parameters. Technical report, Technical Report.

Dette, H. and Haines, L. M. (1994). E-optimal designs for linear and nonlinear models with two parameters. *Biometrika*, 81(4):739–754.

Dobson, A. J. and Barnett, A. (2008). *An introduction to generalized linear models*. CRC press.

Dror, H. A. and Steinberg, D. M. (2005). Approximate local d&optimal experimental design for binary.

Dror, H. A. and Steinberg, D. M. (2006). Robust experimental design for multivariate generalized linear models. *Technometrics*, 48(4):520–529.

Dunteman, G. H. and Ho, M.-H. R. (2005). *An introduction to generalized linear models*, volume 145. Sage Publications.

Eddelbuettel, D. and François, R. (2011). Rcpp: Seamless R and C++ integration. *Journal of Statistical Software*, 40(8):1–18.

Eddelbuettel, D. and Sanderson, C. (2014). Rcpparmadillo: Accelerating r with high-performance c++ linear algebra. *Computational Statistics and Data Analysis*, 71:1054–1063.

Fedorov, V. (1972). *Theory of optimal experiments*. New York: Academic Press.

Firth, D. (1992). Generalized linear models and jeffreys priors: an iterative weighted least-squares approach. In *Computational statistics*, pages 553–557. Springer.

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, 80(1):27–38.

Ford, I., Torsney, B., and Wu, C. J. (1992). The use of a canonical form in the construction of locally optimal designs for non-linear problems. *Journal of the Royal Statistical Society: Series B (Methodological)*, 54(2):569–583.

Garthwaite, P. H., Jolliffe, I. T., and Jones, B. (2002). *Statistical inference*. Oxford University Press on Demand.

Gonzalez Estrada, E. and Villasenor Alva, J. (2009). mvshapirotest: generalized shapiro–wilk test for multivariate normality. *R package version 0.0*, 1.

Goos, P. (2012). *The optimal design of blocked and split-plot experiments*, volume 164. Springer Science & Business Media.

Gotwalt, C. M., Jones, B. A., and Steinberg, D. M. (2009). Fast computation of designs robust to parameter uncertainty for nonlinear settings. *Technometrics*, 51(1):88–95.

Harman, R. and Štulajter, F. (2010). Optimal prediction designs in finite discrete spectrum linear regression models. *Metrika*, 72(2):281–294.

Hedayat, A., Yan, B., and Pezzuto, J. M. (1997). Modeling and identifying optimum designs for fitting dose-response curves based on raw optical density data. *Journal of the American Statistical Association*, 92(439):1132–1140.

Johnson, M. E. and Nachtsheim, C. J. (1983). Some guidelines for constructing exact d-optimal designs on convex design spaces. *Technometrics*, 25(3):271–277.

Kiefer, J. (1959). Optimum experimental designs. *Journal of the Royal Statistical Society: Series B (Methodological)*, 21(2):272–304.

Krantz, S. G. and Parks, H. R. (2012). *The implicit function theorem: history, theory, and applications*. Springer Science & Business Media.

Lin, D. C. and Tang, B. (2015). Latin hypercube and space-filling designs. In Dean, A., Morris, M., Stufken, J., and Bingham, D., editors, *Handbook of Design and Analysis of Experiments*. Boca Raton: CRC Press, New York.

Mathai, A. M. and Provost, S. B. (1992). *Quadratic forms in random variables: theory and applications*. M. Dekker New York.

Mathew, T. and Sinha, B. K. (2001). Optimal designs for binary data under logistic regression. *Journal of Statistical Planning and Inference*, 93(1):295–307.

McCullagh, P. (1984). Generalized linear models. *European Journal of Operational Research*, 16(3):285–292.

McGree, J. M. and Eccleston, J. A. (2012). Robust designs for poisson regression models. *Technometrics*, 54(1):64–72.

Meyer, R. K. and Nachtsheim, C. J. (1995). The coordinate-exchange algorithm for constructing exact optimal experimental designs. *Technometrics*, 37(1):60–69.

