A Bayes Decision Theoretic Approach to the Optimal Design of Screens

David J Laws, BSc.

Thesis submitted to the University of Newcastle upon Tyne
for the degree of Doctor of Philosophy

September 29, 1997

Acknowledgements

My thanks go to E.P.S.R.C. for funding me to carry out this research, my supervisor Richard Boys and mentor Kevin Glazebrook. Thanks also to Michael White, Geoffrey Horrocks, Tony O'Hagan and Frank Ball for their help.

Many thanks to my employers at the University of Nottingham, for their patience and belief that I'd finish this, and to all my friends who kept me sane and drunk at all the right times, I think.

Abstract

An item may be said to reach a standard suitable for use if it has some prescribed attributes. Suppose that a variable $\underline{T}$ measures the standard and $\underline{T} \in C_{\underline{T}}$ if an item has the desired attributes. The variable $\underline{T}$ may be very expensive to measure and so, some cheaper to measure screening variables, $\underline{X}$ say, correlated to $\underline{T}$ may be used to classify items. The purpose of screen design is to determine $C_{\underline{X}}$, the region of $\underline{X}$ space, for which an item should be said to reach the standard.

If the error probabilities of classifying an item based on $\underline{X}$ are very high it may be economical to measure $\underline{T}$. Chapter 2 deals with this idea in the context of a very simple two–stage set-up in which, at the first stage of the screen a univariate screening variable X is measured. Some items are sentenced as acceptable or unacceptable, and the remainder are passed on to the second stage at which $\underline{T}$ is determined. The optimal screen is found that minimises cost, where costs are given for misclassifying items and for measuring the variables. The variable T is assumed binary and the model for $T|X$ is a probit regression model.

In designing a two–stage screen, Chapter 3 considers: (a) a general stochastic structure for $(\underline{T}, \underline{X})$, (b) a general loss function set up for misclassification costs and (c) assumes no fixed form for the screen. Also in Chapter 3, we consider a scenario in which a statistical goal or constraint is imposed in addition to the decision–theoretic target of minimising expected cost.

In Chapter 4 we consider a sequential screen that operates as follows. At each stage of a sequence a covariate is measured and items may be accepted as suitable, discarded or passed on to the next stage. At the final stage the performance variable T is measured.

Returning to the simple one–stage screen based solely on measuring covariates, Chapter 5 poses the question of how many and which covariates to include as part of the screen.

Contents

Chapter 1

Screening

In quality control, screening is the action of inspecting all items to check quality before shipment. With the development of automated manufacturing systems, such inspection has become more cost–effective and A.S.Q.C. (1987) reports that, at that time, about 85% of organisations used screening in their quality control procedures. Tang & Tang (1994) give a review of screening procedures in this context. Screening is also used in many other contexts, for example, medicine (Gastwirth (1987) and Geisser & Johnson (1992)), education (Thomas *et al.* (1977)) and in personnel selection, where employers often select only those candidates who score above a given level on a predetermined scoring scheme. In quality control (and analogously in other applications), the quality of the item, or the presence of attributes the item must have for it to be useful, can be described by a univariate or multivariate performance variable $\underline{T}$. The item has the necessary attributes if $\underline{T}$ takes any of a known set of values, say $\underline{T} \in C_{\underline{T}}$, where $C_{\underline{T}} \subseteq \Omega_{\underline{T}}$, and $\Omega_{\underline{T}}$ is the sample space of $\underline{T}$. For example, for an electronic device to perform satisfactorily it may be the case that the voltage at an internal point must be within a known range of values. Here the purpose of screening is to assess whether the voltage, which is the performance variable, falls within the required range. If it is easy and inexpensive to do so, a screening procedure should simply measure the performance variable and discard items if appropriate. However, in many cases it may be expensive or destructive to measure $\underline{T}$, and so the assessment

of the suitability of an item is based on the measurement of some cheap to measure variables $\underline{X}$ that are correlated to $\underline{T}$. Classification using correlated variables will be cheaper but prone to error and so we have a trade–off between accuracy and cost. To measure the voltage at an internal point in an electronic device will involve disassembling the device and so instead, to assess whether the item might fail, one might measure the voltage at an external point on the device which is easier to measure and strongly correlated to the voltage at the internal point. The correlated variables are called *screening variables* and we denote as $C_{\underline{X}}$, the set of values of $\underline{x}$ for which an item is passed as acceptable by the screen. Many screens are based on correlated variables only, and here screen design is concerned with finding an appropriate $C_{\underline{X}}$. Another set–up that involves measuring $\underline{T}$ on some items will be described in section 1.2 and is the focus of Chapters 2 & 3.

Three main factors are considered in screen design:

Objective or criteria. Two types of objective have been proposed as the purpose of a screen. One type considers statistical goals, that is, items accepted and rejected by the screen should attain goals in terms of $P(\underline{T} \in C_{\underline{T}})$ and/or $P(\underline{T} \notin C_{\underline{T}})$. For example, an item accepted by the screen should have a high probability of being suitable, or the rate of conforming items accepted by the screen should be high. The other type of objective follows a decision theoretic (economic) approach and sets a target of minimising cost for screen design. These two types of objective are, in some sense, interchangeable. A decision–theoretic set–up is implicit in the choice of statistical goal and a statistical goal will be explicit in the solution to the decision–theoretic objective. It may be the case that one wants to explicitly set both economic and statistical objectives. For example, in section 3.4, we design a screen to minimise expected cost and to satisfy a constraint on the proportion of items left unclassified by the part of the screen based on correlated variables. However, we see that imposing the constraint is essentially equivalent to increasing one of the cost parameters.

Structure for $(\underline{T}, \underline{X})$. The components of the performance variable can be dis-

crete or continuous. In the discrete case, the component is typically a binary variable denoting whether or not a desired attribute is present. Continuous components are often divided into three types: the–nominal-the-best, the–smaller-the-better or the-larger-the-better. For the purpose of analysis it is simple to convert the former two types into the latter type. A nominal–the–best variable, by taking the inverse of the distance from the desired nominal value and, a smaller–the–better variable, by multiplying by -1. Such considerations should be part of the model selection process for $(\underline{T}, \underline{X})$. The covariates that are used in the screening variable $\underline{X}$ can also be discrete or continuous and it may be necessary to transform the variable in some way to suit the assumptions behind a chosen model or design.

The backbone of screen design is the model for $(\underline{T}, \underline{X})$. We denote the joint probability model for $(\underline{T}, \underline{X})$ as $p(\underline{t}, \underline{x}|\underline{\theta})$, where $\underline{\theta}$ is a vector of parameters. We learn about the relationship between $\underline{T}$ and $\underline{X}$ through the parameters $\underline{\theta}$ but, in a sense, they are nuisance variables and we wish to deal with the unconditional (predictive) model $p(\underline{t}, \underline{x})$. There are three main approaches for obtaining such information,

(i) Parameters known. In some rare cases the parameters of the probability model may be known, for example, they may be a function of some quantities in the manufacturing system. Here we just write

$$p(\underline{t}, \underline{x}) = p(\underline{t}, \underline{x}|\underline{\theta} = \tilde{\underline{\theta}}),$$

where $\tilde{\underline{\theta}}$ is the known value of the parameters $\underline{\theta}$. Owen, McIntyre & Seymour (1975) develop a screening procedure assuming parameters are known. Clearly, when there is any uncertainty about the value of the parameters one should construct a distribution to model the beliefs about their value (see (iii) below).

(ii) Estimative approach. Suppose we have available training data in the form of a random sample $(\underline{t}_1, \underline{x}_1), (\underline{t}_2, \underline{x}_2), \ldots, (\underline{t}_n, \underline{x}_n)$ from the unscreened population. One approach is to use the sample to obtain an estimate or

confidence interval for $\underline{\theta}$. The unconditional model may then be estimated by

$$p(\underline{t}, \underline{x}) = p(\underline{t}, \underline{x}|\underline{\theta} = \hat{\underline{\theta}}),$$

where $\hat{\underline{\theta}}$ is the estimate of $\underline{\theta}$. Given a confidence interval for $\underline{\theta}$, intervals for various probabilities determined by $p(\underline{t}, \underline{x})$ could also be given. Such an approach is adopted in the context of screening by Owen & Boddie (1976) and Owen & Su (1977). Their methods generally involve tolerance region analysis.

(iii) Predictive approach. Taking a Bayesian approach, knowledge about the parameters $\underline{\theta}$ is summarised by a probability distribution $\pi(\underline{\theta})$. The distribution $\pi(\underline{\theta})$ is based on a training sample on $(\underline{T}, \underline{X})$ and/or other (prior) information about $\underline{T}$ and $\underline{X}$. For example, in the case of a manufacturing process, knowledge about the interaction of the process parameters can be built into the distribution $\pi(\underline{\theta})$. Now a predictive approach can be taken for the model in which we average over the uncertainty about the parameters $\underline{\theta}$,

$$p(\underline{t}, \underline{x}) = E_{\underline{\theta}}\left[p(\underline{t}, \underline{x}|\underline{\theta})\right].$$

We advocate the third approach as it is realistic, unlike the first approach, and takes full account of the uncertainty about the parameters within the coherent framework of a Bayesian analysis. In the remainder of this chapter we will focus on work that takes the third approach, referencing other works only when they present ideas yet to be covered by a predictive approach.

In their review of screening procedures, Tang & Tang (1994) discuss the possibility of inspection error in the measurement of variables. For the performance variable, a measurement with error can simply be considered as another correlated screening variable, see Tang & Schneider (1990). A screening variable with inspection error is simply another screening variable, with larger variance. This does lead to the question of how to obtain a training sample on the performance variable if it is always measured with error or is measured via destructive

testing. In the latter case, in a quality control context, the training sample may have been taken either (i) before or as the item is constructed, or (ii) once the item has been seen to work well or fail. However, (ii) can lead to the problem of verification bias, see Greenes & Begg (1985) for an explanation in the context a medical application. In such problems, the training sample is not a random sample from the population but is taken from a collection of cases in which $(\underline{t}, \underline{x})$ has been observed. However, the reason why $(\underline{t}, \underline{x})$ was observed may not be independent of $(\underline{T}, \underline{X})$ and so, if the circumstances under which the the sample is taken are not built into the model correctly, the inferences made about the population from which the sample is taken may be incorrect. For example, care must be taken to ensure that observed cases have no common factor that affects $(\underline{T}, \underline{X})$ which will not generally be present in the population under consideration. While these issues are important they are beyond the scope of this thesis.

Logistics. A number of practical considerations need to be taken into account in screen design. Is it possible or economic to measure the performance variable in any part of the screen? Are there any constraints on the measurement of screening variables? Is it appropriate to design a screen in which each screening variable is measured in turn with some items discarded at each stage? The decision–theoretic approach must consider what costs should be imposed and how they relate to each other. We consider some of these topics as the thesis progresses, for further discussion see Tang & Tang (1994).

1.1 Designing screens with statistical goals

In this section we review literature in which the objective of screen design is to find the screening region $C_{\underline{X}}$ that satisfies some statistical criteria.

1.1.1 Global criteria

A common motivation given for screening is to discard items so that those that remain contain a prespecified high proportion of conforming (good) items. That is, if we denote $\gamma = P(\underline{T} \in C_{\underline{T}})$ as the pre-screening conforming rate, then the purpose of screen design is to find $C_{\underline{X}}$, the values of $\underline{x}$ for which the screen accepts an item, so that

$$P(\underline{T} \in C_{\underline{T}} | \underline{X} \in C_{\underline{X}}) = \delta, \tag{1.1}$$

where δ is some specified value with $\delta > \gamma$.

How to attain this goal and whether it is attainable will depend on the performance variable, the specification region $C_{\underline{T}}$ and the correlation between the screening and performance variables. When (1.1) is achievable, there may be multiple solutions for $C_{\underline{X}}$ and, in such cases, it is suggested that the best of these solutions is the one that minimises

$$\epsilon = P(\underline{T} \in C_{\underline{T}} | \underline{X} \notin C_{\underline{X}}), \tag{1.2}$$

that is, the rate of good items that are screened out.

Normal models

Boys & Dunsmore (1986) provide a Bayesian approach to the screening problem based on predictive probabilities. They take the case in which T and X are univariate and consider specification regions C_T of the form $[\ell, \infty)$, $(-\infty, u]$ and $[\ell, u]$ with ℓ and/or u known in advance. For each case the authors take the natural form for the screen when X is positively correlated to T. That is, for $C_T = [\ell, \infty)$ they assume that $C_X = [w, \infty)$ and for $C_T = [\ell, u]$ they assume that $C_X = [v, w]$. Situations in which $C_T = (-\infty, u]$ can be transformed to the case with $C_T = [\ell, \infty)$ by multiplying the performance variable by -1. The target of screen design is then to find the screening parameters w or $(v, w)^T$ that satisfy the screening criteria based on (1.1) and (1.2).

Boys & Dunsmore describe a procedure to find the screening parameters in the case of a bivariate normal model for $p(t, x|\underline{\theta})$ with a conjugate prior distribution for $\underline{\theta}$.

When $\underline{X}$ is multivariate and has $p(>1)$ components $\underline{X} = (X_1, X_2, \ldots, X_p)^T$, the selection of a 'natural' form for the screening region $C_{\underline{X}}$ is not so straightforward. In an unpublished paper, Boys & Dunsmore suggest three forms for $C_{\underline{X}}$,

(i) a linear discriminant in which

$$C_{\underline{X}} = \left\{\underline{x} : \underline{a}^T\underline{x} \geq w\right\}, \tag{1.3}$$

with the p–dimensional vector $\underline{a}$, $\underline{a}^T\underline{a} = 1$, and the constant w as the parameters of the screen design,

(ii) a quadratic discriminant with

$$C_{\underline{X}} = \left\{\underline{x} : \underline{x}^T Q\underline{x} \geq w\right\},$$

where Q is a $p \times p$ matrix, and

(iii) a region made up of separate regions for each component of $\underline{X}$,

$$C_{\underline{X}} = C_{X_1} \cap C_{X_2} \cap \ldots \cap C_{X_p},$$

with each C_{X_i} assumed to be of the form $[w_i, \infty)$.

Boys & Dunsmore investigate the construction of such screening regions in the context of a multivariate normal model for $(T, \underline{X})$.

Binary responses

As the usefulness of an item is defined by a simple dichotomy on the performance variable, $\underline{T} \in C_{\underline{T}}$ for a good item and $\underline{T} \notin C_{\underline{T}}$ otherwise, an alternative approach to a full model for $(\underline{T}, \underline{X})$ is to model the relationship between the dichotomy and $\underline{X}$. In other words, if we define a new univariate performance variable T as $T = 1$ if $\underline{T} \in C_{\underline{T}}$ and $T = 0$ if $\underline{T} \notin C_{\underline{T}}$, we model $(T, \underline{X})$. This is a particularly useful simplification in

situations in which there is a complex multivariate performance variable that might pose difficult modelling problems. Also, such a simplification imposes less structure than the full normality of Boys & Dunsmore (1986). However, care should be taken that the advantages of a simple form for T are not outweighed by a loss in performance of the screen. The full model for $(\underline{T}, \underline{X})$ may be a much more accurate model for prediction purposes than a model relating a binary T and $\underline{X}$. Also, when dealing with a decision–theoretic approach, care should be taken that any simplified cost structure is consistent with the full cost structure for $(\underline{T}, \underline{X})$.

With T as a binary variable, the construction of screening regions is equivalent to designing a discrimination or classification rule with two groups. Hence a large amount of discrimination theory can now be applied to the screening problem. In such a set up, Dawid (1976) and Aitchison & Begg (1976) have distinguished two modelling approaches that factorise the joint model $p(t, \underline{x}|\underline{\theta})$. One is the diagnostic paradigm, which models the relationship between T and $\underline{X}$ through the conditional model for T given $\underline{X}$, with $p(t, \underline{x}|\underline{\theta}) = p(t|\underline{x}, \underline{\xi})p(\underline{x}|\underline{\phi})$ where $\underline{\theta} = (\underline{\xi}, \underline{\phi})^T$. Notice that the parameters $\underline{\xi}$ and $\underline{\phi}$ are independent in the likelihood and so if they are also independent *a priori*, the parameters will be independent *a posteriori*. Under such circumstances the predictive models $p(t|\underline{x})$ and $p(\underline{x})$ can be obtained separately. The second approach follows the sampling paradigm which factorises $p(t, \underline{x}|\theta)$ as $p(\underline{x}|t, \underline{\eta})p(t|\underline{\psi})$, with $\underline{\theta} = (\underline{\eta}, \underline{\psi})^T$. Similar comments apply about the independence of the parameters and consideration of the predictive models. Dawid (1976) shows that diagnostic models are robust to verification bias. Connections have been made between logistic models for $p(t|\underline{x}, \underline{\xi})$ and a number of models for $p(\underline{x}|t, \underline{\eta})$. For example, if $p(\underline{x}|t = i, \underline{\eta})$, $i = 0, 1$, are multivariate normal with identical variance matrices, and $p(t|\underline{\psi})$ is Bernoulli, then the diagnostic model for $p(t|\underline{x}, \underline{\xi})$ is linear logistic. A brief proof follows.

Suppose that $\underline{X}|T=i$ is $N_p\left(\underline{\mu}_i, \Sigma\right)$, for $i=0,1$. Also suppose that T is Bernoulli with $P(T=1|\psi)=\psi$. Then

$$
\begin{aligned}
&P(T=1|\underline{x},\underline{\mu}_0,\underline{\mu}_1,\Sigma,\psi) \\
&\quad= \frac{p\left(\underline{x}|T=1,\underline{\mu}_1,\Sigma\right)P(T=1|\psi)}{p\left(\underline{x}|T=1,\underline{\mu}_1,\Sigma\right)P(T=1|\psi)+p\left(\underline{x}|T=0,\underline{\mu}_0,\Sigma\right)P(T=0|\psi)} \\
&\quad= \frac{\exp\left\{-\left(\underline{x}-\underline{\mu}_1\right)^T\Sigma^{-1}\left(\underline{x}-\underline{\mu}_1\right)/2\right\}\psi}{\exp\left\{-\left(\underline{x}-\underline{\mu}_1\right)^T\Sigma^{-1}\left(\underline{x}-\underline{\mu}_1\right)/2\right\}\psi+\exp\left\{-\left(\underline{x}-\underline{\mu}_0\right)^T\Sigma^{-1}\left(\underline{x}-\underline{\mu}_0\right)/2\right\}(1-\psi)} \\
&\quad= \frac{\exp\left\{\underline{\mu}_1^T\Sigma^{-1}\underline{x}-\underline{\mu}_1^T\Sigma^{-1}\underline{\mu}_1/2\right\}\psi}{\exp\left\{\underline{\mu}_1^T\Sigma^{-1}\underline{x}-\underline{\mu}_1^T\Sigma^{-1}\underline{\mu}_1/2\right\}\psi+\exp\left\{\underline{\mu}_0^T\Sigma^{-1}\underline{x}-\underline{\mu}_0^T\Sigma^{-1}\underline{\mu}_0/2\right\}(1-\psi)}.
\end{aligned}
$$

Now reparameterise as follows,

$$
\begin{aligned}
\alpha_0 &= -\underline{\mu}_0^T\Sigma^{-1}\underline{\mu}_0/2+\log(1-\psi), \\
\alpha_1 &= -\underline{\mu}_1^T\Sigma^{-1}\underline{\mu}_1/2+\log(\psi), \\
\underline{\beta}_0^T &= \underline{\mu}_0^T\Sigma^{-1}, \\
\underline{\beta}_1^T &= \underline{\mu}_1^T\Sigma^{-1}.
\end{aligned}
$$

This results in

$$
\begin{aligned}
P(T=1|\underline{x},\alpha_0,\alpha_1,\underline{\beta}_0,\underline{\beta}_1) &= \frac{\exp\left(\alpha_1+\underline{\beta}_1^T\underline{x}\right)}{\exp\left(\alpha_1+\underline{\beta}_1^T\underline{x}\right)+\exp\left(\alpha_0+\underline{\beta}_0^T\underline{x}\right)} \\
&= \frac{\exp\left\{\alpha_1-\alpha_0+(\underline{\beta}_1-\underline{\beta}_0)^T\underline{x}\right\}}{\exp\left\{\alpha_1-\alpha_0+(\underline{\beta}_1-\underline{\beta}_0)^T\underline{x}\right\}+1}.
\end{aligned}
$$

Reparameterising once more clearly shows that the above is a linear logistic model,

$$
P(T=1|\underline{x},\xi_0,\underline{\xi})=\frac{\exp\left(\xi_0+\underline{\xi}^T\underline{x}\right)}{1+\exp\left(\xi_0+\underline{\xi}^T\underline{x}\right)},
$$

where $\xi_0=\alpha_1-\alpha_0$ and $\underline{\xi}=\underline{\beta}_1-\underline{\beta}_0$. □

It has also been shown that Normal models for $\underline{X}|T,\eta$ with different variance matrices result in a quadratic logistic model for $T|\underline{X},\xi$.

The choice of paradigm often depends on the context of the problem and it is important to distinguish the different sampling schemes that may give rise to the

sample on $(T, \underline{X})$ as they provide different information. Anderson & Blair (1982) distinguish the sampling schemes as follows.

(i) *Mixture sampling.* In an observational study (also known as a natural experiment) all the sample values t_i and $\underline{x}_i$ are undetermined prior to the study and so they make up a random sample on $(T, \underline{X})$. Anderson & Blair term this mixture sampling. In both the diagnostic and sampling paradigms the data can be used to produce posterior updates for all parameters, as we have information on $T|\underline{X}$, $\underline{X}$, $\underline{X}|T$ and T.

(ii) *X-conditional sampling.* In a designed experiment, observations of T might be taken at fixed, predetermined values $\underline{x}_i$. This is termed x–conditional sampling. Here we only learn about the diagnostic conditional model for $T|\underline{X}$ and the sample values can be used to produce posterior updates for the parameters $\underline{\xi}$ in the model $p(t|\underline{x}, \underline{\xi})$. The data can not easily be used to update any other parameters in either paradigm. In particular, nothing is learned about the marginal distribution of $\underline{X}$ and so a separate random sample on $\underline{X}$ may be taken to update the parameters $\underline{\phi}$.

(iii) *T-conditional sampling.* An alternative form of designed experiment fixes the number of observations of $\underline{X}$ taken at each value of t, n_1 measurements are taken with $t = 1$ and n_0 measurements with $t = 0$, where n_0 and n_1 are chosen in advance and $n = n_0 + n_1$. This t–conditional sampling observes $\underline{X}|T$ and so the data can be used to obtain posterior updates of the parameters $\underline{\eta}$ in the conditional model $p(\underline{x}|t, \underline{\eta})$. A separate random sample on T might be taken to learn about the parameters $\underline{\psi}$.

It is plain from the above discussion that the sampling scheme should be taken into account when determining the way in which the model $p(t, \underline{x}|\underline{\theta})$ is factorised. Often the choice of sampling scheme is driven by the context of the problem.

In Boys & Dunsmore (1987) (hereafter B & D) and Dunsmore & Boys (1987) (hereafter D & B), the authors take the case in which T is binary, modelling the

dichotomy that defines whether the item is acceptable. They discuss the construction of screen designs under both a diagnostic approach and a sampling approach.

For the diagnostic approach, the authors assume a logistic model for $T|\underline{X}, \underline{\xi}$ with

$$P(T = 1|\underline{x}, \underline{\xi}) = e^{\underline{\xi}^T \underline{x}_0} \Big/ (1 + e^{\underline{\xi}^T \underline{x}_0}),$$

where $\underline{x}_0 = (1, \underline{x})^T$. D & B assume a multivariate normal model for $p(\underline{x}|\underline{\phi})$ and, in the context of a medical example, find the screening region $C_{\underline{X}}$ of the form (1.3) that satisfies the screening criteria based on (1.1) and (1.2). As an alternative that reduces computation, both B & D and D & B propose dimensionality reduction schemes based on fixing $\underline{a}$ in (1.3) prior to screen design. For example, $\underline{a}^T \underline{X}$ could be chosen as (i) a principal component, (ii) the first crimcoord (Gnanadesikan (1977), p.86) or (iii) Fisher's linear discriminant function. In addition, D & B suggest, for the case of a large training sample on $(T, \underline{X})$, setting $\underline{a}$ as the maximum likelihood estimate of $\underline{\xi}$ in the logistic model for $T|\underline{X}, \underline{\xi}$. With $\underline{a}$ fixed and no longer a parameter of screen design, one can model in terms of the variable $D = \underline{a}^T \underline{X}$ rather than $\underline{X}$ and hence reduce dimensionality. For large data sets D can be assumed to be approximately normal and so a normal model is proposed for $p(d|\underline{\phi})$ with D univariate. Also, if $\underline{X}$ is multivariate normal, then D is normal. The authors discuss finding an appropriate screening region C_D of the form $\{d : d \geq w\}$, where w is the parameter that characterises screen design.

In the sampling approach both papers suppose a normal model for $p(\underline{x}|t = i, \underline{\eta})$, $i = 1, 0$ and a binomial model for the number of successes $(T = 1)$ in the training sample. D & B again assume a screening region of the form (1.3) and, for their example, obtain the region of that form that satisfies the global screening criteria. In D & B, the authors once more propose dimensionality reduction schemes and obtain screening regions of the form $\{d : d \geq w\}$ for $D = \underline{a}^T \underline{X}$, where $\underline{a}$ is given by Fisher's linear discriminant. B & D take the case in which X is univariate (supposing that any dimensionality reduction has already taken place) and obtain C_X that has the form $[w, \infty)$, the natural form of the screening region if X is positively correlated to T.

Nonparametric methods

A nonparametric approach to the screening problem is proposed by Boys (1992). With T binary and X univariate, a kernel estimation method (Copas (1983)) is described for calculating the probability in (1.1) and, with $C_X = [w, \infty)$, a method for finding w is given. Another approach suggested is to smooth the empirical estimate of (1.1) which is given by

$$\frac{\text{number of sample cases with } t_i = 1 \text{ and } x_i \geq w}{\text{number of sample cases with } x_i \geq w}.$$

A note of caution

It may not always be possible to find a $C_{\underline{X}}$ that satisfies (1.1) and Liu (1992) gives an upper bound on achievable values of δ. Liu also considers screening under the bivariate normal model of Boys & Dunsmore (1986) and shows that, when X and T are positively correlated, the screening region that satisfies (1.1) and minimises (1.2) may not be of the form $C_X = [w, \infty)$ when C_T is of the form $= [\ell, \infty)$.

1.1.2 Local screening

So far we have discussed the construction of screens designed to satisfy (1.1) with any indeterminacy avoided by minimising the error probability in (1.2). Suppose that one of the items passed by the screen is picked at random. What is the probability that the item is acceptable given all that we know about it? We know that it is an item that has been passed by the screen and so we might say that the probability that it is suitable is δ in (1.1). However, after screening we also know $\underline{x}$ for each item and so we can calculate (or approximate) $P(\underline{T} \in C_{\underline{T}} | \underline{x})$. This probability may be greater than or less than δ for an item passed by the screen. $P(\underline{T} \in C_{\underline{T}} | \underline{x})$ is a local statement about each individual item given the measurement of $\underline{X}$ for that item. $P(\underline{T} \in C_{\underline{T}} | \underline{X} \in C_{\underline{X}})$ is a global statement about all the items that are passed by the screen. The global statement averages $P(T \in C_{\underline{T}} | \underline{X})$ over the conditional distribution of $\underline{X} | \underline{X} \in C_{\underline{X}}$.

Wong, Meeker & Selwyn (1985) and Dunsmore & Boys (1987,1988) suggest a local alternative to the global criterion given by (1.1). The screening region $C_{\underline{X}}$ is required so that

$$P(\underline{T} \in C_{\underline{T}}|\underline{x}) \begin{cases} \geq \delta_L & \text{for } \underline{x} \in C_{\underline{X}}, \\ < \delta_L & \text{for } \underline{x} \notin C_{\underline{X}}, \end{cases} \tag{1.4}$$

for some specified δ_L. This criterion ensures that each item passed by the screen has a suitably high probability of being acceptable. Note that choosing $\delta_L = \delta$ is unlikely to give the same screening region $C_{\underline{X}}$ for the global and local screens. In fact, a screening region that satisfies (1.4) will give $P(\underline{T} \in C_{\underline{T}}|\underline{X} \in C_{\underline{X}})$ at least as big as δ_L, as under (1.4),

$$\begin{aligned} P(\underline{T} \in C_{\underline{T}}|\underline{X} \in C_{\underline{X}}) &= \frac{E_{\underline{X}}\left\{P(\underline{T} \in C_{\underline{T}}|\underline{X})I(\underline{X} \in C_{\underline{X}})\right\}}{P(\underline{X} \in C_{\underline{X}})} \\ &\geq \frac{E_{\underline{X}}\left\{\delta_L I(\underline{X} \in C_{\underline{X}})\right\}}{P(\underline{X} \in C_{\underline{X}})} \\ &= \delta_L, \end{aligned}$$

where

$$I(\underline{X} \in C_{\underline{X}}) = \begin{cases} 1 & \text{if } \underline{X} \in C_{\underline{X}}, \\ 0 & \text{otherwise.} \end{cases}$$

Normal models

With T univariate and a single screening variable X, Wong, Meeker & Selwyn (1985) describe a predictive approach to screening under (1.4), with $p(t, x|\underline{\theta})$ bivariate normal and conjugate priors for $\underline{\theta}$. When T and X are positively correlated, they obtain C_X that has the natural form $[w, \infty)$ for the one–sided specification region $C_T = [\ell, \infty)$.

With an increase in the number of screening variables $\underline{X}$, the local approach is also taken by Tsai & Moskowitz (1986) who term the probability $P(T \notin C_T|\underline{x})$, the individual nonconforming probability (INP). The INP for those items retained by the screen is called the individual unit misclassification error (IME). The screening criterion is given as:

$$\text{IME} \leq \beta \qquad \equiv \qquad P(T \notin C_T|\underline{x}) \leq \beta \text{ for } \underline{x} \in C_{\underline{X}},$$

where β is predetermined. With $\beta = 1 - \delta_L$, this criterion is equivalent to the rule in (1.4). The authors' approach takes a one sided specification region for univariate T with $C_T = [\ell, \infty)$ but assumes no form for $C_{\underline{X}}$. Instead, once $\underline{X} = \underline{x}$ has been measured, they propose calculating $P(T \in C_T|\underline{x})$ for each item and then comparing $P(T \in C_T|\underline{x})$ with the criterion (1.4) to determine whether the item should be accepted. The screening procedure is described in the context of a multivariate normal model for $p(t, \underline{x}|\underline{\theta})$, with parameters assumed known. Under the same modelling assumptions, Moskowitz & Tsai (1988) and Moskowitz, Plante & Tsai (1993) extend this approach to multi–stage screening procedures in which a covariate is measured at each stage of the screen. If we denote as ${}^i\underline{x}$ the values of the covariates measured up to and including stage i of the screen, then, at stage i, an item is accepted as useful if $P(T \in C_T|{}^i\underline{x}) \geq \delta_1$ and an item is rejected if $P(T \in C_T|{}^i\underline{x}) \leq \delta_2$, where δ_1 and δ_2 are prespecified. Those items that are unsentenced are passed on to stage $i + 1$. At the final stage of the screen all items are sentenced by the rule in (1.4) with $\delta_L = \delta_1$. The motivation behind this sequential procedure is to save cost — the screen may sentence an item without the need to measure all of the covariates. In the conclusion of the paper, the authors suggest that δ_1 and δ_2 might be chosen using decision theory.

Binary responses

For the case in which a univariate T is binary and X is univariate, Dunsmore & Boys (1988) compare screening regions constructed under the local and global screening criteria. The global approach is as described in Boys & Dunsmore (1987) and the local screen is constructed under the same modelling assumptions. In the local approach, all that is required is an expression for the probability $P(T = 1|x)$ with C_X then given as those values of x for which $P(T = 1|x)$ is greater than δ_L. Under the diagnostic paradigm the authors take a logistic model for $T|X, \underline{\xi}$ and assume that the posterior distribution of $\underline{\xi}$ is approximately normal. Under three different approximations to $P(T = 1|x)$, they show that the form of $P(T = 1|x)$ is reasonably straightforward

and, supposing that C_X is of the form $[w, \infty)$, the screening parameter w is easy to obtain. It is seen that an advantage of a local screen designed under a diagnostic approach is that no model for X is required. The sampling approach takes a normal model for $X|T = i, \underline{\eta}$, $i = 1, 0$, a binomial model for the number of cases with $t = 1$ and vague priors for the parameters. Again, there is a simple form for $P(T = 1|x)$. The local and global screens are compared for a numerical example with $\delta = \delta_L$. In both approaches the authors assume that C_X is of the form $\{x : x \geq w\}$ and it is seen that w for the local screen is larger than w for the global screen. Hence, in their example, the local screen will discard more items than the global screen.

Nonparametric methods

A nonparametric approach for obtaining a local screening procedure can be adopted by taking a kernel estimate (Copas (1983)) for the regression function $P(T = 1|x)$ and then solving (1.4); see Boys (1992).

1.1.3 Other statistical screening criteria

Turkman & Amaral Turkman (1989) report that the global screening condition (1.1) has drawbacks. Rule (1.1) may not be attainable and when attainable, the misclassification probability (1.2) may be raised to unacceptable levels. An alternative is proposed in which the screening region $C_{\underline{X}}$ that maximises $P(\underline{T} \in C_{\underline{T}}|\underline{X} \in C_{\underline{X}})$ is found, subject to a given acceptance rate $\alpha = P(\underline{X} \in C_{\underline{X}})$. The rate α is termed the *size* of the screening region. A solution is produced that assumes no fixed form for the region $C_{\underline{X}}$ and that is equivalent to the local screening condition in (1.4) with $\delta_L = k\gamma$, where k is chosen to ensure the appropriate size α. The authors propose that the size of the screen might be chosen using decision theory, see section 1.2. The method is illustrated under a predictive approach with a bivariate normal model for $p(t, x|\underline{\theta})$.

Another screening problem is posed by Owen, Li & Chou (1981). Values of ζ,

m and k are given and a screen is constructed so that, with probability ζ, at least m of k items retained by the screen are acceptable. Writing V for the unknown number of acceptable items among the k that are retained, the first step is to find $\delta_m = P(\underline{T} \in C_{\underline{T}} | \underline{X} \in C_{\underline{X}})$, the probability that a retained item is acceptable, that satisfies

$$P(V \geq m | k, \delta_m) = \zeta.$$

The random variable V is binomial with k 'trials' and probability δ_m of a 'success'. Hence, δ_m is required so that

$$\sum_{j=m}^{k} \binom{k}{j} \delta_m^j (1 - \delta_m)^{k-j} = \zeta.$$

After solving for δ_m, the screening region C_X that satisfies $P(\underline{T} \in C_{\underline{T}} | \underline{X} \in C_{\underline{X}}) = \delta_m$ can be found using methods referenced in section 1.1.1.

Tang & Tang (1994) report on two interesting ideas which we will not discuss beyond this section. One is group testing, in which it is possible to simultaneously test a group of items to determine whether the group contains one or more bad items. If the group fail the test, the items are tested individually. The other is the idea of burn–in testing which is appropriate for types of items that are likely to fail only in early use and so they are submitted to a rigorous burn–in period of use before shipment.

1.2 Designing screens using decision theory

In a decision–theoretic approach to screen design, losses are assigned depending on both the action of the screen (retain or discard) and the performance of the item. A screen based on covariates retains an item if $\underline{X} \in C_{\underline{X}}$ and so the possible losses incurred by screening an item can be described by the following table.

	$\underline{T} \in C_{\underline{T}}$	$\underline{T} \notin C_{\underline{T}}$
$\underline{x} \in C_{\underline{X}}$	$c_{aa}(\underline{t}, \underline{x})$	$c_{ar}(\underline{t}, \underline{x})$
$\underline{x} \notin C_{\underline{X}}$	$c_{ra}(\underline{t}, \underline{x})$	$c_{rr}(\underline{t}, \underline{x})$

The costs $c_{aa}(\underline{t}, \underline{x})$ and $c_{rr}(\underline{t}, \underline{x})$ are paid when the screen accepts a good item or rejects a bad item, respectively. Misclassification costs are paid when the screen wrongly assigns an item, these are: $c_{ar}(\underline{t}, \underline{x})$, the cost of retaining an unacceptable item and $c_{ra}(\underline{t}, \underline{x})$, the cost of discarding a good item. Additionally, there may be a cost for operating the screen denoted c_s.

The ranking by size of these costs and the choice of appropriate loss functions will depend on the context of the problem. However, it is generally agreed that $c_{aa}(\underline{t}, \underline{x})$ denotes the smallest of the costs (in reality it is usually negative and hence a profit) and it is usually set at zero, with the other costs assessed relative to it.

In a medical context $c_{rr}(\underline{t}, \underline{x})$ might also be set to zero indicating that a correct diagnosis by the screen, whether positive or negative, is equally beneficial. In any case, $c_{rr}(\underline{t}, \underline{x})$ will usually be small in comparison with the costs of misdiagnosis given by $c_{ar}(\underline{t}, \underline{x})$ and $c_{ra}(\underline{t}, \underline{x})$, and so is often considered negligible.

Quality control literature often assumes that the loss function for rejection is the same whether the item is good or bad, with $c_{ra}(\underline{t}, \underline{x}) = c_{rr}(\underline{t}, \underline{x}) = c_r(\underline{t}, \underline{x})$. In such a scenario, the cost of rejecting an item $c_r(\underline{t}, \underline{x})$ represents the cost of sending the item for repair, or the loss in sales from the item being discarded or sold at a reduced price.

Normal models

Under a bivariate normal model for $p(t, x|\underline{\theta})$, Boys & Dunsmore (1986) suggest a decision–theoretic approach to screening with penalties for misclassification only. For the case in which $C_T = [\ell, \infty)$, the screening region is assumed to be of the form $C_X = [w, \infty)$, and w is chosen to minimise the expected cost of screening an item, given by

$$\begin{aligned} \mathcal{K}(w) &= \text{Expected cost of wrongly discarding the item} \\ &\quad + \text{Expected cost of wrongly retaining the item,} \\ &= \int_{\ell}^{\infty} \int_{-\infty}^{w} c_{ra}(t,x)p(t,x)\,dx\,dt + \int_{-\infty}^{\ell} \int_{w}^{\infty} c_{ar}(t,x)p(t,x)\,dx\,dt, \end{aligned}$$

where $p(t,x)$ is the predictive distribution for (T,X). The solution for the case in which $c_{ra}(t,x)$ and $c_{ar}(t,x)$ are positive constants is described.

Tang (1987,1988a) also assumes bivariate normality but supposes that parameters are known. In the former paper, Tang (1987), the author assumes that C_T and C_X are of the same form as Boys & Dunsmore (1986) but considers a different loss function set–up. With $c_{aa}(t,x) = 0$, $c_{ar}(t,x)$ is the cost of acceptance which is paid when a customer is dissatisfied with a substandard item. Three loss functions are considered for $c_{ar}(t,x)$, namely,

$$\begin{aligned} c_{ar}(t,x) &= \nu, \\ c_{ar}(t,x) &= b(\ell - t), \\ c_{ar}(t,x) &= k(\ell - t)^2, \end{aligned} \tag{1.5}$$

where ν, b and k are positive constants. The first cost function denotes cases in which dissatisfaction with an item is the same however bad it is. The latter two loss functions denote increased dissatisfaction with items as performance level decreases. Tang supposes that the cost of rejection is constant for all items, with $c_{ra}(\underline{t},\underline{x}) = c_{rr}(\underline{t},\underline{x}) = c_r$.

In Tang (1988a) the performance variable has an 'ideal' value, τ say, and acceptable items are those with t close to τ. The screening region C_X is assumed to be of the form $[v,w]$ and a *cost for quality* is assigned to those items that are accepted by the screen. This cost is given by various functions of $|\tau - t|$ similar to those in (1.5). A constant cost for rejection is again imposed. Based on this approach, Tang & Tang (1989) make a technical note that concerns situations with a multivariate screening variable $\underline{X}$ and suggest dimensionality reduction schemes based on taking a linear combination of the components of the screening variable as in (1.3). In the context of a multivariate normal model for $T,\underline{X}|\underline{\theta}$ with parameters known, for constructing a screening region of the form (1.3), it is shown that the optimal linear combination of the screening variables has coefficients $\underline{a}$ that maximise $corr(\underline{a}^T\underline{X},T)$.

Binary responses

In the simple case with constant misclassification costs c_{ar} and c_{ra}, Kim & Bai (1990) examine the case with T binary, X univariate, and $C_X = [w, \infty)$. Under the diagnostic modelling approach they take a logistic model for $p(t|x, \underline{\xi})$ and in the sampling approach use normal models for $p(x|t, \underline{\eta})$. They consider cases in which the parameters are known or are estimated from a sample on $(T, \underline{X})$.

Boys & Glazebrook (1992) consider the same set up as Kim & Bai but restrict themselves to the diagnostic modelling paradigm with $p(t|x, \underline{\xi})$ modelled by a probit regression. The advantage of this choice of link function is that, in a predictive approach, the probability $P(T = 1|x)$ has a closed form when the distribution for $\underline{\xi}$ is normal. The parameters of the probit model are approximately normal when their distribution is based on a large sample and relatively weak prior information (see Lindley (1961)). Under a normality assumption for $\underline{\xi}$, the authors give a solution which is simple and intuitive in terms of model parameters. Also, a robustness study shows that the expected cost of screening an item under their optimal designs is robust to departures from the normality of $\underline{\xi}$. Note that, here, under the diagnostic paradigm, no model for $p(x)$ is needed to construct C_X. Recall that this was also the case in the solution for the local screening criterion (1.4) under a diagnostic modelling approach. However, a model for $p(x)$ is needed to compute expected costs for comparison purposes.

The optimal screening region under costs for misclassification

The Turkman & Turkman (1989) paper referred to in section 1.1.3 suggests a decision–theoretic approach to the screening problem. With losses for misclassification, the authors show that the optimal screening region $C_{\underline{X}}$ contains those $\underline{X}$ for which acceptance is cheaper then rejection, a very intuitive and simple result. They also show that, when c_{ar} and c_{ra} are constants, their decision–theoretic solution is equivalent to the local condition (1.4) with $\delta_L = c_{ar}/(c_{ar} + c_{ra})$.

A two–stage screen

The classification of items based on measurements of correlated variables is prone to misclassification errors. Although the performance variable is usually expensive to measure it may still be worthwhile doing so if the probability of a misclassification based on $\underline{X}$ is high enough. For the unrealistic case in which model parameters are known, Tang (1988b) develops a two–stage procedure in which the first stage is based on a correlated variable and at the second stage the performance variable is measured for those items left unsentenced by the first stage. A decision–theoretic approach is taken with a cost for measuring the performance variable, c_m say, added to the cost structure assumed in Tang (1987). Under a bivariate normal model with known parameters and with T and X positively correlated, the first stage of the screen is assumed to take the form: if $\underline{x} \in (-\infty, v]$ then reject the item, if $\underline{x} \in [w, \infty)$ then accept the item, otherwise pass the item to the second stage where T is measured. In this set up, values for the design parameters $(v, w)^T$, $v \leq w$ are obtained with $c_{ar}(t, x)$ given by each of the three loss functions in (1.5). The paper concludes with a study of the sensitivity of the expected cost of an optimal design to the choice of loss function.

1.3 Overview of the thesis

Here we briefly describe the content of the thesis with reference to the literature reviewed above. Further information and motivation is given in the introduction to each chapter.

In a decision–theoretic framework, Chapter 2 considers the two–stage screening procedure suggested by Tang (1988b) with T binary and X univariate. As already mentioned, we advocate predictive models and the simple and robust solution obtained by Boys & Glazebrook (1992) prompts us to follow their approach with constant costs for misclassification and a probit regression model for $T|X, \underline{\xi}$. For X positively correlated to T, the form of the first stage of the screen is assumed to

be the same as in Tang (1988b). Simple solutions are obtained for the screening parameters v and w and the method is illustrated by an example.

Whereas Chapter 2 concerns the simple case of a two–stage screen under specific cost and modelling assumptions, Chapter 3 allows: (a) a general stochastic structure for $(\underline{T}, \underline{X})$, (b) a more general loss function set up for misclassification costs and (c) assumes no fixed form for the screen. The optimal screening regions are obtained and the solution is illustrated in the context of the probit regression model of Chapter 2. The solutions obtained in Chapter 2 are shown to be cost–optimal under conditions on the parameters of the model and the misclassification costs. Also in Chapter 3, we consider a scenario in which a statistical goal or constraint is imposed in addition to the decision–theoretic target of minimising expected cost. We suppose that there are limited resources for measuring the performance variable and so a constraint is placed on the proportion of items passed to the second stage of the screen. The solution obtained for this constrained problem is intuitive: if too high a proportion of items are passed through to the second stage, the cost of measuring the performance variable is increased. The example of Chapter 2 is extended to illustrate the methods presented in Chapter 3.

In Chapter 4 we propose a sequential screen similar in form to that described in Moskowitz, Plante & Tsai (1993). At each stage of a sequence a covariate is measured and items may be accepted as suitable, discarded or passed on to the next stage. At the final stage the performance variable T is measured. The simple designs found in Chapter 2 are used in a heuristic design to the sequential screen and the heuristic solution is described in full for the case of two covariates. The performance of the heuristic is assessed in a simulation study and in an illustrative example.

Returning to the simple one–stage screen based solely on measuring covariates, Chapter 5 poses the question of how many and which covariates to include as part of the screen. As there may be a restriction on cost, it may not be possible to use all available covariates. Heuristics for choosing screening components are proposed that avoid the computationally intensive task of comparing expected costs and the

performance of these heuristics is assessed in a numerical study.

Chapter 6 makes some final remarks about the techniques presented in the preceding chapters and suggests some avenues for further research.

Chapter 2

The economic design of a simple two–stage screen

2.1 Introduction

In this chapter we develop solutions for a two–stage screen as considered in Tang (1988b). A univariate correlated variable X is measured at the first stage and used as a screening variable. At the second stage a single performance variable T is measured for those items for which the first stage is inconclusive. The motivation here is that there may be economic advantages in reducing error probabilities by measuring T even when it is expensive to do so.

In Tang's model all parameters are assumed known, a situation that is unrealistic. For the case in which T is binary, Boys & Glazebrook (1992) assume a probit model for $T|X$ and, with asymptotic posterior distributions, yield an optimal design for the one–stage screen based only on the screening variable X. Their solution is easy to use and interpret. Here we use the model of Boys & Glazebrook in the context of a two–stage screen.

In section 2.2 we develop Bayes optimal solutions for a two–stage screen when the model for (T, X) is as in Boys & Glazebrook. We assume large values of X imply

that $T = 1$ and so investigate screens assumed to be of the form:

$$\begin{aligned} X \in [w, \infty) \quad &\rightarrow \quad \text{accept the item,} \\ X \in (v, w) \quad &\rightarrow \quad \text{measure } T; \\ X \in (-\infty, v] \quad &\rightarrow \quad \text{reject the item.} \end{aligned}$$

We also show that, under certain conditions, an optimal screen of this form is preferable to at least one other alternative. Section 2.3 reviews some of the model assumptions of section 2.2 and discusses the robustness of our solutions to departures from these model assumptions. Also, section 2.3 gives mathematical details omitted from section 2.2. We give an illustrative example of the application of our method in section 2.4. The example would be more commonly thought of as a discrimination problem but screening theory is equally applicable.

2.2 A Bayes optimal two–stage screen

Suppose that the performance variable T is a binary response variable with $T = 1$ if and only if an item meets the standard and should be accepted. Otherwise $T = 0$ and the item is not fit for use and should be rejected. We think of T as a "gold standard" measurement, and so assume that measurement of T is error-free. Further suppose that the screening variable X is continuous with marginal density function $\psi(x)$ and that the dependence of T on X is expressed through a generalised linear model with

$$P(T = 1|x, \underline{\xi}) = F(\xi_0 + \xi_1 x),$$

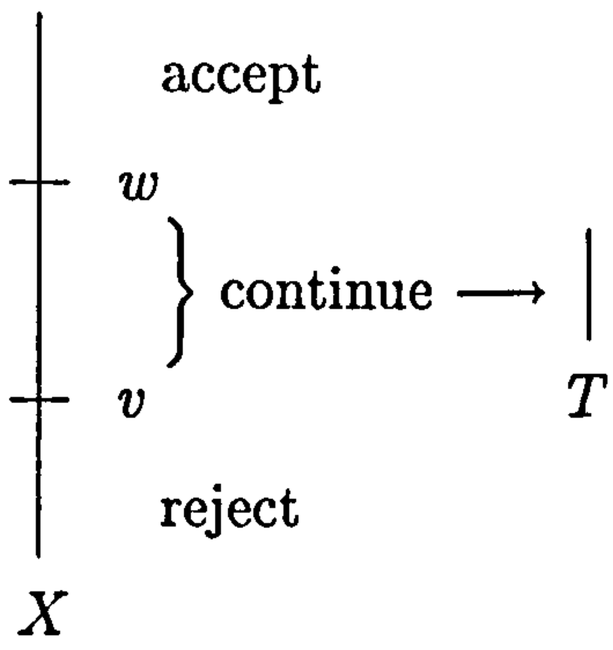
where F is a link function. Here F is required to be a monotonically increasing function that maps $\mathbb{R}$ into $(0, 1)$. Hence

$$P(T = 1|x) = \int_{\underline{\xi}} F(\xi_0 + \xi_1 x)\pi(\underline{\xi})d\underline{\xi},$$

where the distribution of the regression parameters, $\pi(\underline{\xi})$, summarises our posterior beliefs about ξ_0 and ξ_1. The distribution $\pi(\underline{\xi})$ may be based on a sample of data,

from past measurements of (T, X), and/or prior information about $\underline{\xi}$. Note that $\pi(\underline{\xi})$ does not change after observing x.

For the moment we shall assume that large values of X tend to indicate that the item meets the standard ($T = 1$). This larger–the–better assumption corresponds to the model assumption that $\xi_1 > 0$. Since measuring T is expensive, we shall consider a screen which, where possible, eliminates the need for T to be measured. Taking into consideration the assumptions made above, it is natural to design a screen in which, when X is above a cut-off point, w say, the item is accepted. Similarly, the design should include a cut-off point, v say, such that if $X \leq v$ the item is rejected. This eliminates the need to measure T for items that we are reasonably sure of correctly classifying using the screening variable only. When $v < X < w$, we remain unsure whether to accept or reject the item and it is necessary to measure T. Graphically, we consider a screen of the form



that is, we accept an item only if $X \geq w$ or $v < X < w$ and $T = 1$. The above screening procedure is characterised by the values of v and w and we shall determine optimal values for these quantities using Bayes decision theory.

Suppose that the cost structure of the screening mechanism is as follows. The unit cost incurred by rejecting a good item using the X-screen is c_r and that incurred by accepting a bad item is c_a. The costs (per item) of measuring X and T are c_s and c_m respectively. The Bayes cost (expected cost) of screening an item is therefore

$$\begin{aligned}\mathcal{K}(v, w) = c_r\, P(\text{reject a good item}) + c_a\, P(\text{accept a bad item}) \\ + c_m\, P(\text{need to measure } T) + c_s\, P(\text{need to measure } X).\end{aligned}$$

In terms of the screen design above,

$$\begin{aligned} \mathcal{K}(v,w) = c_r\, P(T=1, X \le v) + c_a\, P(T=0, X \ge w) \\ + c_m\, P(v < X < w) + c_s\, I(v,w), \qquad -\infty \le v \le w \le \infty, \end{aligned} \tag{2.1}$$

where

$$I(v,w) = \begin{cases} 1, & \text{if either } v \text{ or } w \text{ are finite,} \\ 0, & \text{otherwise.} \end{cases}$$

Note that, if $-\infty \le v = w \le \infty$ then we have a reduction to a one–stage screen in which sentence is passed on the basis of X only. If both v and w take limiting values X is not measured. In the three cases in which both v and w are not finite the screen acts as follows:

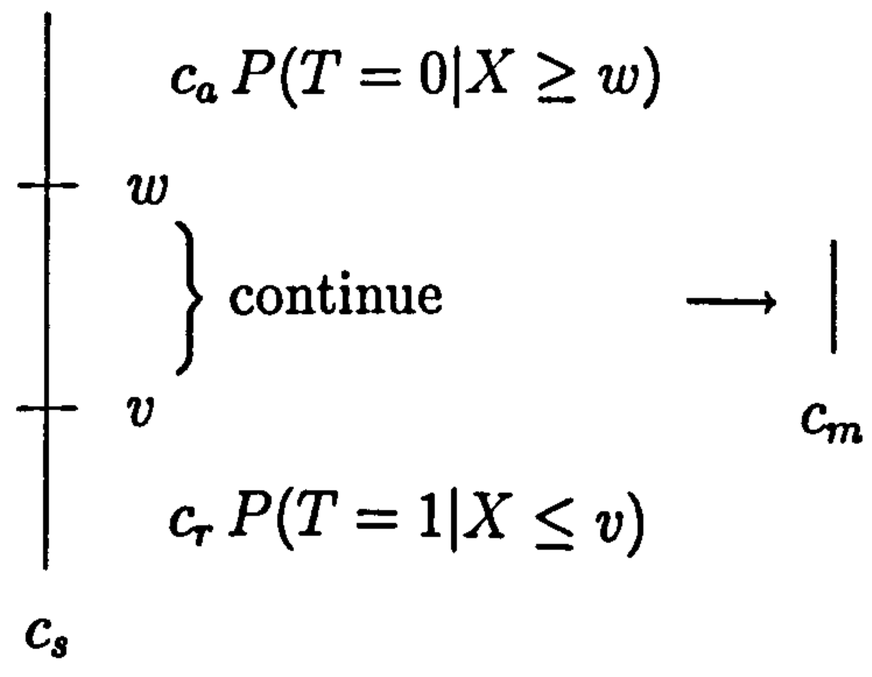
(i) in the limits $v \to \infty$ and $w \to \infty$, the screen rejects all items without measuring X or T,

(ii) in the limits $v \to -\infty$ and $w \to -\infty$, the screen accept all items as reaching the standard without measuring X or T,

(iii) in the limits $v \to -\infty$ and $w \to \infty$, the screen measures the performance variable T on all items.

In all of these reductions (2.1) gives the appropriate cost.

The penalties incurred given the decision based on the X–screen for finite v and w are illustrated below.



Definition 2.1 (v^*, w^*) is a *Bayes design* if

$$\mathcal{K}(v^*, w^*) = \inf_{v,w} \mathcal{K}(v, w),$$

where the above infimum is over all (v, w) satisfying $-\infty \leq v \leq w \leq \infty$.

Note from (2.1) that, once X has been measured, a decision to measure T will never be optimal if $c_m > \min(c_r, c_a)$. When $c_m > \min(c_r, c_a)$ and for any choice of v and w, clearly c_m will always be greater than $\min\{c_r P(T = 1|X \leq v), c_a P(T = 0|X \geq w)\}$, and so T should never be measured.

When $\mathcal{K}(v, w)$ is not minimised in the limit as $v \to \pm\infty$ and $w \to \pm\infty$ consideration of the turning points of

$$\mathcal{K}'(v, w) = \mathcal{K}(v, w) - c_s I(v, w)$$

will yield a Bayes design. $\mathcal{K}'(v, w)$ is the Bayes cost per item, the cost of measuring X excepted. Expressing this cost in terms of the statistical model we obtain

$$\begin{aligned} \mathcal{K}'(v, w) &= c_r \int_{-\infty}^{v} \int_{\underline{\xi}} F(\xi_0 + \xi_1 x)\pi(\underline{\xi})\psi(x)d\underline{\xi}dx \\ &\quad + c_a \int_{w}^{\infty} \int_{\underline{\xi}} \{1 - F(\xi_0 + \xi_1 x)\}\pi(\underline{\xi})\psi(x)d\underline{\xi}dx \\ &\quad + c_m \int_{v}^{w} \psi(x)dx\,, \qquad -\infty \leq v, w \leq \infty. \end{aligned}$$

Note from the above that we deem $\mathcal{K}'$ to be defined by this formula for all pairs (v, w) and not only those with $v \leq w$. However, from the perspective of the development of Bayes designs, it will plainly be of interest to determine when $\mathcal{K}'$ has a minimum (v', w') such that $-\infty < v' < w' < \infty$. From the subsequent analysis it emerges that a necessary condition for this is that $c_a^{-1} + c_r^{-1} < c_m^{-1}$. See comment 2 of Theorem 2.1.

Trivially,

$$\begin{aligned} \frac{\partial}{\partial v}\{\mathcal{K}'(v, w)\} &= \psi(v)\left\{c_r \int_{\underline{\xi}} F(\xi_0 + \xi_1 v)\pi(\underline{\xi})d\underline{\xi} - c_m\right\} \qquad (2.2) \\ \frac{\partial}{\partial w}\{\mathcal{K}'(v, w)\} &= \psi(w)\left\{c_a \int_{\underline{\xi}} F(\xi_0 + \xi_1 w)\pi(\underline{\xi})d\underline{\xi} - c_a + c_m\right\} \end{aligned}$$

and hence the turning points (v, w) satisfy

$$\int_{\underline{\xi}} F(\xi_0 + \xi_1 v)\pi(\underline{\xi})d\underline{\xi} = \frac{c_m}{c_r}, \quad \int_{\underline{\xi}} F(\xi_0 + \xi_1 w)\pi(\underline{\xi})d\underline{\xi} = 1 - \frac{c_m}{c_a}. \tag{2.3}$$

Note that $\psi(x)$, the marginal density of X, plays no part in the solution. Also, as equations (2.3) are not coupled and are of a similar form, the determination of the turning points of $\mathcal{K}'$ is fairly straightforward.