Monahan, J. and Genz, A. (1997). Spherical-radial integration rules for bayesian computation. *Journal of the American Statistical Association*, 92(438):664–674.

Müller, W. G. and Pázman, A. (1999). An algorithm for the computation of optimum designs under a given covariance structure. *Computational Statistics*, 14(2):197–211.

Müller, W. G. and Pázman, A. (2003). Measures for designs in experiments with correlated errors. *Biometrika*, 90(2):423–434.

Myers, W. R., Myers, R. H., and Carter, W. H. (1994). Some alphabetic optimal designs for the logistic regression model. *Journal of Statistical Planning and Inference*, 42(1-2):57–77.

Mysovskikh, I. (1980). The approximation of multiple integrals by using interpolatory cubature formulae. In *Quantitative approximation*, pages 217–243. Elsevier.

Nelder, J. A. and Baker, R. J. (1972). Generalized linear models. *Encyclopedia of Statistical Sciences*.

Nguyen, N.-K. and Miller, A. J. (1992). A review of some exchange algorithms for constructing discrete d-optimal designs. *Computational Statistics & Data Analysis*, 14(4):489–498.

Niaparast, M. (2009). On optimal design for a poisson regression model with random intercept. *Statistics & Probability Letters*, 79(6):741–747.

Overstall, A., Woods, D., and Adamou, M. (2016). acebayes: Optimal bayesian experimental design using the ace algorithm. *URL https://cran. r-project. org/web/packages/acebayes/index. html. R package version*, 1.

Overstall, A., Woods, D., and Adamou, M. (2017a). acebayes: An r package for bayesian optimal design of experiments via approximate coordinate exchange. *arXiv preprint arXiv:1705.08096*.

Overstall, A. M., McGree, J. M., and Drovandi, C. C. (2017b). An approach for finding fully bayesian optimal designs using normal-based approximations to loss functions. *Statistics and Computing*, pages 1–16.

Overstall, A. M. and Woods, D. C. (2017). Bayesian design of experiments using approximate coordinate exchange. *Technometrics*, pages 1–13.

Pázman, A. (1986). *Foundations of optimum experimental design*, volume 14. Springer.

Pronzato, L. and Pázman, A. (2013). Design of experiments in nonlinear models. *Lecture notes in statistics*, 212.

Pukelsheim, F. (2006). *Optimal design of experiments*. SIAM.

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Robert, C. P. (2004). *Monte carlo methods*. Wiley Online Library.

Russell, K. G., Eccleston, J., Lewis, S. M., and Woods, D. C. (2009a). Design considerations for small experiments and simple logistic regression. *Journal of Statistical Computation and Simulation*, 79(1):81–91.

Russell, K. G., Woods, D. C., Lewis, S., and Eccleston, J. (2009b). D-optimal designs for poisson regression models. *Statistica Sinica*, pages 721–730.

Ryan, T. P. and Morgan, J. (2007). Modern experimental design. *Journal of Statistical Theory and Practice*, 1(3-4):501–506.

Silvey, S. (2013). *Optimal design: an introduction to the theory for parameter estimation*, volume 1. Springer Science & Business Media.

Van De Ven, P. and Woods, D. (2014). Optimal blocked minimum-support designs for non-linear models. *Journal of Statistical Planning and Inference*, 144:152–159.

Wang, Y., Myers, R. H., Smith, E. P., and Ye, K. (2006). D-optimal designs for poisson regression models. *Journal of statistical planning and inference*, 136(8):2831–2845.

Welch, W. J. (1983). A mean squared error criterion for the design of experiments. *Biometrika*, 70(1):205–213.

Woods, D. C., Lewis, S. M., Eccleston, J. A., and Russell, K. (2006). Designs for generalized linear models with several variables and model uncertainty. *Technometrics*, 48(2):284–292.

Yang, M., Stufken, J., et al. (2012). Identifying locally optimal designs for nonlinear models: A simple extension with profound consequences. *The Annals of Statistics*, 40(3):1665–1681.