We now investigate, in detail, the solution when $T|X$ follows a linear probit regression, that is, $F \equiv \Phi$, where Φ is the standard normal cumulative distribution function. Hence

$$P(T = 1|x, \underline{\xi}) = \Phi(\xi_0 + \xi_1 x) = \int_{-\infty}^{\xi_0 + \xi_1 x} \phi(z)dz,$$

where $\phi(.)$ is the standard normal density function. We also assume that the regression parameters $\underline{\xi}$ follow a bivariate normal distribution, $N_2(\underline{m}, \boldsymbol{S})$, say. This would be (approximately) the case if $\pi(\underline{\xi})$ were a posterior distribution based on a sufficiently large sample of (T, X) so that $\pi(\underline{\xi})$ may be well approximated by its asymptotic normal form. In this case, $\underline{m}$ is the maximum likelihood estimate for $\underline{\xi}$ and $\boldsymbol{S}$ is the inverse of Fisher's information matrix evaluated at $\underline{m}$. To make the notation more explicit we write

$$\underline{m} = \begin{pmatrix} m_0 \\ m_1 \end{pmatrix}, \quad \boldsymbol{S} = \begin{pmatrix} s_0^2 & rs_0s_1 \\ rs_0s_1 & s_1^2 \end{pmatrix}.$$

So, for given v and w,

$$\begin{aligned} \eta = \xi_0 + \xi_1 v &\sim N(m_0 + m_1 v, s_0^2 + 2vrs_0s_1 + v^2 s_1^2); \\ \zeta = \xi_0 + \xi_1 w &\sim N(m_0 + m_1 w, s_0^2 + 2wrs_0s_1 + w^2 s_1^2). \end{aligned} \tag{2.4}$$

The following result enables us to show that, by inserting the probit link function and under a normality assumption for $\underline{\xi}$, the left hand sides in equations (2.3) are in a closed form.

Lemma 2.1 If $Y \sim N(\mu, \sigma^2)$ then

$$E\{\Phi(Y)\} = \Phi\left\{\frac{\mu}{(1+\sigma^2)^{1/2}}\right\} \tag{2.5}$$

Proof

$$\begin{aligned} E\{\Phi(Y)\} &= \sigma^{-1}\int_{-\infty}^{\infty}\Phi(y)\phi\{(y-\mu)/\sigma\}\,dy, \\ &= P(Z<Y), \text{ where } Z\sim N(0,1) \text{ and } Y \text{ are independent}, \\ &= P(Z-Y<0), \\ &= \Phi\left\{\mu/\left(1+\sigma^2\right)^{1/2}\right\}. \end{aligned}$$ □

With a probit link function, the left hand sides in (2.3) are $E\{(\Phi(\eta)\}$ and $E\{(\Phi(\zeta)\}$ respectively. As η and ζ both have normal distributions (see (2.4)) then, by Lemma 2.1, equations (2.3) become

$$\begin{aligned} \Phi\left\{\frac{m_0+m_1v}{(1+s_0^2+2vrs_0s_1+v^2s_1^2)^{1/2}}\right\} &= k_1; \qquad (2.6) \\ \Phi\left\{\frac{m_0+m_1w}{(1+s_0^2+2wrs_0s_1+w^2s_1^2)^{1/2}}\right\} &= k_2, \end{aligned}$$

where $k_1 = c_m/c_r$ and $k_2 = 1 - c_m/c_a$.

Without loss of generality we may assume that $m_0 = 0$, $s_1 = 1$ and $m_1 > 0$. If $s_1 \neq 1$ or $m_0 \neq 0$ then a change of regression parameters from (ξ_0, ξ_1) to $(\xi_0 - m_0\xi_1/m_1, \xi_1/s_1)$ ensures that $s_1 = 1$ and $m_0 = 0$. If $m_1 < 0$ then a further change in parameters from (ξ_0, ξ_1) to $(\xi_0, -\xi_1)$ results in a case with $m_1 > 0$. See section 2.3.8 for the resultant data transformations. Note that when $m_0 = m_1 = 0$, equations (2.6) have no solutions in which v and w are finite. Under the assumptions that $m_0 = 0$, $s_1 = 1$ and $m_1 > 0$ equations (2.6) simplify to give,

$$g(v) = c_1 \qquad \text{and} \qquad g(w) = c_2, \qquad (2.7)$$

where

$$g(x) = \frac{m_1x}{(1+s_0^2+2xrs_0+x^2)^{1/2}}$$

and $c_i = \Phi^{-1}(k_i)$, $i = 1, 2$. Under these conditions, the following result describes exactly when there exist turning points (v, w) minimising $\mathcal{K}'(v, w)$. Since equations (2.7) are not coupled, the turning points (v, w) can be determined by considering each variable separately, and so we analyse the functions $\mathcal{K}'(u, w)$ for fixed $w \in (-\infty, \infty)$ and $\mathcal{K}'(v, u)$, for fixed $v \in (-\infty, \infty)$, in turn.

Theorem 2.1 Statements (a), (b) and (c) describe the turning points of $\mathcal{K}'(u, w)$ for fixed $w \in (-\infty, \infty)$.

(a) When $c_1 > 0$,

(i) if $m_1 > c_1$ then

$$u_1^+ = c_1^2[rs_0 + \{r^2 s_0^2 + (m_1^2 - c_1^2)(1 + s_0^2)c_1^{-2}\}^{1/2}]/(m_1^2 - c_1^2)$$

achieves a (global) minimum;

(ii) if $m_1 = c_1$, $r < 0$ then $u_1' = -(1+s_0^2)/(2rs_0)$ achieves a (global) minimum;

(iii) if $0 < m_1 < c_1$, $r < 0$ and $r^2 s_0^2 + (m_1^2 - c_1^2)(1 + s_0^2)c_1^{-2} > 0$ then u_1^+ in (i) achieves a (local) minimum;

(iv) in all other cases with $m_1 > 0$, there are no turning points.

(b) When $c_1 < 0$,

(i) if $m_1 > -c_1$ then

$$u_1^- = c_1^2[rs_0 - \{r^2 s_0^2 + (m_1^2 - c_1^2)(1 + s_0^2)c_1^{-2}\}^{1/2}]/(m_1^2 - c_1^2)$$

achieves a (global) minimum;

(ii) if $m_1 = -c_1$, $r > 0$ then $u_1' = -(1 + s_0^2)/(2rs_0)$ achieves a (global) minimum;

(iii) if $0 < m_1 < -c_1$, $r > 0$ and $r^2 s_0^2 + (m_1^2 - c_1^2)(1 + s_0^2)c_1^{-2} > 0$ then u_1^- in (i) achieves a (local) minimum;

(iv) in all other cases with $m_1 > 0$, there are no turning points.

(c) When $c_1 = 0$, $u_1^\dagger = 0$ achieves a (global) minimum.

A description of the turning points of $\mathcal{K}'(v, u)$ for fixed $v \in (-\infty, \infty)$ can be found by replacing $(c_1, u_1^+, u_1', u_1^-, u_1^\dagger)$ in the above by $(c_2, u_2^+, u_2', u_2^-, u_2^\dagger)$.

Proof

The turning points of $K'(u,w)$ for fixed w are solutions to

$$g(u) \equiv \frac{m_1 u}{(1+s_0^2+2urs_0+u^2)^{1/2}} = c_1. \tag{2.8}$$

Since $m_1 > 0$, when $c_1 = 0$ the only real u satisfying (2.8) is zero. Otherwise, the sign of c_1 and of any real solution u of (2.8) coincide. We square both sides of (2.8) and look for real roots which are positive for $c_1 > 0$ and negative for $c_1 < 0$. The roots of the resulting quadratic are

$$m_1 \neq c_1, \qquad u_1^{\pm} = c_1^2[rs_0 \pm \{r^2 s_0^2 + (m_1^2 - c_1^2)(1+s_0^2)c_1^{-2}\}^{1/2}]/(m_1^2 - c_1^2);$$

$$m_1 = c_1, \qquad u_1' = -(1+s_0^2)/(2rs_0).$$

Under the conditions stated in (a)(i), (ii) and (b)(i), (ii) there is one real root with the appropriate sign, for (a)(iii) and (b)(iii) there are two candidate roots and for (a)(iv) and (b)(iv) there are none.

To verify which of the above turning points represent a minimum we now look at the second derivatives of $K'(u,w)$ under the probit model (for fixed w). As equations (2.2) are not coupled it is enough to show that $\partial^2 K'(u,w)/\partial u^2 > 0$ at the turning point concerned, for that point to represent a minimum. Now

$$\frac{\partial^2 K'(u,w)}{\partial u^2} = c_r \Phi\{g(u)\} \frac{\partial \psi(u)}{\partial u} + c_r \psi(u) \frac{\partial \Phi\{g(u)\}}{\partial u} - c_m \frac{\partial \psi(u)}{\partial u}.$$

However, at a turning point, u^* say, equation (2.6) gives $\Phi\{g(u^*)\} = c_m/c_r$, so

$$\left.\frac{\partial^2 K'(u,w)}{\partial u^2}\right|_{u=u^*} = c_r \psi(u^*) \left.\frac{\partial \Phi\{g(u)\}}{\partial u}\right|_{u=u^*}.$$

Therefore, if we can show that $g(u)$ is increasing in u at a stationary point then there exists a minimum of $K'(u,w)$ for fixed w at that stationary point.

Now

$$\frac{\partial g(u)}{\partial u} = \frac{m_1(1+s_0^2+urs_0)}{(1+s_0^2+2urs_0+u^2)^{3/2}}$$

and so $g(u)$ has only one turning point at $u = -(1+s_0^2)/rs_0$. To determine the nature of this turning point we examine the second derivative of $g(u)$ which is:

$$\frac{\partial^2 g(u)}{\partial u^2} = \frac{m_1 rs_0}{(1+s_0^2+2urs_0+u^2)^{3/2}} - \frac{3m_1(1+s_0^2+urs_0)(rs_0+u)}{(1+s_0^2+2urs_0+u^2)^{5/2}}$$

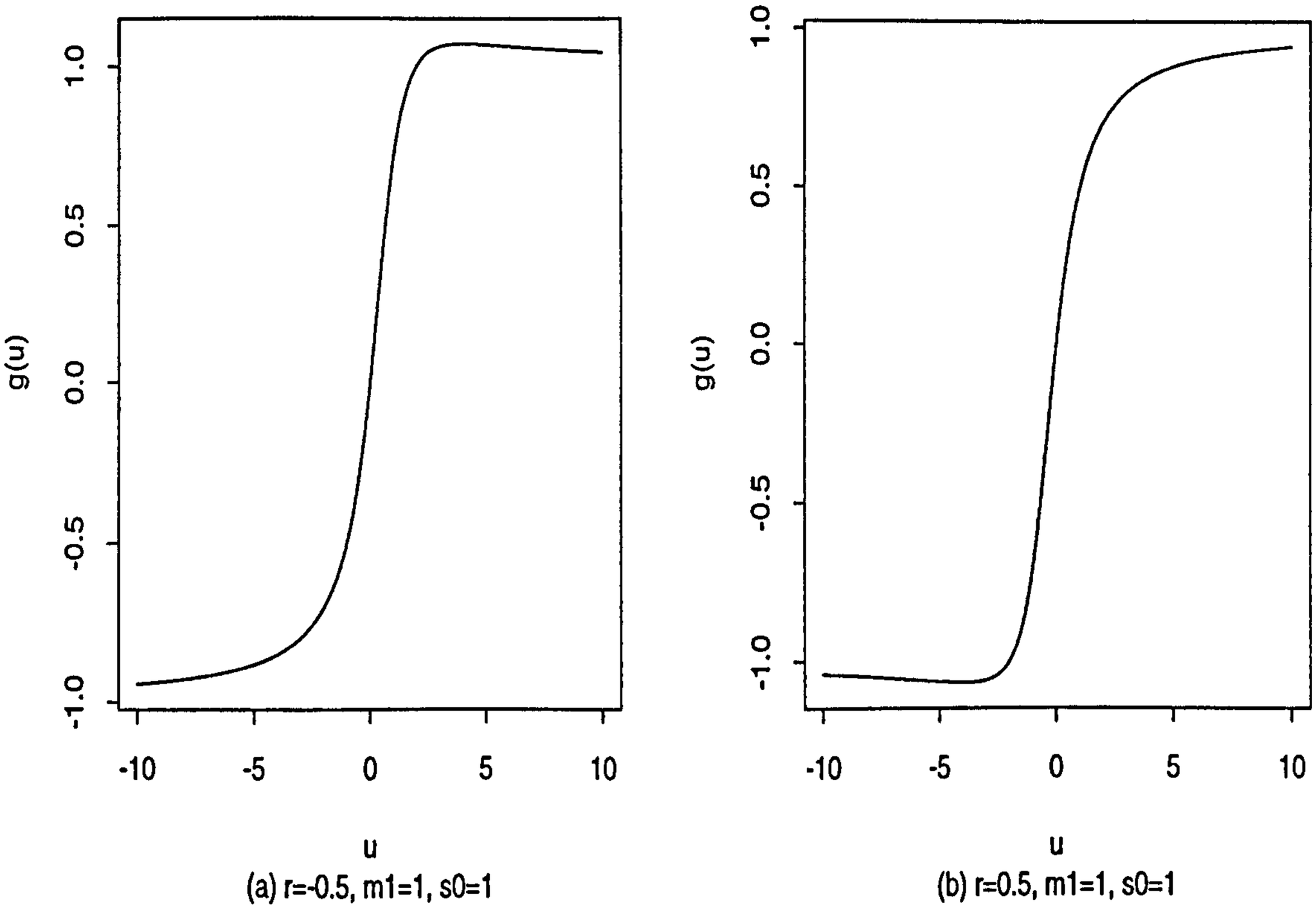


Figure 2.1: Plots of $g(u)$

The second term of this second derivative is zero at the turning point. The square root in the denominator of the first term is a standard deviation term (see equation (2.5)) and we take the positive root. Hence, for a real $g(u)$, the denominator of the first term is positive, as are m_1 and s_0, and the sign of r determines the nature of the turning point.

If $r < 0$, a maximum of $g(u)$ occurs at $u = u_{max} \equiv -(1 + s_0^2)/rs_0$. As $g(u)$ has no other turning points for $m_1 > 0$, $g(u)$ is increasing in u for all $u < u_{max}$ and decreasing in u for all $u > u_{max}$. See Figure 2.1(a). It is easy to see that

$$g(u_{max}) = m_1 \left\{ \frac{1 + s_0^2}{1 + s_0^2 - r^2 s_0^2} \right\}^{1/2} > m_1 \quad \text{and} \quad \lim_{u \to \infty} g(u) = m_1.$$

As there is only one turning point, we have $g(u) > m_1 > 0$ for all $u > u_{max}$. Hence $g(u)$ is increasing in u for all non-positive and unique positive solutions of (2.8), (u_1^+ in (a)(i), u_1' in (a)(ii), u_1^- in (b)(i) and $u_1^\dagger$ in (c)). There are two positive real

roots under conditions given in (a)(iii), these are u_1^+ and u_1^-. As u_{max} is the only turning point of $g(u)$ one of the roots must be larger than u_{max} and the other smaller. Under the conditions stated in (a) (iii) it is easy to see that $u_1^+ < u_1^-$ and conclude that $u_1^+ < u_{max} < u_1^-$. Hence u_1^+ occurs on the increasing slope of $g(u)$ and is the location of a minimum. However the other real root, u_1^- represents a local maximum suggesting that the minimum at u_1^+ may be a local minimum. In this case we must check that $\lim_{u\to\infty} \mathcal{K}'(u,w) > \mathcal{K}'(u_1^+,w)$.

Similarly, if $r > 0$, we can establish that $g(u)$ is increasing in u for all non-negative and unique negative solutions of (2.8) (u_1^+ in (a)(i), u_1^- in (b)(i), u_1' in (b)(ii) and $u_1^\dagger$ in (c)). See Figure 2.1(b). Of the two negative real roots, u_1^- and u_1^+, found under the conditions stated in (b)(iii), the larger (smaller in magnitude) of the two, u_1^-, represents a (local) minimum of $\mathcal{K}'(u,w)$. Here we must check that $\lim_{u\to-\infty} \mathcal{K}'(u,w) > \mathcal{K}'(u_1^-,w)$.

The analysis of turning points of $\mathcal{K}'(v,u)$ for fixed $v \in (-\infty,\infty)$ proceeds in a similar fashion and so we omit the details. □

Comments

1. In those circumstances where no turning points exist, $\mathcal{K}'(v,w)$ is minimised in the limit as $v \to \pm\infty$ and/or $w \to \pm\infty$.

2. Suppose now that turning points do exist. The existence of a minimum satisfying (2.7) requires that the function $g(.)$ be increasing at both v and w. Since g has only one turning point, it must follow that g is increasing at all values between v and w. Hence, for the pair (v,w) with $v < w$ to minimise $\mathcal{K}'(v,w)$, we require that $c_1 < c_2$, that is

$$c_a^{-1} + c_r^{-1} \leq c_m^{-1}. \tag{2.9}$$

When $c_1 \geq c_2$ it will follow that minimising $\mathcal{K}'(v,w)$ over the design space $-\infty \leq v \leq w \leq \infty$ will be achieved at the boundary of the space. Either we

will have $v = w$ or both of v and w taking limiting values.

3. As $\mathcal{K}'(v, w)$ has at most one local minimum, the global minimum is easily found using Theorem 2.1. For example, if $c_2 > m_1 > c_1 > 0$ then the global minimum of $\mathcal{K}'(v, w)$ is either at (u_1^+, u_2^+) or (u_1^+, ∞), whereas if $-c_1 > -c_2 > m_1 > 0$, the global minimum is at one of (u_1^-, u_2^-), $(-\infty, u_2^-)$, and $(-\infty, -\infty)$. Note that $(u_1^-, -\infty)$ is not a solution of the constrained problem $(v < w)$.

4. These optimal designs arise from the minimisation of Bayes cost $\mathcal{K}'(v, w)$. However, $\mathcal{K}$ and $\mathcal{K}'$ coincide when both v and w take infinite values. This corresponds to the degenerate case of not using the screening variable, either by rejecting, accepting or measuring the performance variable on all items. Plainly when $\mathcal{K}'$ is minimised in the limit as $v \to \pm\infty$ and $w \to \pm\infty$ then $\mathcal{K}$ will be minimised by taking the same limit. Where one or both of the minima of $\mathcal{K}'$ are finite, it may still be cheaper not to use the screen and thereby save the screening cost c_s. Therefore, before claiming the minima of $\mathcal{K}'$ as a Bayes design we must verify that the associated Bayes cost $\mathcal{K}$ is less than

$$\min\left\{ \lim_{\substack{v\to-\infty \\ w\to-\infty}} \mathcal{K}(v,w),\ \lim_{\substack{v\to-\infty \\ w\to\infty}} \mathcal{K}(v,w),\ \lim_{\substack{v\to\infty \\ w\to\infty}} \mathcal{K}(v,w) \right\}.$$

5. We have described designs that minimise Bayes cost for the case in which the moments of the regression parameters are such that $m_0 = 0$ and $s_1 = 1$. In section 2.3.8 we describe data transformations which change a case with general regression parameters to a problem of this form. Hence, once we have implemented these changes and found the optimal designs using Theorem 2.1, to recover the designs relating to the original general case of parameters, the data transformation should be reversed. Rearranging equations (2.16), the Bayes design (v^*, w^*) in the general case of regression parameters can be recovered from the Bayes design in the case with $m_0 = 0$ and $s_1 = 1$, given by (v', w') say, by the following equations,

$$v^* = \frac{v'}{s_1} - \frac{m_0}{m_1} \qquad \text{and} \qquad w^* = \frac{w'}{s_1} - \frac{m_0}{m_1}. \tag{2.10}$$

If the original problem has $m_1 < 0$ and a data transformation has been performed to obtain a case in which $m_1 > 0$, then the Bayes design for the original problem can be recovered via the transformations $v \to -v$ and $w \to -w$. In this case it is likely that $\xi_1 < 0$ and it is intuitive that the screen is of the form in (2.11) below, that is: Accept the item if $X < w$, reject the item if $X > v$ and measure T if $w < X < v$.

6. The Bayes designs calculated using Theorem 2.1 are simple and easy to understand in terms of the cost parameters and the first and second moments of the regression parameters in the probit model.

So far we have identified the optimal screening regions of the form: accept the item if $X \in [w, \infty)$, reject it if $X \in [-\infty, v)$ and measure T if $X \in (v, w)$. This is the natural form for the screen when $\xi_1 > 0$. An important question is whether regions of the above form are Bayes optimal when $m_1 > 0$. For example, is it possible that the procedure "accept the item if $X \in [-\infty, w)$, reject it if $X \in [v, \infty)$ and measure T if $X \in (v, w)$" could achieve a lower expected cost when $m_1 > 0$? The following result shows that this cannot happen when $\psi(x)$ is symmetric (even when we ignore the constraint $v \leq w$).

Theorem 2.2 If $m_1 > 0$ and $\psi(x)$ is symmetric then a screening procedure of the form "(II): accept the item if $X \in [-\infty, w)$, reject it if $X \in [v, \infty)$ and measure T if $X \in (w, v)$"

(2.11)

can always be improved on by using a procedure of the form "(I): accept the item if $X \in [w, \infty)$, reject it if $X \in [-\infty, v)$ and measure T if $X \in (v, w)$" or by using no screen at all.

The proof of Theorem 2.2 is given in section 2.3.9. Note that the assumed symmetry of ψ is a much stronger condition than is actually needed for the theorem to hold. The crucial determinant of whether or not to screen is the size of m_1 (Theorem 2.1), that is, in terms of non-standardised quantities, the size of m_1/s_1. It is plain that it will never be optimal to screen when $m_1 = 0$. In section 3.3.1 we provide more detailed conditions under which the form of screen assumed in this chapter is optimal.

2.3 Details and discussion

In section 2.2 we omitted discussion of the assumptions made in forming the solution. Here we comment on those assumptions and also give some mathematical details that were intentionally left out.

2.3.1 The performance variable

We assume that the performance variable, T, is binary, emphasising the divide between an item reaching the standard ($T = 1$) and not ($T = 0$). In quality control this may relate to the dichotomy of items that will work well in use ($T = 1$) and those that will quickly fail in use ($T = 0$). When T is not binary and possibly multivariate, our method can be used in cases in which the dichotomy $T \in C_T$ versus $T \notin C_T$ can be summarised by a binary variable without loss in screen performance. Here the attributes necessary for an item to perform well are given by C_T and the binary variable is set at 1 if the attributes are present and 0 otherwise.

In section 2.2 we further assume that measurement of T is error free. When this is not the case and inspection error is present, Tang & Schneider (1990) show that

the observed value of the performance variable can be treated as a screening variable. In such a case we may wish to use this measurement of the performance variable as a second screening variable in a sequential screen (see Chapter 4) or in a batch screen (see Chapters 3 & 5).

2.3.2 The screening variable

In our solution no assumption is made about the distribution of the screening variable X. Hence we can choose as the screening variable any function of a measurement on an item without concern for its distributional form. The choice of which function of the variable to use in the screen should be a made when ensuring linearity in the probit regression model for $T|X$.

2.3.3 The relationship between T and X

We have taken the diagnostic approach to factorising the joint model for (T, X), with the joint density $p(t, x|\underline{\theta})$ written as $p(t|x, \underline{\xi})p(x|\underline{\phi})$, where the parameters of the model are given by $\underline{\theta}^T = (\underline{\xi}^T, \underline{\phi}^T)$. Hence we use the standard form of binary response models. The design of a simple two–stage screen under the sampling paradigm is discussed in section 6.1. Questions of model adequacy and goodness of fit to data are always important but we will omit them here, see Collett (1991) for a thorough coverage of the issues. We further assume that ξ_1, the coefficient of X, is positive. This modelling set up is the natural way to describe a situation in which larger values of X imply that an item is more likely to reach the standard. In Chapter 1 we note that the form of a performance variable can be changed by a suitable transformation. The screening variable can be adjusted in the same way. For example, when X is a smaller-the-better screening variable, multiplying by -1 creates a larger-the-better variable. Also, when X close to some known nominal value implies that the item is more likely to reach the standard, then taking minus the distance from the nominal value gives a larger-the-better screening variable. Further problems involve the case

in which X very large or very small implies that the item does not reach the standard ($T = 0$). The natural solution here would involve quadratic regression models for $T|X$ — a possible future extension of this work.

Note that the suitability and fit of any particular model may be crucial in screen design. Here we use models that are appropriate to many situations and variables. However, we note later that our solutions are robust to departures from some of these modelling assumptions.

2.3.4 Screen design

The proposed screen design is simple and intuitive to the assumptions made about the relationship between the screening and performance variables. Theorem 2.2 shows that, under certain conditions, our design is preferable to one other design and Theorem 3.2 in section 3.3.1 gives conditions based on model and cost parameters for the designs obtained in section 2.2 to be optimal.

2.3.5 Cost structure

We assume a cost structure that is based on misclassification and screening costs.

The cost of wrongly rejecting a good item, a type I error, is c_r. Typically, c_r includes the cost of unnecessary repair and return of items or, in a different context, the costs of unnecessary patient stress and further treatment when a patient is wrongly diagnosed as unwell.

The cost of wrongly accepting a bad item, a type II error, is c_a. As well as any warranty, repair or handling costs incurred in industrial procedures, c_a may account for loss in sales due to a loss of reputation from selling below standard items. In medical applications c_a is the cost of wrongly diagnosing that a patient is healthy. If the patient is not intending to take further action that may reveal the illness, c_a may be very high.

The costs c_s and c_m represent those associated with any tests or procedures that are used to assess the screening and performance variables respectively.

Note that here the costs do not depend on the observed value of the screening variable X: they only depend on whether $X < v$, $v < X < w$ or $X > w$ and the value of the performance variable T. Costs dependent on the explicit value of X and the (unknown) value of a continuous performance variable are considered in Boys & Dunsmore (1987), Turkman & Amaral Turkman (1989) and, in the context of a two–stage screen, in Chapter 3. Boys & Dunsmore (1987) and Tang (1988b) also examine, in detail, commonly used loss functions that are functions of a continuous performance variable.

Tang (1988b) proposes a fixed cost for rejecting an item, independent of whether the item is of good quality or not. In quality control applications there is an arguable case in favour of a cost representing payments due to rejecting a bad item but it is intuitive that these payments would not be equivalent to the loss in revenue incurred from rejecting a good item. In applications in which all rejected items are thrown away there will be a loss in revenue from discarding good items but not bad items. On the other hand, when all rejected items are sent for repair, good items will not need repair whereas bad items will. Clearly, the cost set up should be chosen carefully to suit the application.

If an all encompassing rejection cost, c_{rej} say, is used to build a simple design similar to that given in section 2.2, the natural form of the X-screen would have the following two regions. If X was above a cut-off point, w say, then an item would be accepted, otherwise ($X \leq w$) the item would be rejected (if $c_{rej} < c_m$) or passed on to the T-screen (if $c_m < c_{rej}$). If $X < w$ and $c_m = c_{rej}$ there would be an arbitrary choice between rejection and measuring T to minimise cost.

Note that any procedure to elicit costs is simplified by the fact that it is only necessary to assess costs relative to one another. For example, it would be enough to define costs such as $c_a = 5\,c_r$, $c_s = c_r/100$ and $c_m = c_r/10$.

2.3.6 Choice of link function

Our use of a probit link function is mainly motivated by the closed form that results when a probit link is used in conjunction with a normality assumption for $\underline{\xi}$, see Lemma 2.1. In any case, we invoke the well-known robustness properties of link functions to suggest that our solutions are unlikely to be sensitive to our choice of a probit link for $T|X$. Boys & Dunsmore (1987) discuss in detail the use of a logistic model, which is a common alternative to the probit. Here

$$P(T=1|x,\underline{\xi}) = \frac{\exp(\xi_0+\xi_1 x)}{1+\exp(\xi_0+\xi_1 x)}.$$

Gauss–Hermite quadrature techniques, a normal approximation to the logistic distribution and a method described in an unpublished Valencia report by J M Bernardo are all proposed to evaluate $P(T=1|x)$ when the posterior distribution of $\underline{\xi}$ is bivariate normal. Another frequently used link function is the complementary log-log form. For a discussion of its use in this context see Boys (1985).

2.3.7 The distribution of the regression parameters

We assume the sample on which the screen will be based is large enough so that the posterior distribution $\pi(\underline{\xi})$ may be well approximated by its asymptotic normal form, $N_2(\underline{m}, \boldsymbol{S})$ say (Lindley (1961)). Here $\underline{m}$ is the maximum likelihood estimate of $\underline{\xi}$, and $\boldsymbol{S}$ is the inverse of Fisher's information matrix evaluated at $\underline{m}$. These can be calculated as follows. The details proceed from Dobson (1983).

Suppose we have a sample of n observations (t_i, x_i) on (T, X). The likelihood function assuming a probit link is

$$L(\underline{\xi}) = \prod_{i=1}^{n} \Phi(\xi_0+\xi_1 x_i)^{t_i} \{1-\Phi(\xi_0+\xi_1 x_i)\}^{1-t_i}$$

and hence the log-likelihood is

$$\ell(\underline{\xi}) = \sum_{i=1}^{n} [t_i \log \Phi(\xi_0+\xi_1 x_i) + (1-t_i)\log\{1-\Phi(\xi_0+\xi_1 x_i)\}].$$

Differentiating, the score with respect to ξ_j is:

$$U_j(\underline{\xi}) = \frac{\partial \ell}{\partial \xi_j} = \sum_{i=1}^{n} x_{ij}\phi(\xi_0 + \xi_1 x_i)\left[\frac{t_i}{\Phi(\xi_0 + \xi_1 x_i)} - \frac{1-t_i}{1-\Phi(\xi_0 + \xi_1 x_i)}\right],$$

where $j = 0, 1$ and $x_{i0} = 1$ and $x_{i1} = x_i$. Differentiating further gives the (j,k) element of the *observed* information matrix, $(j,k = 0,1)$, as

$$\begin{aligned} H_{jk}(\underline{\xi}) &= -\frac{\partial^2 \ell}{\partial \xi_j \partial \xi_k} \\ &= \sum_{i=1}^{n} x_{ij}x_{ik}(\xi_0 + \xi_1 x_i)\phi(\xi_0 + \xi_1 x_i)\left\{\frac{t_i}{\Phi(\xi_0 + \xi_1 x_i)} - \frac{1-t_i}{1-\Phi(\xi_0 + \xi_1 x_i)}\right\} \quad (2.12) \\ &+ \sum_{i=1}^{n} x_{ij}x_{ik}\left\{\phi(\xi_0 + \xi_1 x_i)\right\}^2\left[\frac{t_i}{\Phi(\xi_0 + \xi_1 x_i)^2} + \frac{1-t_i}{\left\{1-\Phi(\xi_0 + \xi_1 x_i)\right\}^2}\right]. \end{aligned}$$

Taking the expectation of this matrix over the distribution of $T|X,\underline{\xi}$ gives Fisher's information matrix. As $E_{T|X,\underline{\xi}}(T) = \Phi(\xi_0 + \xi_1 x_i)$, the components of Fisher's information matrix are given by, $(j,k = 0,1)$

$$I_{jk}(\underline{\xi}) = E_{T|X,\underline{\xi}}\left(-\frac{\partial^2 \ell}{\partial \xi_j \partial \xi_k}\right) = \sum_{i=1}^{n} x_{ij}x_{ik}\frac{\left\{\phi(\xi_0 + \xi_1 x_i)\right\}^2}{\Phi(\xi_0 + \xi_1 x_i)\left\{1-\Phi(\xi_0 + \xi_1 x_i)\right\}}. \quad (2.13)$$

Maximum likelihood estimates of $\underline{\xi}$ are found by solving either of the following iterative equations

$$\underline{m}^{(a)} = \underline{m}^{(a-1)} + H(\underline{m}^{(a-1)})^{-1}\underline{U}(\underline{m}^{(a-1)}), \quad (2.14)$$

or

$$\underline{m}^{(a)} = \underline{m}^{(a-1)} + I(\underline{m}^{(a-1)})^{-1}\underline{U}(\underline{m}^{(a-1)}), \quad (2.15)$$

where a indicates the ath approximation, $\underline{m}$ is the vector of estimates, $H(\underline{m})$ is the observed information matrix evaluated at $\underline{m}$ and $I(\underline{m})$ is Fisher's (expected) information matrix evaluated at $\underline{m}$. Equation (2.14) defines a Newton-Raphson scheme for obtaining the maximum likelihood estimate and equation (2.15) is an adjustment of (2.14) in which $H(\underline{m})$ is replaced by $I(\underline{m})$. The scheme defined by (2.15) is known as Fisher's method of scoring. Both procedures converge to the maximum likelihood estimate but we prefer the latter for two reasons. Firstly, our method requires both

the maximum likelihood estimates $\underline{m}$ and Fisher's information matrix evaluated at $\underline{m}$ — these will both be available at convergence of the scheme based on (2.15). Secondly, computer packages such as GLIM and S-PLUS implement schemes based on (2.15) and so we can obtain the information we need relatively painlessly.

A typical procedure that implements the rule (2.15) would involve taking starting values, $m_0^{(0)}$ and $m_1^{(0)}$, and solving the above equation to produce estimate updates. This procedure would be repeated until some convergence criteria is fulfilled (typically based on the size of $\underline{m}^{(a)} - \underline{m}^{(a-1)}$). The value of $\underline{m}$ that satisfies the convergence criteria, $\underline{m}^{(c)}$ say, is the maximum likelihood estimate of $\underline{\xi}$ to an accuracy dependent on the convergence criteria used. Also $I(\underline{m}^{(c)})$ is Fisher's information matrix evaluated at $\underline{m}^{(c)}$.

Boys & Glazebrook (1992) perform a robustness study that shows that their designs are close to optimal under quite large departures from the assumed normality of $\underline{\xi}$. In their study they show that when $\underline{\xi}$ follows a very skew log—normal distribution their solutions continue to perform well. As our designs generalise their designs it is clear that they will inherit this robustness property.

Alternatively, when the normality assumption is questionable, methods of finding $P(T = 1|x)$ are discussed in Boys & Dunsmore (1987). In particular, they refer to Laplace's method (see Tierney & Kadane (1986)).

2.3.8 Moments of the regression parameters

The screening solution for v and w given in Theorem 2.1 assumes that $s_1 = 1$, $m_0 = 0$ and $m_1 > 0$. The following details show that we can assume this without loss of generality.

If $s_1 \neq 1$ or $m_0 \neq 0$ we change parameters from (ξ_0, ξ_1) to (ξ_0', ξ_1'), where $\xi_0' = \xi_0 - \xi_1 m_0/m_1$ and $\xi_1' = \xi_1/s_1$. To ensure consistency of the probit model, a change in regression parameters is accompanied with a transformation in the explanatory (screening) variable. The linear predictor in the probit regression model

with parameters (ξ_0, ξ_1) is

$$\begin{aligned} \xi_0 + \xi_1 x &= \xi_0 - \frac{m_0 \xi_1}{m_1} + \frac{\xi_1}{s_1}\left\{ s_1 \left(x + \frac{m_0}{m_1} \right) \right\} \\ &= \xi_0' + \xi_1' x', \end{aligned}$$

where $x' = s_1(x + m_0/m_1)$. Therefore, when using Theorem 2.1, we can use the moments of (ξ_0', ξ_0'), to obtain optimal choices of

$$v' = s_1(v + m_0/m_1) \qquad \text{and} \qquad w' = s_1(w + m_0/m_1). \tag{2.16}$$

It is straightforward to verify that the mean and variance of ξ_0' and ξ_1' satisfy the conditions assumed in Theorem 2.1, that is, $m_0' \equiv E(\xi_0') = E(\xi_0) - m_0 E(\xi_1)/m_1 = 0$ and $s_1'^2 \equiv Var(\xi_1') = Var(\xi_1)/s_1 = 1$. The other mean and variance parameters of (ξ_0', ξ_1') are

$$\begin{aligned} m_1' \equiv E(\xi_1') &= E(\xi_1)/s_1 \\ &= m_1/s_1 \\ s_0'^2 \equiv Var(\xi_0') &= Var(\xi_0) + m_0^2 Var(\xi_1)/m_1^2 - 2m_0 Cov(\xi_0, \xi_1)/m_1 \\ &= s_0^2 + m_0^2 s_1^2/m_1^2 - 2m_0 r s_0 s_1/m_1 \\ r's_0' \equiv Cov(\xi_0', \xi_1') &= Cov(\xi_0, \xi_1)/s_1 - m_0 Var(\xi_1)/s_1 m_1 \\ &= r s_0 - m_0 s_1/m_1 \end{aligned}$$

If $m_1 < 0$ we change variables from (ξ_0, ξ_1) to $(\xi_0, -\xi_1)$. The corresponding data transformation is,

$$(v, w, m_0, m_1, s_0^2, r s_0 s_1, s_1^2) \to (-v, -w, m_0, -m_1, s_0^2, -r s_0 s_1, s_1^2).$$

By a combination of the two changes in regression variables above we can ensure that $s_1^2 = 1$, $m_0 = 0$ and $m_1 > 0$. The case $m_1 = 0$ will always result in a decision not to screen and is of little interest here.

2.3.9 Proof of Theorem 2.2

Let $\tilde{\mathcal{K}}(v, w)$ denote the screening cost per item for a screening procedure of form (II) and $\mathcal{K}(v, w)$ denote that for a procedure of form (I), where $-\infty \le v, w \le \infty$. Then

$$\tilde{\mathcal{K}}(v, w) = c_r P(T = 1,\ X \ge v) + c_a P(T = 0,\ X < w) + c_m P(w < X < v) + c_s I(v, w).$$

Note that when $\tilde{\mathcal{K}}(v,w)$ is minimised in both limits, this corresponds to a decision not to screen and so a procedure of the form (II) is not optimal.

When $\tilde{\mathcal{K}}(v,w)$ is not minimised in the limits, the turning points of

$$\tilde{\mathcal{K}}'(v,w) = \tilde{\mathcal{K}}(v,w) - c_s I(v,w)$$

will coincide with the turning points of $\tilde{\mathcal{K}}(v,w)$. We denote as $\mathcal{K}'(v,w)$ the equivalent cost for a procedure of form (I), that is $\mathcal{K}'(v,w) = \mathcal{K}(v,w) - c_s I(v,w)$. Then

$$\begin{aligned}
&\tilde{\mathcal{K}}'(v,w) \\
&\quad = c_r P(T=1,\ X \geq v) + c_a P(T=0,\ X < w) + c_m P(w < X < v) \\
&\quad = c_r P(T=1) + c_a P(T=0) + c_m P(w < X < v) + c_m P(v < X < w) - \mathcal{K}'(v,w)
\end{aligned}$$

Here we shall use the convention

$$P(w < X < v) = -P(v < X < w), \quad v \leq w,$$

so that

$$\tilde{\mathcal{K}}'(v,w) = c_r P(T=1) + c_a P(T=0) - \mathcal{K}'(v,w).$$

Hence

$$\inf_{v,w} \tilde{\mathcal{K}}'(v,w) = c_r P(T=1) + c_a P(T=0) - \sup_{v,w} \mathcal{K}'(v,w). \tag{2.17}$$

However, in the proof of Theorem 2.1 we analysed the turning points of $\mathcal{K}'(v,w)$ and found that in the situation where a global minimisation is achieved (possibly in the limit)

$$\begin{aligned}
\sup_{v,w} \mathcal{K}'(v,w) \ = \ \max \Bigg\{ & \lim_{\substack{v\to-\infty \\ w\to-\infty}} \mathcal{K}'(v,w),\ \lim_{\substack{v\to-\infty \\ w\to\infty}} \mathcal{K}'(v,w), \\
& \lim_{\substack{v\to\infty \\ w\to-\infty}} \mathcal{K}'(v,w),\ \lim_{\substack{v\to\infty \\ w\to\infty}} \mathcal{K}'(v,w) \Bigg\}
\end{aligned}$$

and therefore

$$\begin{aligned}
\inf_{v,w} \tilde{\mathcal{K}}'(v,w) \ = \ \min \Bigg\{ & \lim_{\substack{v\to-\infty \\ w\to-\infty}} \tilde{\mathcal{K}}'(v,w),\ \lim_{\substack{v\to-\infty \\ w\to\infty}} \tilde{\mathcal{K}}'(v,w), \\
& \lim_{\substack{v\to\infty \\ w\to-\infty}} \tilde{\mathcal{K}}'(v,w),\ \lim_{\substack{v\to\infty \\ w\to\infty}} \tilde{\mathcal{K}}'(v,w) \Bigg\}
\end{aligned} \tag{2.18}$$

This is a decision not to use the X-screen and will be preferable to any screening procedure of form (II).

In cases where a local minimum of $\mathcal{K}'(v,w)$ is found with respect to either one or both v and w there is also a local maximum. If this maximum is not a global maximum with respect to either v or w then (2.18) holds. However, if this maximum is a global maximum with respect to one of v and w, say at $v = \tilde{v}$, then

$$\inf_{v,w} \tilde{\mathcal{K}}'(v,w) = \min\left\{ \lim_{w\to-\infty} \tilde{\mathcal{K}}'(\tilde{v},w), \lim_{w\to\infty} \tilde{\mathcal{K}}'(\tilde{v},w) \right\}.$$

If the maximum is a global maximum with respect to both v and w, say at $v = \tilde{v}$ and $w = \tilde{w}$ then

$$\inf_{v,w} \tilde{\mathcal{K}}'(v,w) = \tilde{\mathcal{K}}'(\tilde{v},\tilde{w}).$$

It was shown in the proof of Theorem 2.1 that

$$\begin{aligned} \tilde{v} = v^- \quad &\text{and/or} \quad \tilde{w} = w^- \quad (v^- > 0, w^- > 0) \quad \text{if } r < 0, \\ \tilde{v} = v^+ \quad &\text{and/or} \quad \tilde{w} = w^+ \quad (v^+ < 0, w^+ < 0) \quad \text{if } r > 0, \end{aligned}$$

where $v^- = u_1^-$, $w^- = u_2^-$, $v^+ = u_1^+$ and $w^+ = u_2^+$, and u_i^- and u_i^+ are as defined in the statement of Theorem 2.1.

Consider the case $r < 0$. Since $\psi(x)$ is symmetric about $x = 0$ and Φ is an increasing function it follows that

$$\begin{aligned} &\int_{v^-}^{\infty} \psi(v) \left[c_r \Phi\left\{ \frac{m_1 v}{(1+s_0^2+2vrs_0+v^2)^{1/2}} \right\} - c_m \right] dv \\ &\qquad \geq \int_{-\infty}^{-v^-} \psi(v) \left[c_r \Phi\left\{ \frac{m_1 v}{(1+s_0^2+2vrs_0+v^2)^{1/2}} \right\} - c_m \right] dv, \end{aligned}$$

that is

$$\int_{v^-}^{\infty} \frac{d}{dv}\{\mathcal{K}'(v,w)\}dv \geq \int_{-\infty}^{-v^-} \frac{d}{dv}\{\mathcal{K}'(v,w)\}dv.$$

However, in the proof of Theorem 2.1 we saw that v^+ is the smaller of two turning points of $\mathcal{K}'(v,w)$ and is a minimum, so

$$\frac{d}{dv}\{\mathcal{K}'(v,w)\} < 0, \qquad -v^- < v < v^+,$$

and hence

$$\int_{v^-}^{\infty} \frac{d}{dv}\{\mathcal{K}'(v,w)\}dv \geq \int_{-\infty}^{v^+} \frac{d}{dv}\{\mathcal{K}'(v,w)\}dv. \tag{2.19}$$

An equivalent result for w holds:

$$\int_{w^-}^{\infty} \frac{d}{dw}\{\mathcal{K}'(v,w)\}dw \geq \int_{-\infty}^{w^+} \frac{d}{dw}\{\mathcal{K}'(v,w)\}dw. \tag{2.20}$$

When $r > 0$ it follows that

$$\int_{-v^+}^{\infty} \frac{d}{dv}\{\mathcal{K}'(v,w)\}dv \geq \int_{-\infty}^{v^+} \frac{d}{dv}\{\mathcal{K}'(v,w)\}dv.$$

Here v^- is the larger of two turning points of $\mathcal{K}'(v,w)$ and is a minimum and hence

$$\frac{d}{dv}\{\mathcal{K}'(v,w)\} > 0, \quad v^- < v < -v^+,$$

and equation (2.19) follows. Similarly equation (2.20) holds when $r > 0$.

We now look at all cases when a (local) minimum of $\tilde{\mathcal{K}}'(v,w)$ occurs and in each case find a screen of form (I) that is at least as good as the optimal screen of form (II).

(i) $\inf_{v,w} \tilde{\mathcal{K}}'(v,w) = \tilde{\mathcal{K}}'(v^-,w^-)$

Combining equation (2.19) with (2.20) and noting that $\frac{d}{dv}\{\mathcal{K}'(v,w)\}$ is constant for all w and $\frac{d}{dw}\{\mathcal{K}'(v,w)\}$ is constant for all v, gives

$$\begin{aligned}\lim_{w\to\infty} \int_{v^-}^{\infty} \frac{d}{dv}\{\mathcal{K}'(v,w)\}dv + \int_{w^-}^{\infty} \frac{d}{dw}\{\mathcal{K}'(v^-,w)\}dw \\ \geq \int_{-\infty}^{v^+} \frac{d}{dv}\{\mathcal{K}'(v,w^+)\}dv + \lim_{v\to-\infty} \int_{-\infty}^{w^+} \frac{d}{dw}\{\mathcal{K}'(v,w)\}dw.\end{aligned}$$

Hence

$$\lim_{\substack{v\to\infty \\ w\to\infty}} \mathcal{K}'(v,w) - \mathcal{K}'(v^-,w^-) \geq \mathcal{K}'(v^+,w^+) - \lim_{\substack{v\to-\infty \\ w\to-\infty}} \mathcal{K}'(v,w),$$

and

$$c_r P(T=1) - \mathcal{K}'(v^-,w^-) \geq \mathcal{K}'(v^+,w^+) - c_a P(T=0). \tag{2.21}$$

From this we see that

$$c_r P(T=1) + c_a P(T=0) - \mathcal{K}'(v^-, w^-) \geq \mathcal{K}'(v^+, w^+),$$

and it follows from equation (2.17) that

$$\tilde{\mathcal{K}}'(v^-, w^-) = \inf_{v,w} \tilde{\mathcal{K}}'(v, w) \geq \mathcal{K}'(v^+, w^+).$$

(ii) $\inf_{v,w} \tilde{\mathcal{K}}'(v, w) = \tilde{\mathcal{K}}'(v^+, w^+)$

Rearranging equation (2.21) gives

$$c_r P(T=1) + c_a P(T=0) - \mathcal{K}'(v^+, w^+) \geq \mathcal{K}'(v^-, w^-),$$

and so

$$\tilde{\mathcal{K}}'(v^+, w^+) = \inf_{v,w} \tilde{\mathcal{K}}'(v, w) \geq \mathcal{K}'(v^-, w^-).$$

(iii) $\inf_{v,w} \tilde{\mathcal{K}}'(v, w) = \min\left\{\lim_{w\to-\infty} \tilde{\mathcal{K}}'(v^-, w), \lim_{w\to\infty} \tilde{\mathcal{K}}'(v^-, w)\right\}$

From (2.19) and remembering that $\frac{d}{dv}\{\mathcal{K}'(v, w)\}$ is constant for all w

$$\lim_{w\to-\infty} \int_{v^-}^{\infty} \frac{d}{dv}\{\mathcal{K}'(v, w)\}\, dv \geq \lim_{w\to\infty} \int_{-\infty}^{v^+} \frac{d}{dv}\{\mathcal{K}'(v, w)\}\, dv$$

and so

$$\lim_{\substack{v\to\infty \\ w\to-\infty}} \mathcal{K}'(v, w) - \lim_{w\to-\infty} \mathcal{K}'(v^-, w) \geq \lim_{w\to\infty} \mathcal{K}'(v^+, w) - \lim_{\substack{v\to-\infty \\ w\to\infty}} \mathcal{K}'(v, w).$$

Hence

$$c_r P(T=1) + c_a P(T=0) - c_m - \lim_{w\to-\infty} \mathcal{K}'(v^-, w) \geq \lim_{w\to\infty} \mathcal{K}'(v^+, w) - c_m \tag{2.22}$$

and therefore

$$\lim_{w\to-\infty} \tilde{\mathcal{K}}'(v^-, w) \geq \lim_{w\to\infty} \mathcal{K}'(v^+, w). \tag{2.23}$$

Similarly

$$\lim_{w\to\infty}\int_{v^-}^{\infty}\frac{d}{dv}\{\mathcal{K}'(v,w)\}\,dv \geq \lim_{w\to-\infty}\int_{-\infty}^{v^+}\frac{d}{dv}\{\mathcal{K}'(v,w)\}\,dv$$

and

$$\lim_{\substack{v\to\infty\\w\to\infty}}\mathcal{K}'(v,w)-\lim_{w\to\infty}\mathcal{K}'(v^-,w)\geq\lim_{w\to-\infty}\mathcal{K}'(v^+,w)-\lim_{\substack{v\to-\infty\\w\to-\infty}}\mathcal{K}'(v,w).$$

Hence

$$c_aP(T=0)-\lim_{w\to\infty}\mathcal{K}'(v^-,w)\geq\lim_{w\to-\infty}\mathcal{K}'(v^+,w)-c_rP(T=1) \tag{2.24}$$

and therefore

$$\lim_{w\to\infty}\tilde{\mathcal{K}}'(v^-,w)\geq\lim_{w\to-\infty}\mathcal{K}'(v^+,w). \tag{2.25}$$

From equations (2.23) and (2.25) we conclude that

$$\begin{aligned}\min\left\{\lim_{w\to-\infty}\tilde{\mathcal{K}}'(v^-,w),\lim_{w\to\infty}\tilde{\mathcal{K}}'(v^-,w)\right\}&=\inf_{v,w}\tilde{\mathcal{K}}'(v,w)\\&\geq\min\left\{\lim_{w\to-\infty}\mathcal{K}'(v^+,w),\lim_{w\to\infty}\mathcal{K}'(v^+,w)\right\}.\end{aligned}$$

(iv) $\inf_{v,w}\tilde{\mathcal{K}}'(v,w)=\min\left\{\lim_{w\to-\infty}\tilde{\mathcal{K}}'(v^+,w),\lim_{w\to\infty}\tilde{\mathcal{K}}'(v^+,w)\right\}$

Rearranging equation (2.22) gives

$$c_rP(T=1)+c_aP(T=0)-\lim_{w\to\infty}\mathcal{K}'(v^+,w)\geq\lim_{w\to-\infty}\mathcal{K}'(v^-,w)$$

and so

$$\lim_{w\to\infty}\tilde{\mathcal{K}}'(v^+,w)\geq\lim_{w\to-\infty}\mathcal{K}'(v^-,w). \tag{2.26}$$

Similarly equation (2.24) gives

$$c_rP(T=1)+c_aP(T=0)-\lim_{w\to-\infty}\mathcal{K}'(v^+,w)\geq\lim_{w\to\infty}\mathcal{K}'(v^-,w)$$

and hence

$$\lim_{w\to-\infty}\tilde{\mathcal{K}}'(v^+,w)\geq\lim_{w\to\infty}\mathcal{K}'(v^-,w). \tag{2.27}$$

Here, from (2.26) and (2.27) we know that

$$\min\left\{\lim_{w\to-\infty}\tilde{\mathcal{K}}'(v^+,w),\lim_{w\to\infty}\tilde{\mathcal{K}}'(v^+,w)\right\}=\inf_{v,w}\tilde{\mathcal{K}}'(v,w)$$
$$\geq\min\left\{\lim_{w\to-\infty}\mathcal{K}'(v^-,w),\lim_{w\to\infty}\mathcal{K}'(v^-,w)\right\}.$$

(v) $\inf_{v,w}\tilde{\mathcal{K}}'(v,w)=\min\left\{\lim_{v\to-\infty}\tilde{\mathcal{K}}'(v,w^-),\lim_{v\to\infty}\tilde{\mathcal{K}}'(v,w^-)\right\}$

From (2.20) and noting that $\frac{d}{dw}\{\mathcal{K}'(v,w)\}$ is constant for all v

$$\lim_{v\to\infty}\int_{w^-}^{\infty}\frac{d}{dw}\{\mathcal{K}'(v,w)\}\,dw\geq\lim_{v\to-\infty}\int_{-\infty}^{w^+}\frac{d}{dw}\{\mathcal{K}'(v,w)\}\,dw$$

and so

$$\lim_{\substack{v\to\infty\\w\to\infty}}\mathcal{K}'(v,w)-\lim_{v\to\infty}\mathcal{K}'(v,w^-)\geq\lim_{v\to-\infty}\mathcal{K}'(v,w^+)-\lim_{\substack{v\to-\infty\\w\to-\infty}}\mathcal{K}'(v,w).$$

Hence

$$c_rP(T=1)-\lim_{v\to\infty}\mathcal{K}'(v,w^-)\geq\lim_{v\to-\infty}\mathcal{K}'(v,w^+)-c_aP(T=0), \tag{2.28}$$

and therefore

$$\lim_{v\to\infty}\tilde{\mathcal{K}}'(v,w^-)\geq\lim_{v\to-\infty}\mathcal{K}'(v,w^+). \tag{2.29}$$

Similarly

$$\lim_{v\to-\infty}\int_{w^-}^{\infty}\frac{d}{dw}\{\mathcal{K}'(v,w)\}\,dw\geq\lim_{v\to\infty}\int_{-\infty}^{w^+}\frac{d}{dw}\{\mathcal{K}'(v,w)\}\,dw$$

and

$$\lim_{\substack{v\to-\infty\\w\to\infty}}\mathcal{K}'(v,w)-\lim_{v\to-\infty}\mathcal{K}'(v,w^-)\geq\lim_{v\to\infty}\mathcal{K}'(v,w^+)-\lim_{\substack{v\to\infty\\w\to-\infty}}\mathcal{K}'(v,w).$$

Hence

$$c_m - \lim_{v\to-\infty} \mathcal{K}'(v,w^-) \geq \lim_{v\to\infty} \mathcal{K}'(v,w^+) - c_r P(T=1) - c_a P(T=0) + c_m. \tag{2.30}$$

and it follows that

$$\lim_{v\to-\infty} \tilde{\mathcal{K}}'(v,w^-) \geq \lim_{v\to\infty} \mathcal{K}'(v,w^+). \tag{2.31}$$

Therefore (2.29) and (2.31) show that

$$\min\left\{\lim_{v\to-\infty} \tilde{\mathcal{K}}'(v,w^-), \lim_{v\to\infty} \tilde{\mathcal{K}}'(v,w^-)\right\} = \inf_{v,w} \tilde{\mathcal{K}}'(v,w)$$
$$\geq \min\left\{\lim_{v\to-\infty} \mathcal{K}'(v,w^+), \lim_{v\to\infty} \mathcal{K}'(v,w^+)\right\}.$$

(vi) $\inf_{v,w} \tilde{\mathcal{K}}'(v,w) = \min\left\{\lim_{v\to-\infty} \tilde{\mathcal{K}}'(v,w^+), \lim_{v\to\infty} \tilde{\mathcal{K}}'(v,w^+)\right\}$

Rearranging (2.30) gives

$$c_r P(T=1) + c_a P(T=0) - \lim_{v\to\infty} \mathcal{K}'(v,w^+) \geq \lim_{v\to-\infty} \mathcal{K}'(v,w^-)$$

and so

$$\lim_{v\to\infty} \tilde{\mathcal{K}}'(v,w^+) \geq \lim_{v\to-\infty} \mathcal{K}'(v,w^-). \tag{2.32}$$

Equation (2.28) gives

$$c_r P(T=1) + c_a P(T=0) - \lim_{v\to-\infty} \mathcal{K}'(v,w^+) \geq \lim_{v\to\infty} \mathcal{K}'(v,w^-)$$

and

$$\lim_{v\to-\infty} \tilde{\mathcal{K}}'(v,w^+) \geq \lim_{v\to\infty} \mathcal{K}'(v,w^-). \tag{2.33}$$

Therefore we conclude from equations (2.32) and (2.33) that

$$\min\left\{\lim_{v\to-\infty} \tilde{\mathcal{K}}'(v,w^+), \lim_{v\to\infty} \tilde{\mathcal{K}}'(v,w^+)\right\} = \inf_{v,w} \tilde{\mathcal{K}}'(v,w)$$
$$\geq \min\left\{\lim_{v\to-\infty} \mathcal{K}'(v,w^-), \lim_{v\to\infty} \mathcal{K}'(v,w^-)\right\}.$$

Now, recall that

$$\tilde{\mathcal{K}}(v,w) = \tilde{\mathcal{K}}'(v,w) + c_s I(v,w).$$

Hence the design that minimises cost $\tilde{\mathcal{K}}'(v,w)$ also minimises cost $\tilde{\mathcal{K}}(v,w)$ except in the case when $\tilde{\mathcal{K}}(v,w)$ is minimised in the limits of both v and w. However, this corresponds to a decision not to screen and is not of interest here. In cases when the minimum of $\tilde{\mathcal{K}}(v,w)$ occurs when at least one of v and w do not take limiting values we have that

$$\begin{aligned} \tilde{\mathcal{K}}(v,w) &= \tilde{\mathcal{K}}'(v,w) + c_s \\ \mathcal{K}(v,w) &= \mathcal{K}'(v,w) + c_s \end{aligned}$$

and so if we find that $\mathcal{K}'(\tilde{v},\tilde{w}) \leq \inf_{v,w} \tilde{\mathcal{K}}'(v,w)$ for some $\tilde{v}$ and $\tilde{w}$, we also know that $\mathcal{K}(\tilde{v},\tilde{w}) \leq \inf_{v,w} \tilde{\mathcal{K}}(v,w)$. Hence in each of the cases (i)—(vi) above we have shown that a design of form (I) is at least as good as the optimal design of form (II). □

2.4 A numerical example

To illustrate the method, we consider the construction of an optimal screen for Conn's syndrome, a rare syndrome of hypertension. It is known that the syndrome is due to either a benign tumour ($T = 1$) or to a more diffuse condition of the adrenal glands ($T = 0$). The cause can be determined precisely by means of an exploratory operation. Screening for these different causes is important as the treatment given differs radically between them: a *total adrenalectomy* is required for a tumour whereas only drug therapy is used for the gland condition. We will focus on the design of a screen to identify patients to be given a total adrenalectomy, that is, the $T = 1$ group. We have available measurements of the concentration of potassium in the blood plasma on which we can base the screen. We use data given in Aitchison & Dunsmore (1975), pp.10–11. Inspection of the data reveals that a linear probit regression for $T|X$ is plausible when using log-concentrations as covariates. Also, the calculation of Bayes cost is more straightforward if we use log-concentrations as a

Table 2.1: Cases of Conn's syndrome

Cause of syndrome	log(concentration) of K X^*				Standardised screening variable X			
$T=1$	0.833	1.131	1.099	1.030	−1.275	0.136	−0.019	−0.346
	1.281	1.131	0.916	0.916	0.842	0.136	−0.881	−0.881
	0.876	1.065	0.833	0.789	−1.074	−0.180	−1.275	−1.485
	0.993	1.131	1.065	1.131	−0.517	0.136	−0.180	0.136
	0.642	1.308	0.789	0.993	−2.178	0.972	−1.485	−0.517
$T=0$	1.459	1.163	1.281	1.099	1.682	0.286	0.842	−0.019
	1.435	1.224	1.281	1.335	1.571	0.572	0.842	1.098
	1.194	1.281	1.482		0.431	0.842	1.791	

normality assumption for X is then plausible. Data from 31 cases of Conn's syndrome is given in Table 2.1. In these cases the level of potassium has been measured and the cause of syndrome is known.

Throughout the thesis we shall assume that any screening variables are standardised (mean zero and variance one). This will allow simple comparisons between the estimates of the parameters of any probit regression models that are used. We standardise the screening variable by subtracting the sample mean and dividing by the adjusted sample standard deviation. So here the screening variable is given by

$$X = \frac{X^* - 1.102}{0.2115}.$$

The results of this transformation are given in Table 2.1. Note that large values of X^* and hence X occur more frequently when $T = 0$. The misclassification and measurement costs are:

c_r = the cost of wrongly diagnosing that a patient has bilateral hyperplasia

c_a = the cost of wrongly diagnosing that a patient has a benign tumour

c_s = the cost of measuring the concentration of potassium in blood plasma

c_m = the cost of correct diagnosis using surgery

We will assume that these (relative) costs have been elicited and are as follows:

$$c_r = c_r, \qquad c_a = 3\,c_r/4, \qquad c_m = 3\,c_r/40, \qquad c_s = 0.$$

These costs reflect the fact that the clinician would prefer not to identify a tumour patient as requiring drug therapy and that the exploratory operation is only moderately expensive. Note also that these costs satisfy equation (2.9). The cost of measuring the screening variable is set at zero indicating that the measurement may already be available through a routine test on the patient or c_s is so small relative to c_r, c_a and c_m that it can be thought of as negligible. In any case the only effect of c_s on screen design is on the decision of whether to perform the X-screen.

Fitting a probit regression model to the data using S-PLUS gives:

$$\underline{m} = \begin{pmatrix} 0.8812 \\ -1.863 \end{pmatrix}, \qquad \boldsymbol{S} = \begin{pmatrix} 0.1612 & -0.1546 \\ -0.1546 & 0.4103 \end{pmatrix}. \tag{2.34}$$

Tests of model adequacy and goodness of fit are omitted. In order to use Theorem 2.1 we change regression parameters so that $m_0 = 0$, $s_1 = 1$ and $m_1 > 0$. Following section 2.3.8, moments for the new variables are,

$$\widehat{E(\underline{\xi}')} = \begin{pmatrix} m_0' \\ m_1' \end{pmatrix} = \begin{pmatrix} 0 \\ 2.909 \end{pmatrix}; \qquad \widehat{Var(\underline{\xi}')} = \begin{pmatrix} 0.1067 & -0.06165 \\ -0.06165 & 1 \end{pmatrix}. \tag{2.35}$$

From equations (2.6),

$$c_1 = \Phi^{-1}\left(\frac{c_m}{c_r}\right) = \Phi^{-1}\left(\frac{3}{40}\right) = -1.440, \qquad c_2 = \Phi^{-1}\left(1 - \frac{c_m}{c_a}\right) = \Phi^{-1}\left(\frac{9}{10}\right) = 1.282.$$

As $m_1' > -c_1 > c_2 > 0$ we refer to (b)(i) in Theorem 2.1 to find the optimal lower (transformed) cut-off point v' and (a)(i) for the upper (transformed) cut-off point w'. Here

$$(v', w') = (u_1^-, u_2^+) = (-0.6193, 0.5016) \tag{2.36}$$

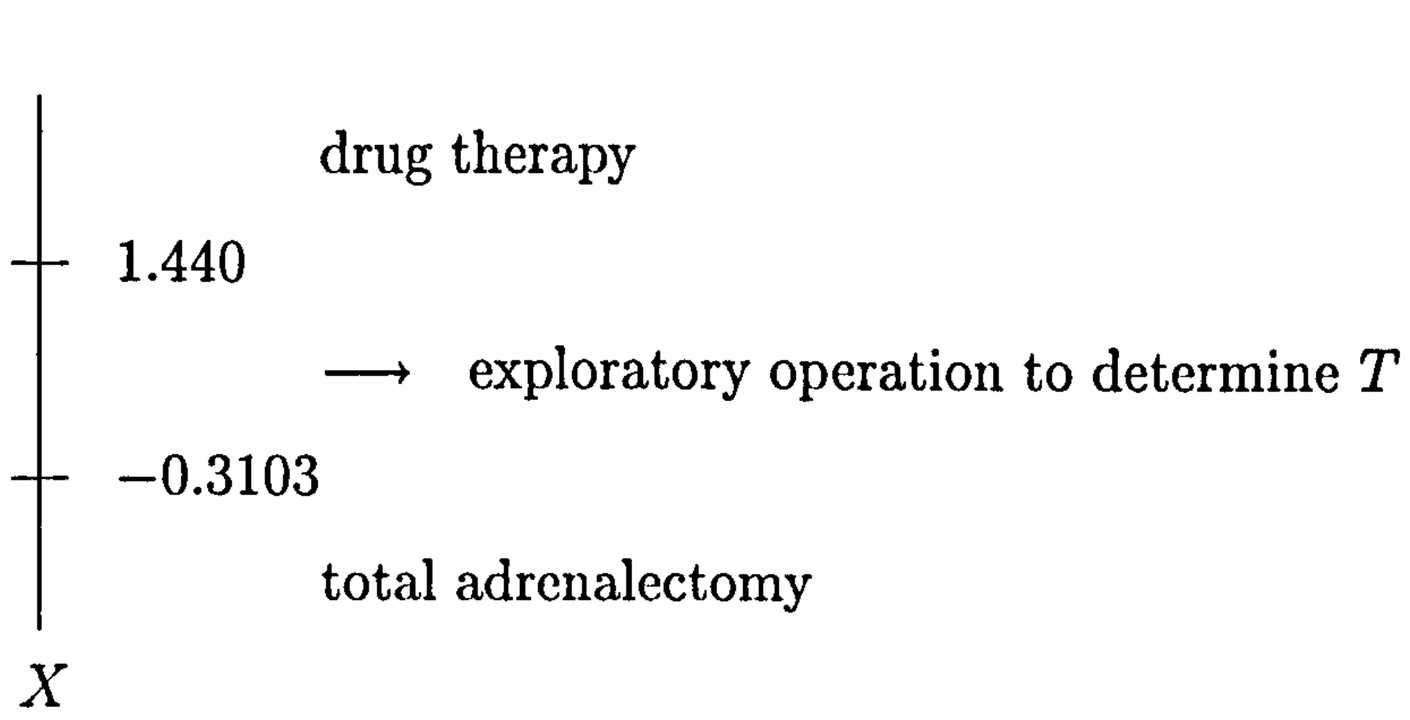


Figure 2.2: The optimal two–stage screen for the Conn's syndrome data.

Now we reverse the change of variables, returning to the original regression parameters (ξ_0, ξ_1) that relate to the screening variable X. Following comment 5 in section 2.2 the Bayes design (v^*, w^*) is given by

$$\begin{aligned} v^* &= -\left(\frac{v'}{s_1} - \frac{m_0}{m_1}\right) = 1.440; \\ w^* &= -\left(\frac{w'}{s_1} - \frac{m_0}{m_1}\right) = -0.3103. \end{aligned}$$

Note that we have a problem in which the data indicates that X small implies that $T = 1$ and so we assume that the screen is of the form: Accept $(T = 1)$ if the patient has $X < w$, reject $(T = 0)$ if the patient has $X > v$ and measure T if $w < X < v$. As the screening cost c_s is zero and the above solution represents a global minimum of $\mathcal{K}'(v, w)$, it is not necessary to check that our design is cheaper than not using the X-screen (see comment 4, section 2.2). The Bayes design is presented graphically in Figure 2.2.

Submitting the 31 cases given in Table 2.1 to the screen shown in Figure 2.2, results in the following classifications:

Actual	Total	X–screen classification		
		$T = 1$	$T = 0$	measure T
$T = 1$	20	11	0	9
$T = 0$	11	0	3	8

Note that none of the cases would have been misdiagnosed. A better assessment of

the performance of the screen would be obtained via the use of the design on further observations of (T, X) or by a simulation study.

Chapter 3

A class of Bayes optimal two–stage screens

3.1 Introduction

In the previous chapter, we constructed two–stage screens designed to assess whether an item will satisfy some pre-defined criteria. The criteria are satisfied if $\underline{T} \in C_{\underline{T}}$ where $\underline{T}$ is the measurement of a performance variable on the item. At the first stage of the screen, the measurement of a covariate X was used to classify some items and those that remain unsentenced are passed to the second stage where a "gold standard" measurement of $\underline{T}$ is taken. For the case in which T is binary and $T|X$ is modelled by a probit regression, we obtained Bayes optimal designs using a decision theoretic approach based on misclassification costs and the costs of measuring the covariate and the performance variable. We also assumed that the screen took a pre–specified form and optimised within that class of screens. Hence, Chapter 2 was concerned with building designs useful in many scenarios but which cover only a subclass of the many stochastic structures that may be useful in modelling situations in which screening may be applied.

In this chapter we suppose that for each item we can measure a batch of covariates

$\underline{X}$ on which to base the first stage of the screen. We assume a general stochastic structure for $(\underline{T}, \underline{X})$ and optimise without assuming a form for the screen. Also, we adopt a more general cost set-up in which the misclassification costs may be functions of $\underline{X}$ and $\underline{T}$. The cost structure assumed in Chapter 2 is retained as a special case. In Chapter 2 we saw that the cost associated with measuring $\underline{X}$ is only important when deciding whether it is optimal to perform the $\underline{X}$–stage of the screen. Here we shall assume *a priori* that it will always be better to operate the first stage of the screen than not to do so. Hence we do not include a cost for measuring $\underline{X}$.

Section 3.2 describes the form of the optimal Bayes two-stage screen for a fully general model on $(\underline{T}, \underline{X})$. The solution is straightforward and some of the results of Turkman & Amaral Turkman (1989) can be recovered as special cases. The simplicity of the solution is illustrated within a probit regression model in section 3.3. We have already seen in Chapter 2 that this model combines a reasonable level of generality with a capacity to yield simple, intuitive screen designs when the posterior distribution for the regression parameters takes its asymptotic Normal form. Section 3.4 considers Bayes optimal screens for situations when, due to limited resources, there is an upper bound on the proportion of patients (or items) that can have their performance variable measured. The solution to this constrained problem is shown to take the same form as that of the unconstrained problem described earlier, and is again illustrated within a probit regression model in section 3.5.

3.2 Bayes optimal two–stage screens

Suppose we are interested in screening for items with attributes described by $\underline{T} \in C_{\underline{T}}$ using a p–dimensional screening variable $\underline{X}$. We shall denote the sample spaces of $\underline{T}$ and $\underline{X}$ by $\Omega_{\underline{T}}$ and $\Omega_{\underline{X}}$ respectively. The set of attributes $C_{\underline{T}}$ is a subset of $\Omega_{\underline{T}}$, and $\underline{T}$ may be univariate or multivariate. A two–stage screen partitions $\Omega_{\underline{X}}$ into the three regions $\{\Omega_A, \Omega_R, \Omega_M\}$. At the first stage of the screen $\underline{X}$ is measured and if

$$\begin{aligned}
\underline{X} \in \Omega_A &\longrightarrow \text{item is accepted} \\
\underline{X} \in \Omega_R &\longrightarrow \text{item is rejected} \\
\underline{X} \in \Omega_M = \Omega_{\underline{X}} \setminus \{\Omega_A \cup \Omega_R\} &\longrightarrow \text{item is passed onto the second stage.}
\end{aligned}$$

At the second stage, $\underline{T}$ is measured and the item classified accordingly. Recall that classification based on the screening variable $\underline{X}$ is open to error and so it may be economically advisable to incorporate such a second stage based on $\underline{T}$ even though it may be expensive to do so.

The costs associated with screening are determined by c_a, c_r and c_m, all functions from $\Omega_{\underline{T}} \times \Omega_{\underline{X}}$ into $\mathbb{R}^+$. These yield the costs associated with acceptance, rejection and measurement of $\underline{T}$ respectively. We do not include a cost associated with the measurement of $\underline{X}$. We shall suppose that

$$c_a(\underline{t}, \underline{x}) = 0, \qquad \underline{t} \in C_{\underline{T}}, \underline{x} \in \Omega_{\underline{X}} \qquad \text{and} \qquad c_r(\underline{t}, \underline{x}) = 0, \qquad \underline{t} \notin C_{\underline{T}}, \underline{x} \in \Omega_{\underline{X}},$$

that is, no costs are incurred when accepting an item which meets the standard or when rejecting an item which does not. This assumption may not be reasonable in some applications and is discussed in section 2.3.5. Finally, we shall refer to the following choices of c_a, c_r and c_m as the *standard case*:

$$c_a(\underline{t}, \underline{x}) = c_a I(\underline{t} \notin C_{\underline{T}}), \qquad c_r(\underline{t}, \underline{x}) = c_r I(\underline{t} \in C_{\underline{T}}), \qquad c_m(\underline{t}, \underline{x}) = c_m,$$

where c_a, c_r and c_m are constants and

$$I(\underline{t} \in A) = \begin{cases} 1, & \underline{t} \in A \\ 0, & \text{otherwise.} \end{cases}$$

Chapter 2 assumes this simple cost structure. It is based on the assumption that all items with $\underline{T} \in C_{\underline{T}}$ are equally valuable and all items with $\underline{T} \notin C_{\underline{T}}$ are equally worthless. Tang (1988b) and other authors describe loss functions that are linear or quadratic in the distance of $\underline{T}$ from the boundary of $C_{\underline{T}}$ and argue that such loss functions are more realistic. The method given by this chapter is capable of incorporating such loss functions.

The relationship between $\underline{T}$ and $\underline{X}$ is given by the conditional model $\underline{T}|\underline{X},\underline{\xi}$ where $\underline{\xi}$ are unknown parameters with posterior distribution $\pi(\underline{\xi})$. The goal is to choose a screen to minimise the Bayes cost

$$\begin{aligned}\mathcal{K} &= E(\text{Cost of the option chosen at the first stage of the screen})\\ &= E\,(\text{Acceptance cost if item is accepted}\\ &\qquad + \text{Rejection cost if item is rejected}\\ &\qquad + \text{Cost of measuring } \underline{T} \text{ if measured})\,.\end{aligned}$$

That is, we choose $\{\Omega_A, \Omega_R, \Omega_M\}$ to minimise

$$\begin{aligned}\mathcal{K}(\Omega_A,\Omega_R,\Omega_M) &= E_{\underline{T},\underline{X},\underline{\xi}}\{c_a(\underline{T},\underline{X})I(\underline{X}\in\Omega_A)\\ &\qquad + c_r(\underline{T},\underline{X})I(\underline{X}\in\Omega_R) + c_m(\underline{T},\underline{X})I(\underline{X}\in\Omega_M)\}\end{aligned} \tag{3.1}$$

Definition 3.1 A *Bayes two-stage screen* $(\Omega_A^*,\Omega_R^*,\Omega_M^*)$ is a partition of $\Omega_{\underline{X}}$ satisfying

$$\mathcal{K}\,(\Omega_A^*,\Omega_R^*,\Omega_M^*) = \inf_{(\Omega_A,\Omega_R,\Omega_M)} \mathcal{K}\,(\Omega_A,\Omega_R,\Omega_M)\,, \tag{3.2}$$

the infimum in (3.2) being over all threefold partitions of $\Omega_{\underline{X}}$.

In order to describe Bayes two-stage screens we introduce $\tilde{c}_a$, $\tilde{c}_r$ and $\tilde{c}_m$, all functions from $\Omega_{\underline{X}}$ into $\mathbb{R}^+$ summarising the conditional expected costs incurred when an item with $\underline{X}=\underline{x}$ is accepted, rejected or passed onto the second stage respectively, viz.

$$\begin{aligned}\tilde{c}_a(\underline{x}) &= E_{\underline{\xi}}\left[E_{\underline{T}|\underline{X}=\underline{x},\underline{\xi}}\{c_a(\underline{T},\underline{x})\}\right]\\ \tilde{c}_r(\underline{x}) &= E_{\underline{\xi}}\left[E_{\underline{T}|\underline{X}=\underline{x},\underline{\xi}}\{c_r(\underline{T},\underline{x})\}\right]\\ \tilde{c}_m(\underline{x}) &= E_{\underline{\xi}}\left[E_{\underline{T}|\underline{X}=\underline{x},\underline{\xi}}\{c_m(\underline{T},\underline{x})\}\right].\end{aligned} \tag{3.3}$$

Theorem 3.1 Any partition of $\Omega_{\underline{X}}$ such that

$$\begin{aligned}
\Omega_A^* &\subseteq \left\{\underline{x} \in \Omega_{\underline{X}} : \tilde{c}_a(\underline{x}) = \min\left[\tilde{c}_a(\underline{x}), \tilde{c}_r(\underline{x}), \tilde{c}_m(\underline{x})\right]\right\}, \\
\Omega_R^* &\subseteq \left\{\underline{x} \in \Omega_{\underline{X}} : \tilde{c}_r(\underline{x}) = \min\left[\tilde{c}_a(\underline{x}), \tilde{c}_r(\underline{x}), \tilde{c}_m(\underline{x})\right]\right\}, \\
\Omega_M^* &\subseteq \left\{\underline{x} \in \Omega_{\underline{X}} : \tilde{c}_m(\underline{x}) = \min\left[\tilde{c}_a(\underline{x}), \tilde{c}_r(\underline{x}), \tilde{c}_m(\underline{x})\right]\right\},
\end{aligned}$$

is a Bayes two–stage screen.

Proof

From (3.1) and (3.3) we have that

$$\begin{aligned}
\mathcal{K}(\Omega_A, \Omega_R, \Omega_M) &= E_{\underline{X}}\Big(E_\xi\Big[E_{\underline{T}|\underline{X},\xi}\{c_a(\underline{T}, \underline{X})I(\underline{X} \in \Omega_A) \\
&\qquad + c_r(\underline{T}, \underline{X})I(\underline{X} \in \Omega_A) + c_m(\underline{T}, \underline{X})I(\underline{X} \in \Omega_A)\}\Big]\Big) \\
&= E_{\underline{X}}\{\tilde{c}_a(\underline{X})I(\underline{X} \in \Omega_A) + \tilde{c}_r(\underline{X})I(\underline{X} \in \Omega_R) + \tilde{c}_m(\underline{X})I(\underline{X} \in \Omega_M)\}
\end{aligned}$$

Given $\underline{X}$, the cost of each option (acceptance, rejection, measuring $\underline{T}$) must be at least as big as the cost of the cheapest option and so the Bayes cost

$$\begin{aligned}
\mathcal{K}(\Omega_A, \Omega_R, \Omega_M) \geq\ & E_{\underline{X}}\{\min\{\tilde{c}_a(\underline{X}), \tilde{c}_r(\underline{X}), \tilde{c}_m(\underline{X})\}\, I(\underline{X} \in \Omega_A) \\
&+ \min\{\tilde{c}_a(\underline{X}), \tilde{c}_r(\underline{X}), \tilde{c}_m(\underline{X})\}\, I(\underline{X} \in \Omega_R) \qquad (3.4) \\
&+ \min\{\tilde{c}_a(\underline{X}), \tilde{c}_r(\underline{X}), \tilde{c}_m(\underline{X})\}\, I(\underline{X} \in \Omega_M)\},
\end{aligned}$$

for any partition $(\Omega_A, \Omega_R, \Omega_M)$. Under such a partition, for any $\underline{X} \in \Omega_{\underline{X}}$,

$$I(\underline{X} \in \Omega_A) + I(\underline{X} \in \Omega_R) + I(\underline{X} \in \Omega_M) = 1,$$

as $\underline{X}$ must be contained in one and only one region of the partition. The bound in (3.4) simplifies to

$$\mathcal{K}(\Omega_A, \Omega_R, \Omega_M) \geq E_{\underline{X}}\left[\min\{\tilde{c}_a(\underline{X}), \tilde{c}_r(\underline{X}), \tilde{c}_m(\underline{X})\}\right]. \qquad (3.5)$$

Plainly any partition $(\Omega_A^*, \Omega_R^*, \Omega_M^*)$, as described in the statement of the theorem, achieves the lower bound in (3.4) and (3.5). □

Theorem 3.1 confirms that, once $\underline{X}$ has been measured, the optimal decision is to choose the option with the lowest expected cost. However, there are multiple

solutions for cases in which the minimisation concerned is attained by more than one of the arguments for any $\underline{x}$, that is, two or more options achieve the lowest expected cost. In order to remove this problem, we will adopt a tie–breaking convention which ranks the regions in the order, Ω_A, Ω_R and Ω_M. Thus, those values of $\underline{x}$ resulting in a three–way tie will be placed in the region Ω_A, those with a tie between Ω_R and Ω_M, in Ω_R, and so on. We will assume this convention throughout the remainder of this chapter, noting that, in any case, Bayes cost is not affected.

We note that if $\tilde{c}_m(\underline{x}) \geq \min\{\tilde{c}_a(\underline{x}), \tilde{c}_r(\underline{x})\}$ for all $\underline{x} \in \Omega_{\underline{X}}$ then, from Theorem 3.1, an optimal choice is $\Omega_M = \emptyset$ and we have the one–stage screen as considered by Turkman & Amaral Turkman (1989). Note also from Theorem 3.1 that knowledge of the marginal distribution of $\underline{X}$ is not necessary for a Bayes two–stage screen. The following result summarises Theorem 3.1 for the standard case.

Lemma 3.1 For the standard case, if $c_m/c_r < 1 - c_m/c_a$, then a Bayes two–stage screen is

$$\begin{aligned}
\Omega_A^* &= \left\{\underline{x} \in \Omega_{\underline{X}} : E_{\underline{\xi}}\left[P(\underline{T} \in C_{\underline{T}}|\underline{x}, \underline{\xi})\right] \geq 1 - \frac{c_m}{c_a}\right\}, \\
\Omega_R^* &= \left\{\underline{x} \in \Omega_{\underline{X}} : E_{\underline{\xi}}\left[P(\underline{T} \in C_{\underline{T}}|\underline{x}, \underline{\xi})\right] \leq \frac{c_m}{c_r}\right\}, \\
\Omega_M^* &= \left\{\underline{x} \in \Omega_{\underline{X}} : \frac{c_m}{c_r} < E_{\underline{\xi}}\left[P(\underline{T} \in C_{\underline{T}}|\underline{x}, \underline{\xi})\right] < 1 - \frac{c_m}{c_a}\right\},
\end{aligned}$$

otherwise the one–stage screen

$$\begin{aligned}
\Omega_A^* &= \left\{\underline{x} \in \Omega_{\underline{X}} : E_{\underline{\xi}}\left[P(\underline{T} \in C_{\underline{T}}|\underline{x}, \underline{\xi})\right] \geq \frac{c_a}{c_a + c_r}\right\}, \\
\Omega_R^* &= \left\{\underline{x} \in \Omega_{\underline{X}} : E_{\underline{\xi}}\left[P(\underline{T} \in C_{\underline{T}}|\underline{x}, \underline{\xi})\right] < \frac{c_a}{c_a + c_r}\right\}, \\
\Omega_M^* &= \emptyset.
\end{aligned}$$

is optimal.

Proof

In the standard case

$$\begin{aligned}
\tilde{c}_a(\underline{x}) &= E_{\underline{\xi}}\left[E_{\underline{T}|\underline{X}=\underline{x},\underline{\xi}}\left\{c_a I(\underline{T} \notin C_{\underline{T}})\right\}\right] \\
&= c_a\left[1 - E_{\underline{\xi}}\left\{P(\underline{T} \in C_{\underline{T}}|\underline{x}, \underline{\xi})\right\}\right].
\end{aligned}$$

Similarly

$$\tilde{c}_r(\underline{x}) = c_r E_{\underline{\xi}} \left\{ P(\underline{T} \in C_{\underline{T}} | \underline{x}, \underline{\xi}) \right\}$$

and

$$\tilde{c}_m(\underline{x}) = E_{\underline{\xi}} \left\{ E_{\underline{T}|\underline{X}=\underline{x},\underline{\xi}} (c_m) \right\} = c_m.$$

Theorem (3.1) gives the set Ω_M^* as the following requirements on $\underline{x}$

$$c_m < c_a \left[1 - E_{\underline{\xi}} \left\{ P(\underline{T} \in C_{\underline{T}} | \underline{x}, \underline{\xi}) \right\} \right],$$

and

$$c_m < c_r E_{\underline{\xi}} \left\{ P(\underline{T} \in C_{\underline{T}} | \underline{x}, \underline{\xi}) \right\}.$$

Rearranging, the region Ω_M^* is made up of those $\underline{x}$ for which

$$\frac{c_m}{c_r} < E_{\underline{\xi}} \left\{ P(\underline{T} \in C_{\underline{T}} | \underline{x}, \underline{\xi}) \right\} < 1 - \frac{c_m}{c_a}.$$

If $c_m/c_r \geq 1 - c_m/c_r$ this requirement will not be satisfied by any $\underline{x}$ and $\Omega_M^* = \emptyset$. Here a one–stage screen based on the covariates $\underline{X}$ is preferable to a two–stage screen. When this is not the case $c_m/c_r < 1 - c_m/c_a$ and so

$$c_m < \frac{c_a c_r}{c_a + c_r} \tag{3.6}$$

is a condition for Ω_M^* to be non-empty.

The set Ω_A^* must satisfy

$$c_a \left[1 - E_{\underline{\xi}} \left\{ P(\underline{T} \in C_{\underline{T}} | \underline{x}, \underline{\xi}) \right\} \right] \leq c_m,$$

and

$$c_a \left[1 - E_{\underline{\xi}} \left\{ P(\underline{T} \in C_{\underline{T}} | \underline{x}, \underline{\xi}) \right\} \right] \leq c_r E_{\underline{\xi}} \left\{ P(\underline{T} \in C_{\underline{T}} | \underline{x}, \underline{\xi}) \right\},$$

which, when rearranged, become

$$E_{\underline{\xi}} \left\{ P(\underline{T} \in C_{\underline{T}} | \underline{x}, \underline{\xi}) \right\} \geq 1 - \frac{c_m}{c_a}, \tag{3.7}$$

and

$$E_{\underline{\xi}}\left\{P(\underline{T} \in C_{\underline{T}}|\underline{x}, \underline{\xi})\right\} \geq \frac{c_a}{c_a + c_r}. \tag{3.8}$$

Under condition (3.7) and the requirement that Ω_M^* is non-empty, (3.6), it is clear that (3.8) is satisfied. Similarly, when Ω_M^* is empty and (3.6) is not met, the requirement (3.8) is a sufficient condition for Ω_A^* and condition (3.7) is obsolete.

By a similar argument, the set Ω_R^* must satisfy

$$E_{\underline{\xi}}\left\{P(\underline{T} \in C_{\underline{T}}|\underline{x}, \underline{\xi})\right\} \leq \frac{c_m}{c_r}$$

when Ω_M^* is non-empty and

$$E_{\underline{\xi}}\left\{P(\underline{T} \in C_{\underline{T}}|\underline{x}, \underline{\xi})\right\} < \frac{c_a}{c_a + c_r}.$$

when $\Omega_M^* = \emptyset$. All inequalities follow the tie-breaking convention. □

3.3 Probit regression model

Here we illustrate our designs within a probit regression model for $T|\underline{X}$. Here $\Omega_T = \{0, 1\}$ with $T = 1$ if and only if the item has the required attributes. In Chapter 2, such a model was used to obtain optimal screen designs when $\underline{X}$ is univariate and the screen is of a fixed form. Here we shall begin by taking $\underline{X}$ to be continuous and p–dimensional and give the Bayes two–stage screen for the standard case. Then, in section 3.3.1, we look at the case of $\underline{X}$ univariate and recover some of the results obtained in section 2.2. We also describe the conditions under which the form of the screen assumed in Chapter 2 is optimal. Section 3.3.2 considers the case when $\underline{X}$ is bivariate. In section 3.3.3 we give the Bayes two–stage screen for the numerical example introduced in section 2.4 and extend the example to the $p = 2$ case by including a second covariate related to the cause of Conn's syndrome.

The probit model has

$$P(T = 1|\underline{x}, \underline{\xi}) = \Phi\left(\underline{\xi}^T \underline{x}_0\right)$$

where $\Phi(a) = P(Z < a)$, $Z \sim N(0,1)$ and $\underline{x}_0^T = (1, \underline{x}^T)$. As most standard link functions give qualitatively similar results, the choice of probit link function is in part motivated by the closed form for $E_{\underline{\xi}}\{P(T = 1|\underline{x}, \underline{\xi})\}$ which results when $\pi(\underline{\xi})$ takes its asymptotic posterior form. Here the regression parameters follow a multivariate normal distribution, $\underline{\xi} \sim N_{p+1}(\underline{m}, \boldsymbol{S})$, as would approximately be the case if $\pi(\underline{\xi})$ were a posterior distribution based on a moderately sized data set. In this latter scenario, $\underline{m}$ would be the maximum likelihood estimate of ξ and $\boldsymbol{S}$, the inverse of Fisher's information matrix evaluated at $\underline{m}$.

Lemma 3.2

$$E_{\underline{\xi}}\left\{P(T = 1|\underline{x}, \underline{\xi})\right\} = \Phi\left\{\frac{\underline{m}^T\underline{x}_0}{(1 + \underline{x}_0^T\boldsymbol{S}\underline{x}_0)^{1/2}}\right\}.$$

Proof

Given $\underline{x}_0$, $\eta = \underline{\xi}^T\underline{x}_0 \sim N(\underline{m}^T\underline{x}_0, \underline{x}_0^T\boldsymbol{S}\underline{x}_0)$, and, by Lemma 2.1,

$$E_{\underline{\xi}}\left\{P(T = 1|\underline{x}, \underline{\xi})\right\} = E_{\eta}\left\{\Phi(\eta)\right\} = \Phi\left\{\underline{m}^T\underline{x}_0/\left(1 + \underline{x}_0^T\boldsymbol{S}\underline{x}_0\right)^{1/2}\right\}. \qquad \square$$

Hence, following Lemma 3.1, when $c_m/c_r < 1 - c_m/c_a$, the following partition yields a Bayes two–stage screen for the standard case,

$$\begin{aligned}
\Omega_A^* &= \left\{\underline{x} \in \Omega_{\underline{X}} : \frac{\underline{m}^T\underline{x}_0}{(1 + \underline{x}_0^T\boldsymbol{S}\underline{x}_0)^{1/2}} \geq c_2\right\} \\
\Omega_R^* &= \left\{\underline{x} \in \Omega_{\underline{X}} : \frac{\underline{m}^T\underline{x}_0}{(1 + \underline{x}_0^T\boldsymbol{S}\underline{x}_0)^{1/2}} \leq c_1\right\} \qquad (3.9) \\
\Omega_M^* &= \left\{\underline{x} \in \Omega_{\underline{X}} : c_1 < \frac{\underline{m}^T\underline{x}_0}{(1 + \underline{x}_0^T\boldsymbol{S}\underline{x}_0)^{1/2}} < c_2\right\},
\end{aligned}$$

where $c_2 = \Phi^{-1}(1 - c_m/c_a)$ and $c_1 = \Phi^{-1}(c_m/c_r)$, with $c_1 < c_2$. Notice that c_1 and c_2 are as defined in equations (2.7) in section 2.2.

3.3.1 Screening with one covariate

We now examine the case $p = 1$ more closely. Here X is univariate and $\underline{\xi}$ is bivariate with mean and covariance written as,

$$\underline{m} = \begin{pmatrix} m_0 \\ m_1 \end{pmatrix} \quad \text{and} \quad \boldsymbol{S} = \begin{pmatrix} s_0^2 & rs_0s_1 \\ rs_0s_1 & s_1^2 \end{pmatrix}.$$

From (3.9) a Bayes two–stage screen for the standard case, with $c_m/c_r < 1 - c_m/c_a$, is given by

$$\begin{aligned} \Omega_A^* &= \left\{ x \in \Omega_X : \frac{m_0 + m_1 x}{(1 + s_0^2 + 2xrs_0s_1 + x^2s_1^2)^{1/2}} \geq c_2 \right\}, \\ \Omega_R^* &= \left\{ x \in \Omega_X : \frac{m_0 + m_1 x}{(1 + s_0^2 + 2xrs_0s_1 + x^2s_1^2)^{1/2}} \leq c_1 \right\}, \\ \Omega_M^* &= \left\{ x \in \Omega_X : c_1 < \frac{m_0 + m_1 x}{(1 + s_0^2 + 2xrs_0s_1 + x^2s_1^2)^{1/2}} < c_2 \right\}. \end{aligned} \tag{3.10}$$

If we employ a suitable data transformation, as in section 2.3.8, we may assume that $m_0 = 0$ and $s_1 = 1$. For this case we now investigate the form of the Bayes two–stage screen and compare it with the form of screen assumed in Chapter 2. First, for $i = 1, 2$, we define

$$\begin{aligned} u_i^{\pm} &= c_i^2[rs_0 \pm \{r^2s_0^2 + (m_1^2 - c_i^2)(1 + s_0^2)c_i^{-2}\}^{1/2}]/(m_1^2 - c_i^2), \\ u' &= -(1 + s_0^2)/2rs_0. \end{aligned}$$

When both u_i^+ and u_i^- are real and distinct, which will always be the case for $m_1^2 > c_i^2$, notice that,

$$\begin{aligned} u_i^+ > u_i^- \quad &\text{for} \quad m_1^2 > c_i^2, \\ u_i^- > u_i^+ \quad &\text{for} \quad m_1^2 < c_i^2. \end{aligned}$$

Lemma 3.3 When the quadratic equation

$$x^2 + 2bx + c = 0$$

(b, c constants) does not have real roots then,

(a) no real x satisfy

$$x^2 + 2bx + c \leq 0, \tag{3.11}$$

(b) all real x satisfy

$$x^2 + 2bx + c > 0. \tag{3.12}$$

Proof

The roots of the quadratic are given by

$$x = -b \pm \sqrt{b^2 - c},$$

where the roots are not real when $c > b^2$.

(a) For inequality (3.11) to hold for some real x, there must exist a $d \geq 0$ such that

$$x^2 + 2bx + c = -d$$

has real roots. The roots are real for $b^2 \geq c + d$. This will never be the case as $c > b^2$ and $d \geq 0$.

(b) Each real value of x must satisfy at least one of (3.11) and (3.12). In part (a) we have shown that no real x satisfy (3.11) and so all real x must satisfy (3.12).

□

Theorem 3.2 Under requirements on costs and regression parameters, the following table gives Ω_A^*, the acceptance region of a Bayes two-stage screen, when $T|X$ is

modelled by a probit regression and X is univariate.

		$m_1 > 0$		$m_1 < 0$	
		$r > 0$	$r < 0$	$r > 0$	$r < 0$
$c_2 > 0$	$m_1^2 > c_2^2$	$x \geq u_2^+$		$x \leq u_2^-$	
	$m_1^2 = c_2^2$	NA	$x \geq u'$	$x \leq u'$	NA
	$m_1^2 < c_2^2$†	NA	$u_2^+ \leq x \leq u_2^-$	$u_2^+ \leq x \leq u_2^-$	NA
	$m_1^2 < c_2^2$‡	NA	NA	NA	NA
$c_2 < 0$	$m_1^2 > c_2^2$	$x \geq u_2^-$		$x \leq u_2^+$	
	$m_1^2 = c_2^2$	$x \geq u'$	AA	AA	$x \leq u'$
	$m_1^2 < c_2^2$†	$x \leq u_2^+$ or $x \geq u_2^-$	AA	AA	$x \leq u_2^+$ or $x \geq u_2^-$
	$m_1^2 < c_2^2$‡	AA	AA	AA	AA
$c_2 = 0$		$x \geq 0$		$x \leq 0$	

where, NA denotes that no items should be accepted by the first stage of the screen, AA denotes that all items should be accepted by the first stage of the screen, † indicates that condition

$$r^2 s_0^2 + (m_1^2 - c_2^2)(1 + s_0^2) c_2^{-2} > 0. \tag{3.13}$$

holds, and ‡ indicates that (3.13) does not hold.

Proof

Under the data transformation, Ω_A^* denotes those values of x for which

$$\frac{m_1 x}{(1 + s_0^2 + 2xrs_0 + x^2)^{1/2}} \geq c_2. \tag{3.14}$$

On squaring both sides of this inequality, the form of the region is determined by the sign of c_2.

For $c_2 > 0$, inequality (3.14) is satisfied when both

$$m_1 x > 0 \tag{3.15}$$

and

$$m_1^2 x^2 / \left(1 + s_0^2 + 2xrs_0 + x^2\right) \geq c_2^2 \tag{3.16}$$

are true. When $m_1^2 > c_2^2$, rearranging (3.16) gives the requirement

$$x^2 - \frac{2rs_0c_2^2}{m_1^2 - c_2^2}x - \frac{(1+s_0^2)c_2^2}{m_1^2 - c_2^2} \geq 0. \tag{3.17}$$

Solving for equality in (3.17) gives $x = u_2^{\pm}$ and (3.17) can be rewritten as

$$\left(x - u_2^-\right)\left(x - u_2^+\right) \geq 0.$$

Hence, x must be either greater than both u_2^- and u_2^+, less than both, or equal to at least one. In this case, u_2^+ is positive and u_2^- is negative and so condition (3.16) is satisfied when either

$$x \geq u_2^+ \qquad \text{or} \qquad x \leq u_2^-.$$

When $m_1 > 0$, rule (3.15) requires that $x > 0$ and so only those $x \geq u_2^+$ meet inequality (3.14). Also, when $m_1 < 0$, (3.15) requires $x < 0$ and $x \leq u_2^-$ is the appropriate rule.

When $m_1^2 = c_2^2$, inequality (3.16) rearranges to give the simple requisite

$$2xrs_0 + s_0^2 + 1 \leq 0.$$

This is met when

$$x \leq u' \quad \text{if } r > 0; \tag{3.18}$$

$$x \geq u' \quad \text{if } r < 0. \tag{3.19}$$

In the first instance, with $r > 0$, u' is negative and in the second, with $r < 0$, u' is positive. However, when $m_1 > 0$, inequality (3.15) demands that $x > 0$, which will not be satisfied for any x that comply with the requirement for $r > 0$. Here $\Omega_A^* = \emptyset$. Similarly, for $m_1 < 0$, (3.15) requires that $x < 0$ which will not be true for any x that fulfil the condition for $r < 0$. In the other cases, $m_1 > 0$ and $r < 0$, and $m_1 < 0$ and $r > 0$, the x that satisfy the appropriate rule from (3.18) and (3.19) will also comply with (3.15).

In the case of $m_1^2 < c_2^2$, the rule given by (3.16) becomes

$$x^2 - \frac{2rs_0c_2^2}{m_1^2 - c_2^2}x - \frac{(1+s_0^2)c_2^2}{m_1^2 - c_2^2} \leq 0. \tag{3.20}$$

and so

$$\left(x - u_2^-\right)\left(x - u_2^+\right) \leq 0,$$

where u_2^- and u_2^+ are real when condition (3.13) holds. When the roots at equality in (3.20) are not real then it follows from Lemma 3.3 that no real x will satisfy (3.20) and no items should be accepted by the first stage of the screen. When the roots are real then condition (3.20) is met when either

$$u_2^- \leq x \leq u_2^+ \qquad \text{or} \qquad u_2^+ \leq x \leq u_2^-. \tag{3.21}$$

Here u_2^- is greater than u_2^+ and so the first rule above is obsolete. When $r > 0$ both of the roots are negative and for $r < 0$ both are positive. Hence, when $m_1 > 0$ and $r > 0$, there will be no x which agree with both the second rule in (3.21) and condition (3.15). Here $\Omega_A^* = \emptyset$. Also, when $m_1 < 0$ and $r < 0$, no x will satisfy both conditions. Otherwise, when $m_1 > 0$ and $r < 0$, or $m_1 < 0$ and $r > 0$, all x that satisfy the second rule in (3.21) will also satisfy (3.15).

For $c_2 < 0$, inequality (3.14) is satisfied when

$$m_1 x \geq 0 \tag{3.22}$$

or both

$$m_1 x < 0 \tag{3.23}$$

and

$$m_1^2 x^2 / \left(1 + s_0^2 + 2xrs_0 + x^2\right) \leq c_2^2 \tag{3.24}$$

are true. When $m_1^2 > c_2^2$, the condition (3.24) becomes inequality (3.20). However, here condition (3.13) always holds and u_2^+ (which is positive) is greater than u_2^- (negative). Hence (3.24) gives the requirement that

$$u_2^- \leq x \leq u_2^+. \tag{3.25}$$

When $m_1 > 0$, condition (3.22) accepts all non–negative x and together, (3.23) and (3.25) accept negative x such that $x \geq u_2^-$. Hence, for (3.14) to hold, it is enough to require $x \geq u_2^-$. Similarly, for $m_1 < 0$, the appropriate rule is $x \leq u_2^+$.

When $m_1^2 = c_2^2$, (3.24) rearranges to give

$$2xrs_0 + s_0^2 + 1 \geq 0.$$

This holds when

$$x \geq u' \qquad \text{if } r > 0; \tag{3.26}$$

$$x \leq u' \qquad \text{if } r < 0. \tag{3.27}$$

Here u' is negative in the first case, when $r > 0$, and positive in the second, when $r < 0$. When $m_1 > 0$ and $r > 0$, condition (3.22) accepts all non-negative x and (3.23) allows negative x. Hence all $x \geq u'$ will satisfy (3.14). For $m_1 > 0$ and $r < 0$, the rules (3.27) and (3.23) accept all negative x and (3.22) accepts all non-negative x. Here all items should be accepted by the first stage of the screen. Similarly, for $m_1 < 0$ and $r < 0$, the rule $x \leq u'$ must be satisfied for (3.14) to be met and, for $m_1 < 0$ and $r > 0$, all items should be accepted.

In the case of $m_1^2 < c_2^2$, the rule given by (3.24) becomes inequality (3.17), where the roots at equality, u_2^- and u_2^+, are real only when condition (3.13) is satisfied. For real u_2^- and u_2^+, the root u_2^- is greater than u_2^+ and so (3.17) requires that either

$$x \leq u_2^+ \qquad \text{or} \qquad x \geq u_2^-. \tag{3.28}$$

When $r < 0$, both roots are positive and for $r > 0$, both are negative. In cases with $m_1 > 0$ and $r < 0$, requirement (3.22) accepts all non-negative x, and rules (3.23) and (3.28) accept all negative x. Here $\Omega_A^* = \Omega_X$. For $m_1 > 0$ and $r > 0$, (3.22) accepts all non-negative x and (3.23) and (3.28) accept all negative x for which (3.28) is true. As both u_2^+ and u_2^- are negative, (3.28) is a sufficient condition for acceptance. Similarly, for $m_1 < 0$ and $r > 0$, all items should be accepted, and when $m_1 < 0$ and $r < 0$, condition (3.28) is sufficient for acceptance. In cases when (3.13) is not satisfied, it follows from Lemma 3.3 that all real x satisfy (3.17) and so (3.22), (3.23) and (3.24) admit all values of x. Hence $\Omega_A^* = \Omega_X$.

For $c_2 = 0$, the acceptance rule (3.14) simply becomes the rule $m_1 x > 0$. For $m_1 > 0$, it follows that all items with $x > 0$ should be accepted and for $m_1 < 0$ all items with $x < 0$ should be accepted. □

Corollary 3.1 The table below gives Ω_R^*, the rejection region of the Bayes two–stage screen.

		$m_1 > 0$		$m_1 < 0$	
		$r > 0$	$r < 0$	$r > 0$	$r < 0$
$m_1^2 > c_1^2$		$x \leq u_1^+$		$x \geq u_1^-$	
$c_1 > 0$	$m_1^2 = c_1^2$	AR	$x \leq u'$	$x \geq u'$	AR
	$m_1^2 < c_1^2$†	AR	$x \leq u_1^+$ or $x \geq u_1^-$	$x \leq u_1^+$ or $x \geq u_1^-$	AR
	$m_1^2 < c_1^2$‡	AR	AR	AR	AR
	$m_1^2 > c_1^2$	$x \leq u_1^-$		$x \geq u_1^+$	
$c_1 < 0$	$m_1^2 = c_1^2$	$x \leq u'$	NR	NR	$x \geq u'$
	$m_1^2 < c_1^2$†	$u_1^+ \leq x \leq u_1^-$	NR	NR	$u_1^+ \leq x \leq u_1^-$
	$m_1^2 < c_1^2$‡	NR	NR	NR	NR
$c_1 = 0$		$x \leq 0$		$x \geq 0$	

where, NR denotes that no items should be rejected by the first stage of the screen, AR denotes that all items should be rejected by the first stage of the screen, † indicates that condition

$$r^2 s_0^2 + (m_1^2 - c_1^2)(1 + s_0^2)c_1^{-2} > 0 \tag{3.29}$$

holds, and ‡ indicates that (3.29) does not hold.

Proof

After the data transformation, Ω_R^* is made up of those x for which

$$\frac{m_1 x}{(1 + s_0^2 + 2xrs_0 + x^2)^{1/2}} \leq c_1. \tag{3.30}$$

Multiplying (3.30) through by -1, the result follows easily from Theorem 3.2, by substituting $-m_1$ for m_1, $-c_1$ for c_2 and 'reject' for 'accept'. □

Comments

1. The conditions on X under which the performance variable T should be measured can be recovered from the relation $\Omega_M^* = \Omega_X \setminus (\Omega_A^* \cup \Omega_R^*)$.

2. When $m_1 > \max(|c_1|, |c_2|)$ the Bayes two–stage screen has the simple form: accept an item if $x \geq w$, say, reject an item if $x \leq v$, say, and pass an item to the second stage if $v < x < w$. This is in accord with Chapter 2 which assumed *a priori* that the two–stage screen was of this simple form and optimised within this restricted class. We have shown, *inter alia*, that when $m_1 > \max(|c_1|, |c_2|)$, the solution given in Chapter 2 is in fact globally optimal.

   Similarly, when $m_1 < -\max(|c_1|, |c_2|)$, we have the simple Bayes two–stage screen: accept an item if $x \leq w$, say, reject an item if $x \geq v$, say, and pass an item to the second stage if $w < x < v$.

3. Notice that, when $|m_1| \leq \min(|c_1|, |c_2|)$, it may be optimal to omit the X–stage of the screen and either accept, reject or measure T for all items.

4. For the case in which the regression parameters are standardised so that $m_0 = 0$ and $s_1 = 1$, Theorem 3.2 and Corollary 3.1 describe the form of the screen conditional on the values of m_1 and r. In terms of unstandardised quantities, these requirements should be made on the values of

$$\frac{m_1}{s_1} \qquad \text{and} \qquad \frac{m_1 r s_0 - m_0 s_1}{(m_1^2 s_0^2 + m_0^2 s_1^2 - 2 m_0 m_1 r s_0 s_1)^{1/2}},$$

   respectively. See section 2.3.8.

3.3.2 Screening with two covariates

The results presented for the $p = 1$ case have analogues for cases with $p > 1$. When $p = 2$ and the standardised mean regression parameters $|m_1/s_1|$ and $|m_2/s_2|$ are large enough, the boundaries of Ω_A^*, Ω_R^* and Ω_M^* in (3.9) are hyperbolae. By way of example, consider a case with mean regression parameter $\underline{m} = (0, 1, 2)^T$, covariance structure $var(\xi_i) = 1$, $i = 0, 1, 2$ and $corr(\xi_i, \xi_j) = 0.2$, $i, j = 0, 1, 2$ $(i \neq j)$, and costs $(c_m, c_a, c_r) = (1, 4, 6)$. Here $c_2 = \Phi^{-1}(0.75) = 0.6745$ and so Ω_A^* in (3.9) contains those $\underline{x} \in \Omega_{\underline{X}}$ for which

$$\frac{x_1 + 2x_2}{\{2 + x_1^2 + x_2^2 + 0.4(x_1 + x_2 + x_1 x_2)\}^{1/2}} \geq 0.6745.$$

By squaring both sides and then following similar reasoning to that used in the proof of Theorem 3.2, the Bayes two–stage screen accepts those items for which

$$x_2 > \frac{-x_1}{2} \tag{3.31}$$

and

$$\left(x_2 - w^+\right)\left(x_2 - w^-\right) \geq 0, \tag{3.32}$$

where

$$w^{\pm} = 0.02567 - 0.5385x_1 \pm \left(0.1363x_1^2 + 0.02369x_1 + 0.2573\right)^{1/2}.$$

Plainly, w^+ will always be larger than w^- and so rule (3.32) will be satisfied when either $x_2 \geq w^+$ or $x_2 \leq w^-$. However, for both $x_2 \leq w^-$ and requirement (3.31) to hold, we must have

$$\frac{-x_1}{2} < w^-.$$

Writing in w^- explicitly and rearranging, this will be true for any x_1 such that

$$\frac{0.02567 - 0.0385x_1}{\left(0.1362x_1^2 + 0.02369x_1 + 0.2573\right)^{1/2}} > 1. \tag{3.33}$$

Analysis of the turning points of the left hand side shows there to be only a maximum of 0.1188 at $x_1 = -3$. Hence the inequality will not hold for any x_1. Therefore, no $x_2 \leq w^-$ will satisfy inequality (3.31). Also, if $w^+ > -x_1/2$ for all x_1 then all values of (x_1, x_2) that satisfy the rule $x_2 \geq w^+$ will also satisfy (3.31) . This will be the case if the inequality

$$w^+ < -x_1/2 \tag{3.34}$$

does not hold for any x_1. Rearranging (3.34) and substituting in for w^+, we again obtain inequality (3.33) which never holds. Hence, $w^+ > -x_1/2$ for all x_1 and requirement (3.31) is always true for $x_2 \geq w^+$. The acceptance region of the Bayes two–stage screen, Ω_A^*, for this example is now given by those (x_1, x_2) for which

$$x_2 \geq 0.02567 - 0.5385x_1 + \left(0.1363x_1^2 + 0.02369x_1 + 0.2573\right)^{1/2}.$$

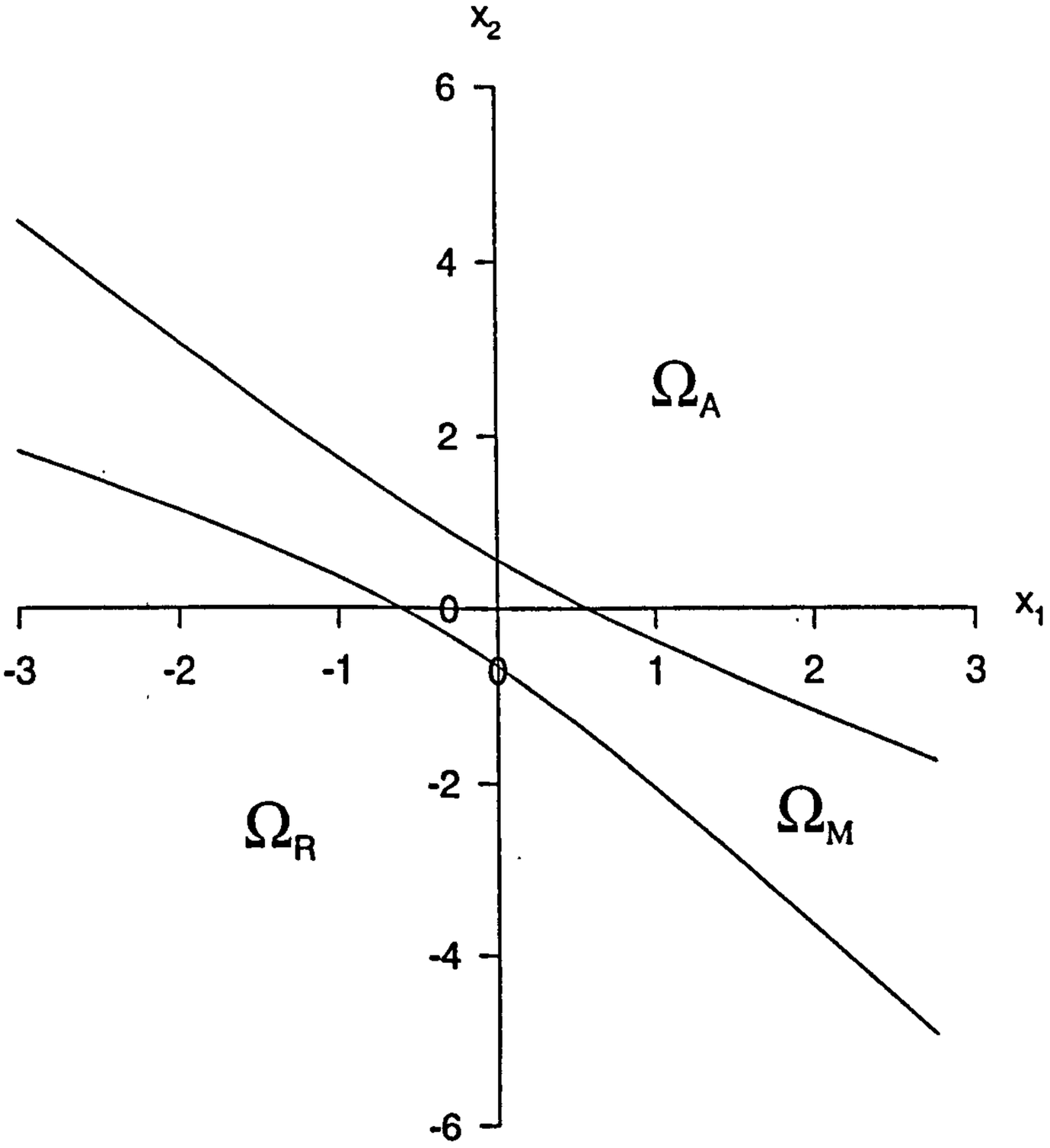


Figure 3.1: Plot of screening partition

Also $c_1 = \Phi^{-1}(1/6) = 0.9674$ and by a similar analysis, Ω_R^* is made up of those (x_1, x_2) for which

$$x_2 \leq 0.06108 - 0.5916x_1 - \left(0.3291x_1^2 + 0.04989x_1 + 0.6146\right)^{1/2}.$$

Figure 3.1 depicts the screen graphically. Note that the regions are not symmetric in x_1 and x_2 because $m_1 \neq m_2$ and $c_a \neq c_r$.

3.3.3 Conn's syndrome example

Now we return to the numerical example presented in section 2.4. There we dealt with the construction of a two–stage screen to determine the cause of Conn's syndrome. The cause of the illness determines the correct treatment. The first stage of the screen was based on a single covariate, the concentration of potassium in the blood plasma. At the second stage, patients were submitted for an exploratory operation which will verify the cause. We gave the Bayes design of the screen within a restricted class of forms for the screen and under probit modelling assumptions. Also, we assumed that the cost structure was as in the standard case with $c_a = 3c_r/4$ and $c_m = 3c_r/40$ and that the regression parameters were standardised so that $m_1 > 0$, $m_0 = 0$ and $s_1 = 1$. In section 3.3.1, we have shown that when $m_1 > \max(|c_1|, |c_2|)$ then the form of the screen assumed in Chapter 2 is Bayes optimal. In this example, we have that $c_1 = -1.440$ and $c_2 = 1.282$, under the standardisation $m_1 = 2.909$ and so the two–stage screen described by (2.36) is Bayes optimal. In section 2.4 we also recovered the Bayes design for unstandardised values of the regression parameters by reversing the data transformation and supposing that, when $m_1 < 0$, the screen is of the form: accept an item if $x \leq w$, say, reject an item if $x \geq v$, say, and pass an item to the second stage if $w < x < v$. In section 3.3.1 we show this to be the correct form of the Bayes two–stage screen for (unstandardised) $m_1/s_1 < -\max(|c_1|, |c_2|)$. Clearly this inequality holds for this case and the two–stage screen given by Figure 2.2 is Bayes optimal.

The Bayes two–stage screen for the $p = 1$ case was based on 31 cases of Conn's

syndrome in which both the cause of the illness had been established and a measurement of the covariate had been taken. For each of these 31 cases, a measurement on a second covariate was also taken, the concentration of carbon dioxide in the blood plasma. We now consider a Bayes two–stage screen using both available covariates ($p = 2$). To compute the Bayes cost of a screen we will need to make an assumption about the distribution of the screening variables. Taking log concentrations makes a bivariate normality assumption reasonable. We recommend standardising the screening variables to ease comparison between the parameters of the regression model. These factors lead us to design a Bayes two-stage screen using

$$\begin{aligned} X_1 &= \text{standardised log concentration of potassium,} \\ X_2 &= \text{standardised log concentration of carbon dioxide.} \end{aligned} \tag{3.35}$$

For the log concentration of carbon dioxide, we standardise using the sample estimates of the mean and variance which are 3.334 and 0.1115^2, respectively. The procedure used to standardise the log concentration of potassium is given in section 2.4. The transformed data is given in Table 3.1.

Estimates of the regression parameters can be found by a simple extension of the iterative maximum likelihood method described in section 2.3.7. Using S-PLUS to implement such a scheme, from the data we have

$$\underline{m} = \begin{pmatrix} 1.241 \\ -1.576 \\ 0.953 \end{pmatrix} \quad \text{and} \quad \boldsymbol{S} = \begin{pmatrix} 0.3030 & -0.2202 & 0.1468 \\ -0.2202 & 0.4874 & 0.0254 \\ 0.1468 & 0.0254 & 0.3129 \end{pmatrix}. \tag{3.36}$$

Assuming the same costs as in the $p = 1$ example, and that a linear probit regression model is appropriate for $T|X_1, X_2$, the Bayes two-stage screen takes the form of (3.9), that is,

$$\begin{aligned} \Omega_A^* &= \{\underline{x} : g(x_1, x_2) \geq 1.282\}, \\ \Omega_R^* &= \{\underline{x} : g(x_1, x_2) \leq -1.440\}, \\ \Omega_M^* &= \{\underline{x} : -1.440 < g(x_1, x_2) < 1.282\}. \end{aligned} \tag{3.37}$$

Table 3.1: Cases of Conn's syndrome

Cause of syndrome	Standardised log concentration of K X_1				Standardised log concentration of CO_2 X_2			
$T=1$	−1.275	0.136	−0.019	−0.346	0.691	−0.305	−0.341	1.462
	0.842	0.136	−0.881	−0.881	−1.363	−0.018	0.484	0.601
	−1.074	−0.180	−1.275	−1.485	1.238	0.448	−0.682	1.650
	−0.517	0.136	−0.180	0.136	1.462	0.332	−0.206	1.013
	−2.178	0.972	−1.485	−0.517	1.596	−0.206	1.462	−0.179
$T=0$	1.682	0.286	0.842	−0.019	−1.623	−1.031	−0.753	−2.179
	1.571	0.572	0.842	1.098	−0.081	−0.018	−1.031	−0.682
	0.431	0.842	1.791		−0.341	−0.682	−0.816	

where

$$g(x_1,x_2)=\frac{1.241-1.576x_1+0.953x_2}{(1.3030-0.4404x_1+0.2936x_2+0.0507x_1x_2+0.4874x_1^2+0.3129x_2^2)^{1/2}}. \tag{3.38}$$

Notice that $|m_1|/s_1 > |m_2|/s_2 > \max(|c_1|,|c_2|)$ and the boundaries of the regions are hyperbolae as shown in the graph of the partition in Figure 3.2.

Computing Bayes cost

With costs as in the standard case the Bayes cost of a Bayes two–stage screen is

$$\begin{aligned}\mathcal{K}(\Omega_A^*,\Omega_R^*,\Omega_M^*) \;=\; & E_{\underline{X}}\Big\{c_a E_{\underline{\xi}}\left[P(T=0|\underline{X},\underline{\xi})\right] I\left(\underline{X}\in\Omega_A^*\right) \\ & + c_r E_{\underline{\xi}}\left[P(T=1|\underline{X},\underline{\xi})\right] I\left(\underline{X}\in\Omega_R^*\right) + c_m I\left(\underline{X}\in\Omega_M^*\right)\Big\}.\end{aligned}$$

With a probit regression model for $T|\underline{X}$ that is based on a moderately sized sample, Lemma 3.2 describes the closed form for $E_{\underline{\xi}}\left[P(T=1|\underline{x},\underline{\xi})\right]$. Hence, for the Conn's

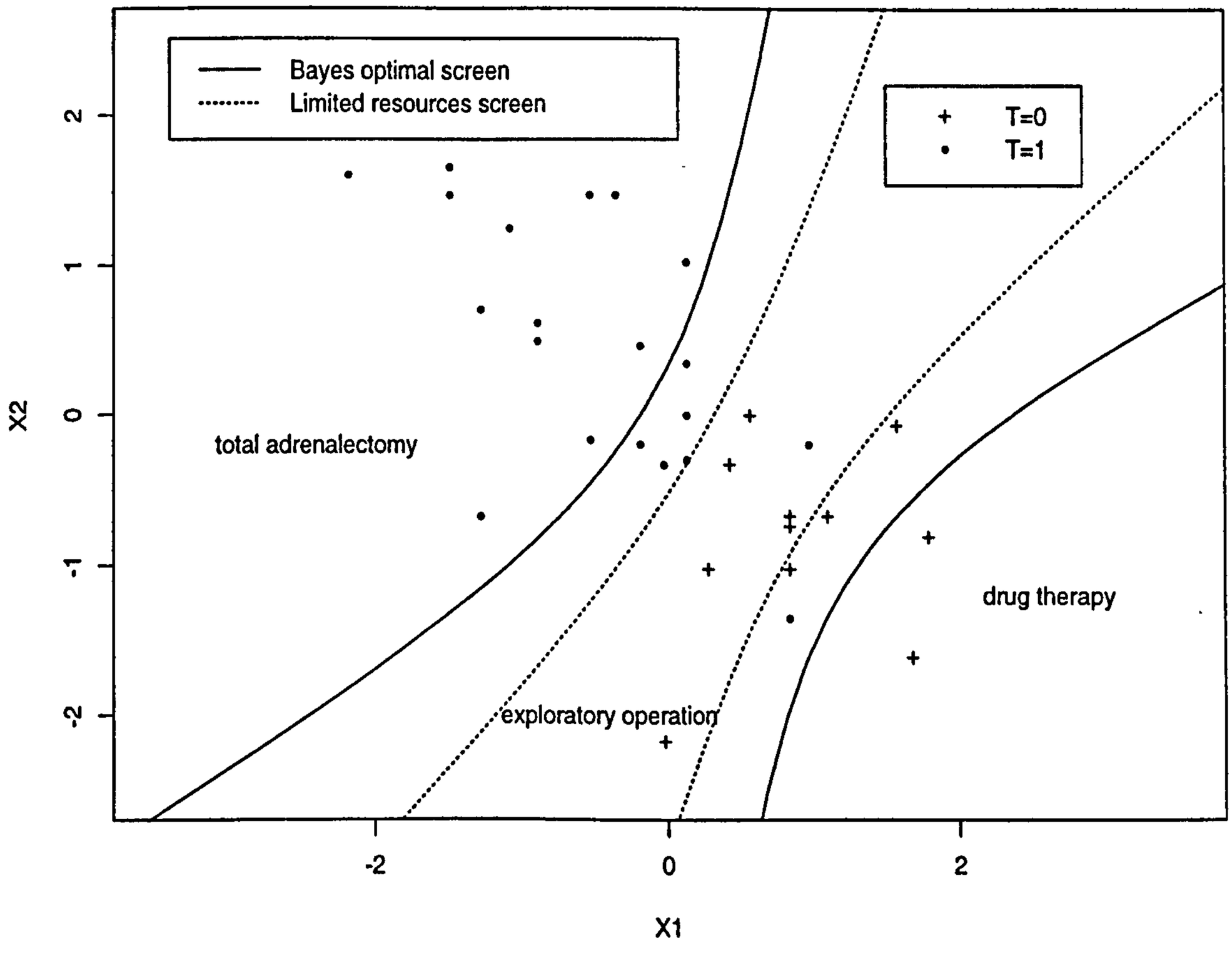


Figure 3.2: Plot of 2–dimensional screening regions

syndrome data, the Bayes cost of the screen is

$$\begin{aligned} \mathcal{K}(\Omega_A^*, \Omega_R^*, \Omega_M^*) &= E_{\underline{X}}[c_a \Phi\{-g(X_1, X_2)\} I(\underline{X} \in \Omega_A^*) \\ &\quad + c_r \Phi\{g(X_1, X_2)\} I(\underline{X} \in \Omega_R^*) + c_m I(\underline{X} \in \Omega_M^*)]. \end{aligned}$$

To compute the Bayes cost of the screen we must make an assumption about the distribution of the screening variables (X_1, X_2). As in the $p = 1$ case, it seems reasonable to assume that the log concentrations are normal. Hence, we suppose that the screening variables (3.35) are bivariate standard normal with correlation coefficient given by the sample value, -0.6783. Under this assumption the Bayes

cost becomes

$$\begin{aligned}\mathcal{K}(\Omega_A^*, \Omega_R^*, \Omega_M^*) &= \int_{\mathbb{R}^2} [c_a \Phi \{-g(x_1, x_2)\} I(\underline{x} \in \Omega_A^*) \\ &\quad + c_r \Phi \{g(x_1, x_2)\} I(\underline{x} \in \Omega_R^*) \\ &\quad + c_m I(\underline{x} \in \Omega_M^*)] \phi_2(x_1, x_2 | r_x = -0.6783)\, d\underline{x},\end{aligned}$$

where $\phi_2(x_1, x_2|r_x)$ is the bivariate normal density function with correlation r_x. A simple way of estimating this integral is to use Monte–Carlo techniques (see, for example, Hammersley and Hanscomb (1964)). We denote

$$\begin{aligned}h(x_1, x_2) &= c_a \Phi \{-g(x_1, x_2)\} I(\underline{x} \in \Omega_A^*) \\ &\quad + c_r \Phi \{g(x_1, x_2)\} I(\underline{x} \in \Omega_R^*) + c_m I(\underline{x} \in \Omega_M^*), \qquad (3.39)\end{aligned}$$

and suppose that we have a random sample, $(X_1^{(1)}, X_2^{(1)}), (X_1^{(2)}, X_2^{(2)}), \ldots, (X_1^{(n)}, X_2^{(n)})$, from $\phi_2(x_1, x_2|r_x = -0.6783)$. Then

$$\widehat{\mathcal{K}} = \frac{1}{n} \sum_{i=1}^{n} h(X_1^{(i)}, X_2^{(i)})$$

is an unbiased estimator of $\mathcal{K}$. The variance of $\widehat{\mathcal{K}}$ is σ^2/n, where

$$\begin{aligned}\sigma^2 &= Var_{\underline{X}} \{h(X_1, X_2)\} \\ &= \int_{\mathbb{R}^2} h(X_1, X_2)^2 \phi_2(x_1, x_2|r_x = -0.6783)\, d\underline{x} - \mathcal{K}^2.\end{aligned}$$

To estimate $\mathcal{K}$ we use a computer to generate a pseudo-random sample $(x_1^{(1)}, x_2^{(1)})$, $(x_1^{(2)}, x_2^{(2)}), \ldots, (x_1^{(n)}, x_2^{(n)})$ from $\phi_2(x_1, x_2|r_x = -0.6783)$ and calculate $\mathcal{K}$ for the sample. The variance σ^2 can be estimated by the (adjusted) sample variance of $h(X_1, X_2)$,

$$\widehat{\sigma^2} = \frac{1}{n-1} \left\{ \sum_{i=1}^{n} h(X_1^{(i)}, X_2^{(i)})^2 - n\widehat{\mathcal{K}}^2 \right\}.$$

By the Central Limit Theorem, for n large enough, the estimator $\widehat{\mathcal{K}}$ will be approximately normal with mean $\mathcal{K}$ and variance σ^2/n. Hence an approximate 95% confidence interval for $\mathcal{K}$ is given by

$$\widehat{\mathcal{K}} \pm 1.96\sqrt{\widehat{\sigma^2}/n}.$$

Notice that we make a further approximation by substituting the sample estimate of the variance for the variance. As n is typically very large this will make little difference. To specify the accuracy of the answer we can give a threshold value, δ_{mc} say, for the relative error and continue sampling until the threshold is breached. That is, we can continue generating samples until

$$\frac{1.96\sqrt{\widehat{\sigma^2}/n}}{\hat{\mathcal{K}}} < \delta_{mc}. \tag{3.40}$$

For a value of $\delta_{mc} = 10^{-a}$, we will be reasonably sure (95%) that the result of the procedure will be accurate to a significant figures. Notice that the left hand side above decreases proportional to $\sqrt{n}$. Hence, to improve the accuracy of the answer by one significant figure the number of simulations must be multiplied by 100. The algorithm used to compute Bayes cost is given below.

Algorithm

1. Initialise at zero two storage bins that will hold Σ, denoting the sum of the sample values of $h(x_1, x_2)$ and Σ^2, denoting the sum of the squared values of $h(x_1, x_2)$.

2. Generate a pseudo random observation, $(x_1^{(i)}, x_2^{(i)})$, from $\phi_2(x_1, x_2 | r_x = -0.6783)$ using a standard routine for generating random numbers from a bivariate normal random variable, for example N.A.G. routine G05EZF, see N.A.G. (1990).

3. Calculate $h(x_1^{(i)}, x_2^{(i)})$ as follows.

   Evaluate $g(x_1^{(i)}, x_2^{(i)})$ and

   (i) if $g(x_1^{(i)}, x_2^{(i)}) \geq 1.282$ then $\underline{x}^{(i)} \in \Omega_A^*$ and $h(x_1^{(i)}, x_2^{(i)}) = c_a \Phi\left\{-g(x_1^{(i)}, x_2^{(i)})\right\}$,

   (ii) else, if $g(x_1^{(i)}, x_2^{(i)}) \leq -1.440$ then $\underline{x}^{(i)} \in \Omega_A^*$ and $h(x_1^{(i)}, x_2^{(i)}) = c_r \Phi\left\{g(x_1^{(i)}, x_2^{(i)})\right\}$,

   (iii) otherwise, $\underline{x}^{(i)} \in \Omega_M^*$ and $h(x_1^{(i)}, x_2^{(i)}) = c_m$.

4. Let $\Sigma = \Sigma + h(x_1^{(i)}, x_2^{(i)})$ and $\Sigma^2 = \Sigma^2 + h(x_1^{(i)}, x_2^{(i)})^2$.

Table 3.2: Screen classifications

		1–d screen classification			2–d screen classification		
Actual	Total	$T=1$	$T=0$	measure T	$T=1$	$T=0$	measure T
$T=1$	20	11	0	9	13	0	7
$T=0$	11	0	3	8	0	2	9

5. Either return to step 2 or, for $i \geq 1000$ say, calculate $\widehat{\mathcal{K}} = \Sigma/n$ and $\widehat{\sigma^2} = \{\Sigma^2 - n\widehat{\mathcal{K}}^2\}/(n-1)$ and test for the required accuracy with requirement (3.40). If the current accuracy of the answer is sufficient then stop, otherwise return to step 2.

In order to calculate the Bayes cost correct to four significant figures, for this example we chose a value of $\delta_{mc} = 10^{-4}$. To achieve such an accuracy took over 2 hours of computer time and over 100 million pairs of (x_1, x_2) were generated. Clearly this method is not very efficient for this problem but it is easy to implement and suffices for our purposes. The Monte-Carlo integration method works well when the h–function in (3.39) does not vary much; plainly, here this will not be the case. The simplicity of the method for our problem is highlighted by the ease with which a change in the distribution of (X_1, X_2) can be incorporated. All that is required is to swap over the routine that generates observations of (X_1, X_2) to generate from the new distribution.

The Bayes cost was calculated as $\mathcal{K} = 0.0488c_r$, a reduction of almost 10% over the 1-dimensional screen which has Bayes cost $\mathcal{K} = 0.0542c_r$, calculated using a simple modification of the above algorithm and under the full two covariate model. Table 3.2 shows that using the additional screening variable increases the number of correctly classified patients at the first stage, with fewer going through to the more expensive second stage. We note in passing that none of the cases would have been wrongly diagnosed using this screen, though a more thorough assessment of the performance of the screen could be obtained by validating its performance on a large training set or via the model using simulation methods.

3.4 Limited resources

In some situations, limited resources of manpower or equipment may force us to impose limits on the proportion of items passed on to the second stage of the screen. Hence, it seems natural to consider the problem of minimising $\mathcal{K}(\Omega_A, \Omega_R, \Omega_M)$ subject to the constraint

$$P(\underline{X} \in \Omega_M) \leq \alpha, \tag{3.41}$$

for some $\alpha \in (0, 1)$. We note in passing that the methods of this section are capable of accommodating other constrained versions of the optimisation problem of section 3.2.

As a necessary preliminary, we consider a problem with an equality constraint, that is

$$\text{minimise } \mathcal{K}(\Omega_A, \Omega_R, \Omega_M) \text{ subject to } P(\underline{X} \in \Omega_M) = \alpha. \tag{3.42}$$

Any partition solving (3.42) will be called *α-optimal* and any partition that satisfies the constraint $P(\underline{X} \in \Omega_M) = \alpha$ will be called *feasible* for (3.42). To solve (3.42) we introduce the Lagrangian

$$\begin{aligned} L(\Omega_A, \Omega_R, \Omega_M, \lambda, \alpha) &= \mathcal{K}(\Omega_A, \Omega_R, \Omega_M) + \lambda \{P(\underline{X} \in \Omega_M) - \alpha\} \\ &= E_{\underline{X}} [\tilde{c}_a(\underline{X}) I(\underline{X} \in \Omega_A) + \tilde{c}_r(\underline{X}) I(\underline{X} \in \Omega_R) \\ &\qquad + \{\tilde{c}_m(\underline{X}) + \lambda\} I(X \in \Omega_M)] - \lambda\alpha \end{aligned}$$

from (3.1) and (3.3). By a simple extension of Theorem 3.1, the minimisation of L is achieved by

$$\begin{aligned} \Omega_A^\lambda &= \{\underline{x} \in \Omega : \tilde{c}_a(\underline{x}) = \min [\tilde{c}_a(\underline{x}), \tilde{c}_r(\underline{x}), \tilde{c}_m(\underline{x}) + \lambda]\}, \\ \Omega_R^\lambda &= \{\underline{x} \in \Omega : \tilde{c}_r(\underline{x}) = \min [\tilde{c}_a(\underline{x}), \tilde{c}_r(\underline{x}), \tilde{c}_m(\underline{x}) + \lambda]\}, \\ \Omega_M^\lambda &= \{\underline{x} \in \Omega : \tilde{c}_m(\underline{x}) + \lambda = \min [\tilde{c}_a(\underline{x}), \tilde{c}_r(\underline{x}), \tilde{c}_m(\underline{x}) + \lambda]\}. \end{aligned} \tag{3.43}$$

Lemma 3.4 $\rho(\lambda) \equiv P(\underline{X} \in \Omega_M^\lambda)$ is decreasing in λ.

Proof

Plainly, from (3.43), $\lambda > \mu \Rightarrow \Omega_M^\lambda \subseteq \Omega_M^\mu$ and the result follows. □

The following corollary is a consequence of the strong Lagrangian principle, see Whittle (1971).

Corollary 3.2 The partition $(\Omega_A^\lambda, \Omega_R^\lambda, \Omega_M^\lambda)$ will be α-optimal if it is feasible for the minimisation problem in (3.42), that is, if $\rho(\lambda) = \alpha$.

Proof

Denote as $(\tilde{\Omega}_A, \tilde{\Omega}_R, \tilde{\Omega}_M)$ any α-optimal partition. We write $\{\Omega_3\}$ for the set of all threefold partitions of $\Omega_{\underline{X}}$ and $\{\Omega_3^\alpha\}$ for the set of all feasible threefold partitions. If $(\Omega_A^\lambda, \Omega_R^\lambda, \Omega_M^\lambda)$ is feasible then clearly, as $(\tilde{\Omega}_A, \tilde{\Omega}_R, \tilde{\Omega}_M)$ solves (3.42),

$$\mathcal{K}(\tilde{\Omega}_A, \tilde{\Omega}_R, \tilde{\Omega}_M) \leq \mathcal{K}(\Omega_A^\lambda, \Omega_R^\lambda, \Omega_M^\lambda) \tag{3.44}$$

Also,

$$\begin{aligned}
\mathcal{K}(\tilde{\Omega}_A, \tilde{\Omega}_R, \tilde{\Omega}_M) &= \min_{\{\Omega_3^\alpha\}} \mathcal{K}(\Omega_A, \Omega_R, \Omega_M) \\
&= \min_{\{\Omega_3^\alpha\}} [\mathcal{K}(\Omega_A, \Omega_R, \Omega_M) + \lambda\{P(\underline{X} \in \Omega_M) - \alpha\}] \\
&\geq \min_{\{\Omega_3\}} [\mathcal{K}(\Omega_A, \Omega_R, \Omega_M) + \lambda\{P(\underline{X} \in \Omega_M) - \alpha\}] \\
&= L(\Omega_A^\lambda, \Omega_R^\lambda, \Omega_M^\lambda) \\
&= \mathcal{K}(\Omega_A^\lambda, \Omega_R^\lambda, \Omega_M^\lambda)
\end{aligned} \tag{3.45}$$

as $P(\underline{X} \in \Omega_M^\lambda) = \alpha$. It follows from (3.44) and (3.45) that the Bayes costs $\mathcal{K}(\tilde{\Omega}_A, \tilde{\Omega}_R, \tilde{\Omega}_M)$ and $\mathcal{K}(\Omega_A^\lambda, \Omega_R^\lambda, \Omega_M^\lambda)$ are equal and so $(\Omega_A^\lambda, \Omega_R^\lambda, \Omega_M^\lambda)$ is α-optimal. □

The key focus, then, is feasibility. In Theorem 3.3 we give conditions which are sufficient to establish the existence of a feasible partition of the form $(\Omega_A^\lambda, \Omega_R^\lambda, \Omega_M^\lambda)$ for any $\alpha \in (0, 1)$.

Theorem 3.3 If $\tilde{c}_a(\underline{x})$, $\tilde{c}_r(\underline{x})$ and $\tilde{c}_m(\underline{x})$ are bounded above and $\rho : \mathbb{R} \mapsto [0, 1]$ is continuous then $\rho(\mathbb{R}) = [0, 1]$.

Proof

Let $\bar{c}_a$, $\bar{c}_r$ and $\bar{c}_m$ be upper bounds for $\tilde{c}_a(\underline{x})$, $\tilde{c}_r(\underline{x})$ and $\tilde{c}_m(\underline{x})$ respectively. Let $\underline{\Lambda} = -\bar{c}_m - \epsilon$ for some $\epsilon > 0$. If $\lambda \leq \underline{\Lambda}$ then

$$\tilde{c}_m(\underline{x}) + \lambda \leq \bar{c}_m + \lambda \leq -\epsilon < \min\{\tilde{c}_a(\underline{x}), \tilde{c}_r(\underline{x})\} \text{ for all } \underline{x} \in \Omega_{\underline{X}},$$

as all costs are bounded below by zero. That is, when $\lambda \leq \underline{\Lambda}$, (3.43) gives $\Omega_M^\lambda = \Omega_{\underline{X}}$. Hence there exists a λ such that $\rho(\lambda) = 1$. Similarly if $\overline{\Lambda} = \min\{\bar{c}_a, \bar{c}_r\} + \epsilon$ then, when $\lambda \geq \overline{\Lambda}$,

$$\tilde{c}_m(\underline{x}) + \lambda \geq \tilde{c}_m(\underline{x}) + \min\{\bar{c}_a, \bar{c}_r\} + \epsilon > \min\{\bar{c}_a, \bar{c}_r\} \geq \min\{\tilde{c}_a(\underline{x}), \tilde{c}_r(\underline{x})\} \text{ for all } \underline{x} \in \Omega_{\underline{X}},$$

and so (3.43) gives $\Omega_M^\lambda = \emptyset$. Hence there also exists a λ such that $\rho(\lambda) = 0$. The result now follows by the continuity of ρ and the intermediate value theorem.

Theorem 3.4 is an immediate consequence of Theorem 3.3 and Corollary 3.2.

Theorem 3.4 Under the hypotheses of Theorem 3.3,

(i) for all $\alpha \in (0,1)$ there exists a $\lambda(\alpha)$ such that $\rho\{\lambda(\alpha)\} = \alpha$;

(ii) any partition $(\Omega_A^{\lambda(\alpha)}, \Omega_R^{\lambda(\alpha)}, \Omega_M^{\lambda(\alpha)})$ is α-optimal where $\lambda(\alpha)$ is as in (i).

Consider now the hypotheses of Theorem 3.3. The boundedness of the costs seems natural and indeed required in practice but the continuity of ρ needs further consideration. This continuity (which *is* satisfied for the probit regression model of section 3.3 when $\underline{X}$ is absolutely continuous) is related to the requirement that sets of the form

$$\{\underline{X} : \tilde{c}_a(\underline{X}) - \tilde{c}_m(\underline{X}) = \lambda\} \qquad \text{and} \qquad \{\underline{X} : \tilde{c}_r(\underline{X}) - \tilde{c}_m(\underline{X}) = \lambda\},$$

for given $\lambda \in \mathbb{R}$, should always have probability zero. This will frequently be violated, *inter alia*, when $\underline{X}$ is discrete. To illustrate this further, suppose X is univariate and discrete. Also, suppose that for some $x_i \in \Omega_X$, we have $P(X = x_i) = p_i$ and $\tilde{c}_r(x_i) > \tilde{c}_a(x_i) > \tilde{c}_m(x_i)$. Then, $\lambda_i = \tilde{c}_a(x_i) - \tilde{c}_m(x_i)$ is greater than zero and the set

$$\{X : \tilde{c}_a(X) - \tilde{c}_m(X) = \lambda_i\}$$

will have a probability of at least p_i. Plainly, $x_i \in \Omega_M^\lambda$ for all $\lambda < \lambda_i$, and $x_i \notin \Omega_M^\lambda$ for all $\lambda \geq \lambda_i$. Also, if we define

$$p \equiv \lim_{\epsilon \to 0} \rho(\lambda_i - \epsilon) = \lim_{\epsilon \to 0} P(X \in \Omega_M^{\lambda_i - \epsilon}),$$

then $\rho(\lambda_i) = p - p_{i+}$ where $p_{i+} \geq p_i$. Hence $\lim_{\epsilon \to 0} \rho(\lambda_i - \epsilon) - \rho(\lambda_i) = p_{i+}$ and there exists a discontinuity of ρ at λ_i.

Hence consider a problem in which the equation $\rho(\lambda) = \alpha$ has no solution, that is, $\alpha \notin \rho(\mathbb{R})$. Following Neyman–Pearson theory, we can come close to α-optimality via a suitable randomisation. Let $\alpha_1, \alpha_2 \in \rho(\mathbb{R})$ and $\alpha_1 < \alpha < \alpha_2$. By Lemma 3.4, $\lambda(\alpha_1) > \lambda(\alpha_2)$. Suppose that p_1 and p_2 satisfy

$$p_1\alpha_1 + p_2\alpha_2 = \alpha \qquad \text{and} \qquad p_1 + p_2 = 1. \tag{3.46}$$

That is,

$$p_1 = \frac{\alpha_2 - \alpha}{\alpha_2 - \alpha_1} \qquad \text{and} \qquad p_2 = \frac{\alpha - \alpha_1}{\alpha_2 - \alpha_1}. \tag{3.47}$$

We consider the randomised two–stage screen

$$p_1 \otimes \left(\Omega_A^{\lambda(\alpha_1)}, \Omega_R^{\lambda(\alpha_1)}, \Omega_M^{\lambda(\alpha_1)}\right) + p_2 \otimes \left(\Omega_A^{\lambda(\alpha_2)}, \Omega_R^{\lambda(\alpha_2)}, \Omega_M^{\lambda(\alpha_2)}\right),$$

where, with probability p_1 we use the α_1-optimal screen and with probability p_2 we use the α_2-optimal screen. Note that the only items affected by the randomisation are those for which $\underline{x} \in \Omega_M^{\lambda(\alpha_2)} \setminus \Omega_M^{\lambda(\alpha_1)}$ and the proportion of such items will be small if $\lambda(\alpha_1) - \lambda(\alpha_2)$ is small.

By (3.46), this randomised screen satisfies the α–constraint in (3.42) and has associated Bayes cost

$$p_1 \mathcal{K}\left(\Omega_A^{\lambda(\alpha_1)}, \Omega_R^{\lambda(\alpha_1)}, \Omega_M^{\lambda(\alpha_1)}\right) + p_2 \mathcal{K}\left(\Omega_A^{\lambda(\alpha_2)}, \Omega_R^{\lambda(\alpha_2)}, \Omega_M^{\lambda(\alpha_2)}\right). \tag{3.48}$$

Lemma 3.5 The Bayes cost in (3.48) comes within

$$\frac{1}{4}(\alpha_2 - \alpha_1)\{\lambda(\alpha_1) - \lambda(\alpha_2)\} \tag{3.49}$$

of the cost of an α-optimal screen, should one exist.

Proof

Let $(\tilde{\Omega}_A, \tilde{\Omega}_R, \tilde{\Omega}_M)$ be an α-optimal screen, assuming one exists. By the properties of $(\Omega_A^{\lambda(\alpha_i)}, \Omega_R^{\lambda(\alpha_i)}, \Omega_M^{\lambda(\alpha_i)})$, $i = 1, 2$, with respect to the Lagrangian we have

$$\begin{aligned} L(\Omega_A^{\lambda(\alpha_1)}, \Omega_R^{\lambda(\alpha_1)}, \Omega_M^{\lambda(\alpha_1)}, \lambda(\alpha_1), \alpha_1) &\leq L(\tilde{\Omega}_A, \tilde{\Omega}_R, \tilde{\Omega}_M, \lambda(\alpha_1), \alpha_1) \\ \Rightarrow \qquad \mathcal{K}\left(\Omega_A^{\lambda(\alpha_1)}, \Omega_R^{\lambda(\alpha_1)}, \Omega_M^{\lambda(\alpha_1)}\right) &\leq \mathcal{K}\left(\tilde{\Omega}_A, \tilde{\Omega}_R, \tilde{\Omega}_M\right) + \lambda(\alpha_1)\left\{P(\underline{X} \in \tilde{\Omega}_M) - \alpha_1\right\} \\ &= \mathcal{K}\left(\tilde{\Omega}_A, \tilde{\Omega}_R, \tilde{\Omega}_M\right) + \lambda(\alpha_1)(\alpha - \alpha_1) \qquad (3.50) \end{aligned}$$

and, similarly,

$$\mathcal{K}\left(\Omega_A^{\lambda(\alpha_2)}, \Omega_R^{\lambda(\alpha_2)}, \Omega_M^{\lambda(\alpha_2)}\right) \leq \mathcal{K}\left(\tilde{\Omega}_A, \tilde{\Omega}_R, \tilde{\Omega}_M\right) + \lambda(\alpha_2)(\alpha - \alpha_2). \qquad (3.51)$$

Multiplying (3.50) by p_1, (3.51) by p_2, the Bayes cost in (3.48) is within

$$p_1\lambda(\alpha_1)(\alpha - \alpha_1) + p_2\lambda(\alpha_2)(\alpha - \alpha_2)$$

of $\mathcal{K}(\tilde{\Omega}_A, \tilde{\Omega}_R, \tilde{\Omega}_M)$. We deduce from (3.47) that this is

$$\{\lambda(\alpha_1) - \lambda(\alpha_2)\}\frac{(\alpha_2 - \alpha)(\alpha - \alpha_1)}{(\alpha_2 - \alpha_1)}.$$

It is easy to show that $(\alpha_2 - \alpha)(\alpha - \alpha_1) \leq (\alpha_2 - \alpha_1)^2/4$, where the bound is attained when $\alpha = (\alpha_1 + \alpha_2)/2$. The result follows. □

From the definition of ρ it is clear that $\rho(\mathbb{R}) \subseteq [0, 1]$. Hence, $\rho(\lambda)$ is real valued for all real λ. By Lemma 3.4, $\rho(\lambda)$ is decreasing in λ and it is easy to establish that when $\alpha \notin \rho(\mathbb{R})$ we must have that

$$\sup\{\lambda;\ \rho(\lambda) > \alpha\} = \inf\{\lambda;\ \rho(\lambda) < \alpha\}.$$

Hence $\lambda(\alpha_1)$, $\lambda(\alpha_2)$ may be chosen to make (3.49) arbitrarily small. Hence we have ϵ-Bayes optimality via this randomisation device.

We are now able to turn to the problem of primary interest to us, that of choosing $(\Omega_A, \Omega_R, \Omega_M)$ to

$$\text{minimise } \mathcal{K}(\Omega_A, \Omega_R, \Omega_M) \text{ subject to } P(\underline{X} \in \Omega_M) \leq \alpha. \qquad (3.52)$$

We shall suppose henceforth that the hypotheses of Theorem 3.3 are satisfied, bearing in mind that any discontinuity in ρ may be accommodated via the randomisation device above. In Theorem 3.5, $(\Omega_A^*, \Omega_R^*, \Omega_M^*)$ is an (unconstrained) Bayes two-stage screen, as in Theorem 3.1.

Theorem 3.5 Under the hypotheses of Theorem 3.3, our constrained problem (3.52) is solved by

(a) the globally optimal Bayes two-stage screen $(\Omega_A^*, \Omega_R^*, \Omega_M^*)$ if $P(\underline{X} \in \Omega_M^*) \leq \alpha$,

(b) any α-optimal screen $(\Omega_A^{\lambda(\alpha)}, \Omega_R^{\lambda(\alpha)}, \Omega_M^{\lambda(\alpha)})$ when $P(\underline{X} \in \Omega_M^*) > \alpha$.

Proof

Part (a) is trivial – if the globally optimal solution is feasible for the constrained problem (3.52) then it must be optimal for it.

Suppose now that the globally optimal solution is infeasible, with

$$P(\underline{X} \in \Omega_M^*) > \alpha. \tag{3.53}$$

Using the notation in (3.43), it is clear from the Lagrangian formulation that (under consistent tie-breaking rules)

$$(\Omega_A^*, \Omega_R^*, \Omega_M^*) = \left(\Omega_A^0, \Omega_R^0, \Omega_M^0\right).$$

Hence (3.53) implies that

$$\rho(0) > \alpha.$$

It then follows from Lemma 3.4 that $\lambda(\alpha) > 0$. Hence if $(\Omega_A, \Omega_R, \Omega_M)$ is any feasible solution to (3.52) it follows that

$$\begin{aligned} & L(\Omega_A^{\lambda(\alpha)}, \Omega_R^{\lambda(\alpha)}, \Omega_M^{\lambda(\alpha)}, \lambda(\alpha), \alpha) \leq L(\Omega_A, \Omega_R, \Omega_M, \lambda(\alpha), \alpha) \\ \Rightarrow \quad & \mathcal{K}(\Omega_A^{\lambda(\alpha)}, \Omega_R^{\lambda(\alpha)}, \Omega_M^{\lambda(\alpha)}) \leq \mathcal{K}(\Omega_A, \Omega_R, \Omega_M) + \lambda(\alpha)\{P(\underline{X} \in \Omega_M) - \alpha\} \\ \Rightarrow \quad & \mathcal{K}(\Omega_A^{\lambda(\alpha)}, \Omega_R^{\lambda(\alpha)}, \Omega_M^{\lambda(\alpha)}) \leq \mathcal{K}(\Omega_A, \Omega_R, \Omega_M) \end{aligned}$$

as required. □

Comments

1. When the globally optimal screen is not feasible, that is, $P(\underline{X} \in \Omega_M^*) > \alpha$, notice that the constrained optimum is obtained by increasing the penalty of measuring the performance variable. In fact, the initial assessment of $\tilde{c}_m(\underline{x})$ should have taken into account any limitation on resources and the procedure in this section would not be needed. Hence the method here is useful as a way of ensuring that costs parameters have been chosen to fully reflect resource limitations.

2. Our method restricts throughput to the second stage of the screen, by constraining at α the probability of an item passing to the second stage. In practice, there may be a more immediate restriction, on the number of items from a batch that may be admitted to the second stage. Here, if the globally optimal procedure is not feasible, that is, it admits too many items from the batch to the second stage, one suggestion would be to increase λ in (3.43) until only the required number of items are passed to the second stage. Further work needs to be done to establish whether such a scheme would be optimal.

3.5 Probit regression model

Here we illustrate screening under limited resources assuming the standard cost structure and a probit regression model. The case in which X is univariate is covered in detail by section 3.5.1 and in section 3.5.2 we look at limited resources in the context of the Conn's syndrome example.

Firstly, we assume that $c_m/c_r < 1 - c_m/c_a$, ensuring that it is optimal to have a two–stage screen (with $\Omega_M^* \neq \emptyset$). In section 3.3.1 we saw that, when X is univariate and m_1/s_1 is small enough, it may be better not to perform the X-stage of the screen and reject, accept or measure T for all items. Here we shall assume that the standardised regression parameters m_i/s_i are large enough so that it will always be optimal to perform the $\underline{X}$-stage of the screen. We limit the number of items passing to the second stage of the screen by the constraint $P(X \in \Omega_M) \leq \alpha$. By

Theorem 3.5, when $\rho(0) > \alpha$, the optimal two–stage screen subject to this constraint is given by any α-optimal two–stage screen. In this case, following (3.9) and (3.43), when $(c_m + \lambda)/c_r < 1 - (c_m + \lambda)/c_a$, the Lagrangian is minimised by

$$\begin{aligned} \Omega_A^\lambda &= \left\{ \underline{x} \in \Omega_{\underline{X}} : \frac{\underline{m}^T \underline{x}_0}{(1 + \underline{x}_0^T \boldsymbol{S} \underline{x}_0)^{1/2}} \geq c_2(\lambda) \right\} \\ \Omega_R^\lambda &= \left\{ \underline{x} \in \Omega_{\underline{X}} : \frac{\underline{m}^T \underline{x}_0}{(1 + \underline{x}_0^T \boldsymbol{S} \underline{x}_0)^{1/2}} \leq c_1(\lambda) \right\} \end{aligned} \tag{3.54}$$

$$\Omega_M^\lambda = \left\{ \underline{x} \in \Omega_{\underline{X}} : c_1(\lambda) < \frac{\underline{m}^T \underline{x}_0}{(1 + \underline{x}_0^T \boldsymbol{S} \underline{x}_0)^{1/2}} < c_2(\lambda) \right\},$$

where $c_2(\lambda) = \Phi^{-1}\{1-(c_m+\lambda)/c_a\}$ and $c_1(\lambda) = \Phi^{-1}\{(c_m+\lambda)/c_r\}$, with $c_1(\lambda) < c_2(\lambda)$. Simple algebra yields that for $\lambda < \hat{\lambda}$,

$$\frac{c_m + \lambda}{c_r} < 1 - \frac{c_m + \lambda}{c_a} \tag{3.55}$$

where $\hat{\lambda} = c_a c_r/(c_a + c_r) - c_m > 0$. To find the α-optimal screen we require $\lambda(\alpha)$, the value of λ that solves $\rho(\lambda) = \alpha$. As X is absolutely continuous, the continuity of ρ is guaranteed and so, by Theorem 3.4, there exists a λ such that $\rho(\lambda) = \alpha$ for all $\alpha \in (0, 1)$. Plainly, $\Omega_M^{\hat{\lambda}} = \emptyset \Rightarrow \rho(\hat{\lambda}) = 0$ and we know that $\rho(0) > \alpha$, hence $\rho(0) > \rho\{\lambda(\alpha)\} > 0$ and by Lemma 3.4, $0 < \lambda(\alpha) < \hat{\lambda}$. Hence (3.55) will hold for $\lambda = \lambda(\alpha)$, and so the α-optimal screen is determined by the value of λ that solves

$$P\left\{ c_1(\lambda) < \frac{\underline{m}^T \underline{X}_0}{(1 + \underline{X}_0^T \boldsymbol{S} \underline{X}_0)^{1/2}} < c_2(\lambda) \right\} = \alpha,$$

where $\underline{X}_0 = (1, \underline{X})^T$.

3.5.1 Constrained screening with one covariate

Here we discuss the effect of limited resources when X is univariate. We consider the case in which $m_0 = 0$ and $s_1 = 1$ and suppose that $m_1 > \max(|c_1|, |c_2|)$, which in turn guarantees that $m_1 > \max(|c_1(\lambda)|, |c_2(\lambda)|)$, $0 \leq \lambda < \hat{\lambda}$. From Theorem 3.2 and Corollary 3.1 we infer that in the range $\lambda \in [0, \hat{\lambda})$, the screen $(\Omega_A^\lambda, \Omega_R^\lambda, \Omega_M^\lambda)$ minimising $L(\Omega_A, \Omega_R, \Omega_M, \lambda, \alpha)$ is given by:

$$\Omega_A^\lambda = [w_\lambda, \infty), \qquad \Omega_R^\lambda = (-\infty, v_\lambda], \qquad \Omega_A^\lambda = (v_\lambda, w_\lambda),$$

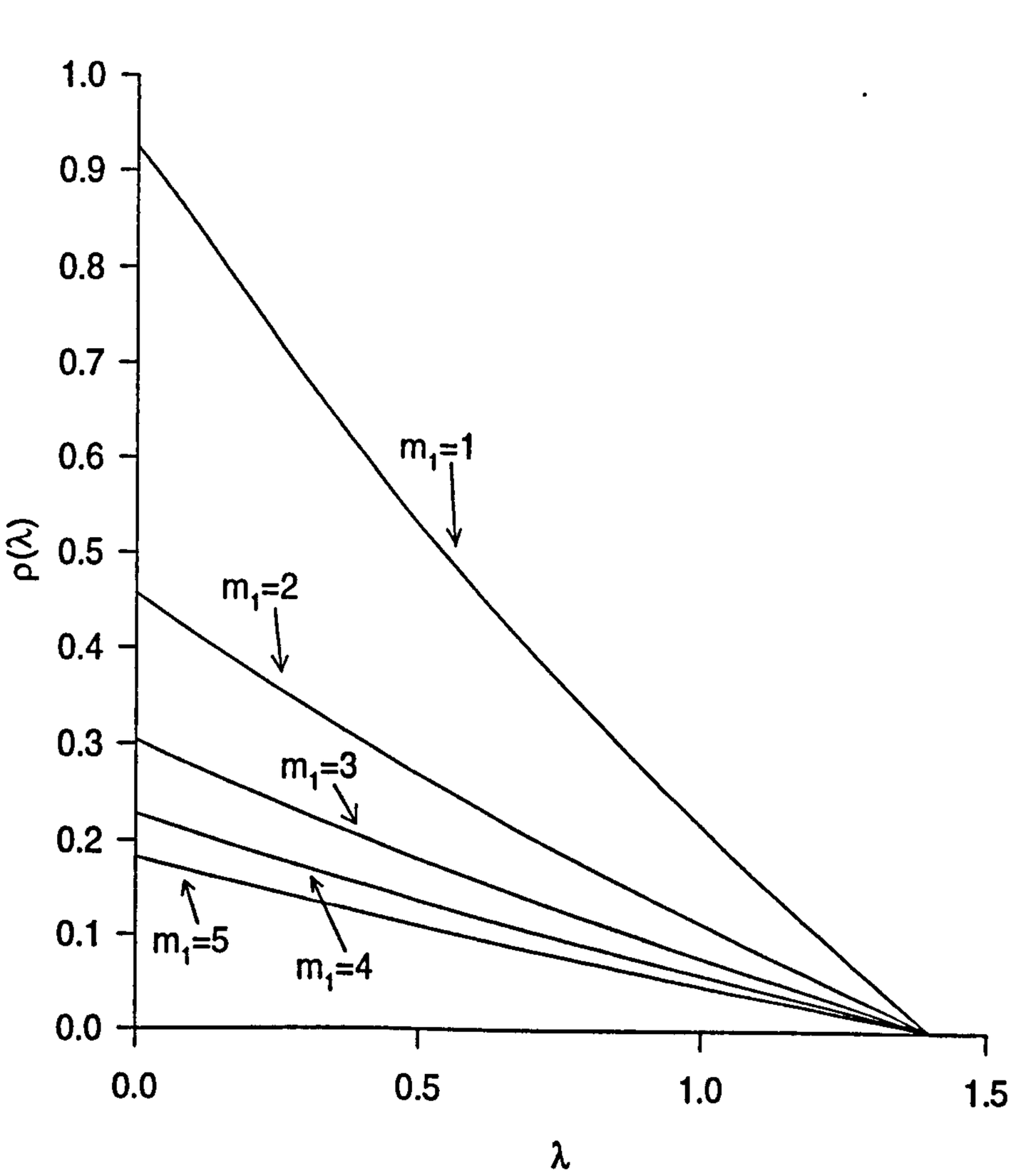


Figure 3.3: Plot of $\rho(\lambda)$ against λ

where w_λ and v_λ can be found by substituting $c_2(\lambda)$ for c_2 in Theorem 3.2 and $c_1(\lambda)$ for c_1 in Corollary 3.1. Hence

$$\rho(\lambda) = P\{X \in (v_\lambda, w_\lambda)\}, \qquad 0 \leq \lambda < \hat{\lambda}.$$

In the case in which the screening variable $X \sim N(0,1)$, for given λ, we can simply compute v_λ and w_λ from Theorem 3.2 and Corollary 3.1, and then obtain

$$\rho(\lambda) = \Phi(w_\lambda) - \Phi(v_\lambda).$$

In Figure 3.3, find plots of $\rho(\lambda)$, $0 \leq \lambda < \hat{\lambda}$, for examples in which $X \sim N(0,1)$, the regression parameters have mean $m_0 = 0$, $m_1 = 1,2,3,4,5$ and covariance structure $s_0 = s_1 = 1$, $r = 0.5$, with costs $(c_m, c_a, c_r) = (1,4,6)$.

We now consider a scenario in which resources are limited in such a way that we are interested in the constrained minimisation problem (3.52) with $\alpha = 0.25$. From Figure 3.3, we see that when $m_1 = 4$ or 5, $\rho(0) \leq 0.25$ and so the globally optimal screen $(\Omega_A^*, \Omega_R^*, \Omega_M^*)$ is feasible and hence is optimal for the constrained problem, see Theorem 3.5(a). However, when $m_1 = 1$, 2 or 3 then $\rho(0) > 0.25$ and, by Theorem 3.5(b), a 0.25-optimal screen $(\Omega_A^{\lambda(0.25)}, \Omega_R^{\lambda(0.25)}, \Omega_M^{\lambda(0.25)})$ solves the problem. We can find $\lambda(0.25)$ directly from the plot or use a numerical equation solving technique, such as halving, to find

$$\lambda(0.25) = (\lambda : \Phi(w_\lambda) - \Phi(v_\lambda) = 0.25).$$

The plot could be used to choose sensible starting points for such a routine. For example, when $m_1 = 2$ we have $\lambda(0.25) = 0.554$ and the optimal screen for the constrained problem is

$$\Omega_A^{\lambda(0.25)} = [0.213, \infty), \qquad \Omega_R^{\lambda(0.25)} = (-\infty, -0.428], \qquad \Omega_M^{\lambda(0.25)} = (-0.428, 0.213).$$

These results are consistent with the notion that large values of m_1 correspond to more effective $\underline{X}$–filters for the attributes. Hence in cases when m_1 is large, less use is made of the second stage.

3.5.2 Conn's syndrome example

Consider again the Conn's syndrome data described in section 2.4 and section 3.3.3. A Bayes optimal design based only on the covariate X_1, the standardised log concentration of potassium, is given by Figure 2.2. Assuming that X_1 is standard Normal, the probability of a patient being passed on to the second stage of the screen is

$$\rho(0) = P(-0.31 < X_1 < 1.44) = 0.55,$$

which may well be too large to be practicable. Suppose that the throughput to the second stage must not exceed 25%. To attain this level of throughput and to maintain Bayes optimality under this constraint, the cost of reaching the second stage must be increased by an amount $\lambda(0.25)$ and the screening parameters recalculated. The amount $\lambda(0.25)$ is found by performing a simple search algorithm:

1. Take as starting values $\lambda_0 = 0$ and $\lambda_1 = \hat{\lambda} = 0.3526c_r$. Hence $\rho(\lambda_0) = 0.55$ and $\rho(\lambda_1) = 0$.

2. Obtain standardised values of the regression parameters that ensure that $m_0 = 0$ and $s_1 = 1$, see section 2.3.8. For this example, the untransformed and transformed regression parameters are given by (2.34) and (2.35) respectively.

3. Calculate the new point as

$$\lambda_2 = \frac{\lambda_0 \{0.25 - \rho(\lambda_1)\} + \lambda_1 \{\rho(\lambda_0) - 0.25\}}{\rho(\lambda_0) - \rho(\lambda_1)}$$

4. Find the values of the cut–points v'_{λ_2} and w'_{λ_2} from Theorem 3.2 and Corollary 3.1, replacing c_2 by $c_2(\lambda_2) = \Phi^{-1}\{1 - (c_m + \lambda_2)/c_a)\}$ and c_1 by $c_1(\lambda_2) = \Phi^{-1}\{(c_m + \lambda_2)/c_r)\}$. These values are cut-points under the data transformation that ensures $m_0 = 0$ and $s_1 = 1$.

5. Reverse the data transformation of section 2.3.8 to obtain cut–points relating to the variable X_1:

$$v_{\lambda_2} = \frac{v'_{\lambda_2}}{s_1} - \frac{m_0}{m_1}$$

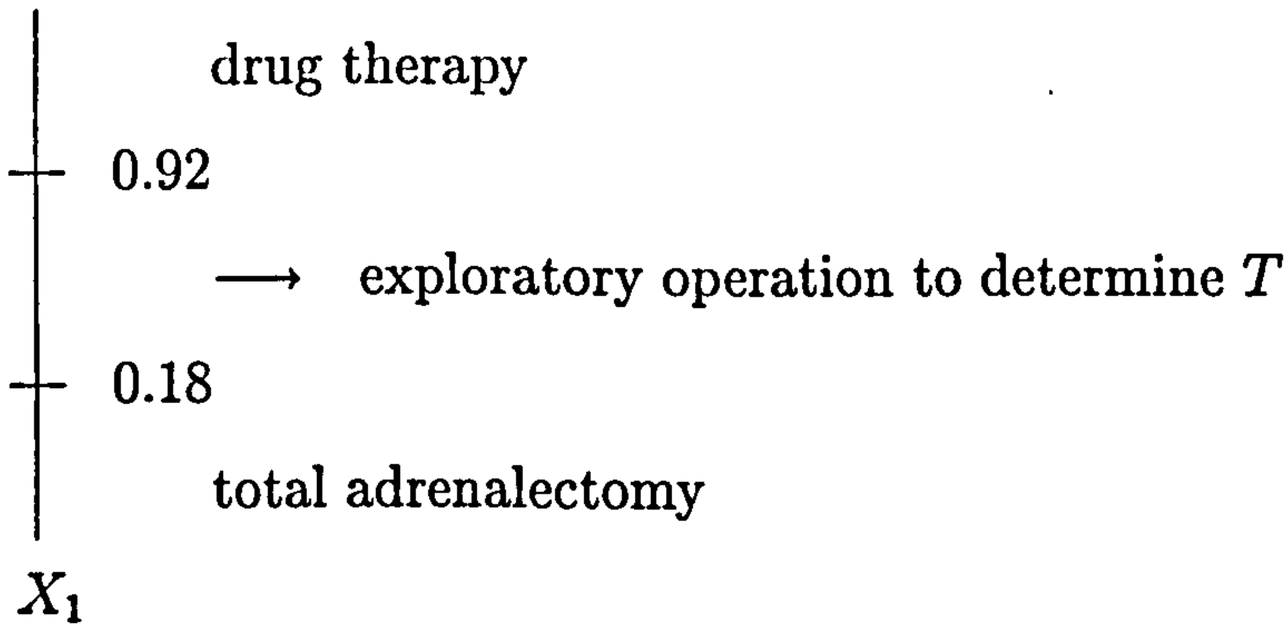
$$w_{\lambda_2} = \frac{w'_{\lambda_2}}{s_1} - \frac{m_0}{m_1}$$

6. Calculate $\rho(\lambda_2) = \Phi(w_{\lambda_2}) - \Phi(v_{\lambda_2})$.

7. If $\rho(\lambda_2) > \alpha$ then set $\lambda_0 = \lambda_2$, otherwise set $\lambda_1 = \lambda_2$.

8. If some convergence criteria is met, say $|\rho(\lambda_1) - \rho(\lambda_0)| < \delta_\lambda$ for some small δ_λ then set $\lambda(0.25) = \lambda_2$ and stop, otherwise return to step 3.

Notice that in each iteration $\rho(\lambda_0) > 0.25$ and $\rho(\lambda_1) < 0.25$. The algorithm computes a new point as the point where a straight line drawn between the two current points crosses the line $\rho(\lambda) = 0.25$. Using the above algorithm on the Conn's syndrome data and imposing the convergence criteria in step 8 with $\delta_\lambda = 10^{-5}c_r$, we found that $\lambda(0.25) = 0.1515c_r$, i.e. $\rho(0.1515c_r) = 0.25$. Reading off the cut–points from step 5 of the final iteration of the algorithm, a 0.25-optimal 1 dimensional Bayes two-stage screen is



and has Bayes cost $\mathcal{K} = 0.0740c_r$ (computed as in section 3.3.3). The overall classification of the 31 cases on this screen is given in Table 3.3. Notice that, by restricting the throughput to the second stage, the Bayes cost has increased by around 37%. The number of cases correctly classified has increased from 14 to 22 but the number of misclassified cases has also increased from 0 to 2.

A bivariate version of this screen based on the two covariates X_1 and X_2 in (3.35) is calculated as follows. Again we shall assume that $(X_1, X_2)^T$ follows a bivariate standard normal distribution with known correlation coefficient $r_x = -0.678$, the sample correlation coefficient. The 2-dimensional Bayes two–stage screen is given by (3.37) in section 3.3.3. That screen passes $\rho(0) = P(\underline{X} \in \Omega_M) = 47\%$ of patients on to the second stage. Again suppose we wish to limit the throughput to not exceed

25%. The search algorithm proceeds as above but with step 2 deleted and steps 4–6 replaced by the calculation of

$$\rho(\lambda_2) = P\left\{c_1(\lambda_2) < g(X_1, X_2) < c_2(\lambda_2)\right\},$$

where $g(x_1, x_2)$ is as defined in (3.38). We calculate this probability using the Monte–Carlo integration method first described in section 3.3.3. We sample pseudo observations from the bivariate normal distribution with density $\phi_2(x_1, x_2|r_x = -0.678)$ and, for a generated pair $\left(x_1^{(i)}, x_2^{(i)}\right)$, the h–function in (3.39) is replaced by:

$$h\left(x_1^{(i)}, x_2^{(i)}\right) = \begin{cases} 1 & \text{if } c_1(\lambda_2) < g\left(x_1^{(i)}, x_2^{(i)}\right) < c_2(\lambda_2), \\ 0 & \text{otherwise.} \end{cases}$$

Here the function $h(X_1, X_2)$ is a Bernoulli random variable with probability $\rho(\lambda_2)$ of a 'success'. Hence its variance is simply given by $\rho(\lambda_2)\{1-\rho(\lambda_2)\}$. An initial search was performed with δ_{mc} set to 10^{-2} in a modified form of the relative error convergence criterion (3.40), so that we find $\rho(\lambda)$ accurate to 2 significant figures and then, using values from that search as starting points, we performed a search with $\delta_{mc} = 10^{-3}$. In this latter search, the number of pairs generated to find each value of $\rho(\lambda_2)$ was typically of the order of $\left\{1.96\sqrt{0.25(1-0.25)}/(0.25 \times 10^{-3})\right\}^2 = 11,524,800$. The search found that $\rho(0.1175c_T) = 0.25$ and so we take $\lambda(0.25) = 0.1175c_T$. Therefore a 0.25-optimal 2-dimensional Bayes two-stage screen is

$$\begin{aligned} \Omega_A^* &= \left\{\underline{x} \in \Omega_{\underline{X}} : g(x_1, x_2) \geq 0.654\right\} \\ \Omega_R^* &= \left\{\underline{x} \in \Omega_{\underline{X}} : g(x_1, x_2) \leq -0.869\right\} \\ \Omega_M^* &= \left\{\underline{x} \in \Omega_{\underline{X}} : -0.869 < g(x_1, x_2) < 0.654\right\}. \end{aligned}$$

This screen is shown graphically in Figure 3.2. As expected, the central region which passes patients on to the second stage has been reduced in size to accommodate the restriction in throughput. The performance of this screen on the 31 cases is given in Table 3.3. The screen has Bayes cost $\mathcal{K} = 0.0600c_T$ (calculated as in section 3.3.3, an improvement of nearly 20% over the 0.25-optimal 1-dimensional screen. In terms of the cases, the use of the extra dimension has resulted in correctly classifying one of the

Table 3.3: Screen classifications under limited resources

		1–d screen classification			2–d screen classification		
Actual	Total	$T=1$	$T=0$	measure T	$T=1$	$T=0$	measure T
$T=1$	20	18	1	1	18	1	1
$T=0$	11	1	4	6	0	5	6

previously misclassified cases. Comparing the limited resources 2-dimensional screen with its unlimited version, we see that restricting the throughput has increased the Bayes cost by around 23%. Also, as in the 1-dimensional comparison, the number of correctly and incorrectly classified cases has increased: the number correctly classified from 15 to 23 and the number incorrectly classified cases from 0 to 1.

Chapter 4

Sequential screening

4.1 Introduction

In Chapter 2 and Chapter 3, we have developed optimal designs for two-stage procedures, the second stage of which is an observation of the performance variable $\underline{T}$. In this chapter, as in Chapter 3, we consider the case in which we have more than one screening variable, or covariate, available with which to assess whether an item reaches the standard. Chapter 3 was concerned with the design of a two–stage screen in which the first stage is a classification based on a batch of covariate measurements. Here we take a different approach and discuss the design of a sequential screen. In such a screen, at each stage of a sequence, a covariate is measured and a decision is made to either (i) sentence the item as acceptable, (ii) sentence the item as unacceptable, or (iii) pass the item on to the next stage. At the final stage of the sequence the performance variable is measured on those items that remain unsentenced. In many practical situations, the only form of the screen that is considered both simple enough to be workable and that makes full use of the available information (that is, full use of the covariates) is a sequential screen. As we shall see, the problem of finding the optimal design of a sequential screen is not trivial and in order to simplify the problem we assume that the order of the screens is fixed in advance of the design stage.

Recall that, for the case in which T is binary, the optimal two stage designs of Chapter 2 were seen to be simple and easily understood in relation to the cost parameters and the nature of the regression model. Also, the performance of those two stage designs has been found to be robust to departures from model assumptions (Boys & Glazebrook (1992)) and so it seems natural to try to use those designs in a sequential screen. Hence, in this chapter, we take advantage of the considerable strengths of the optimal two stage designs given in Chapter 2, using them as building blocks for a heuristic solution to the problem of designing an optimal sequential screen with a fixed screen order.

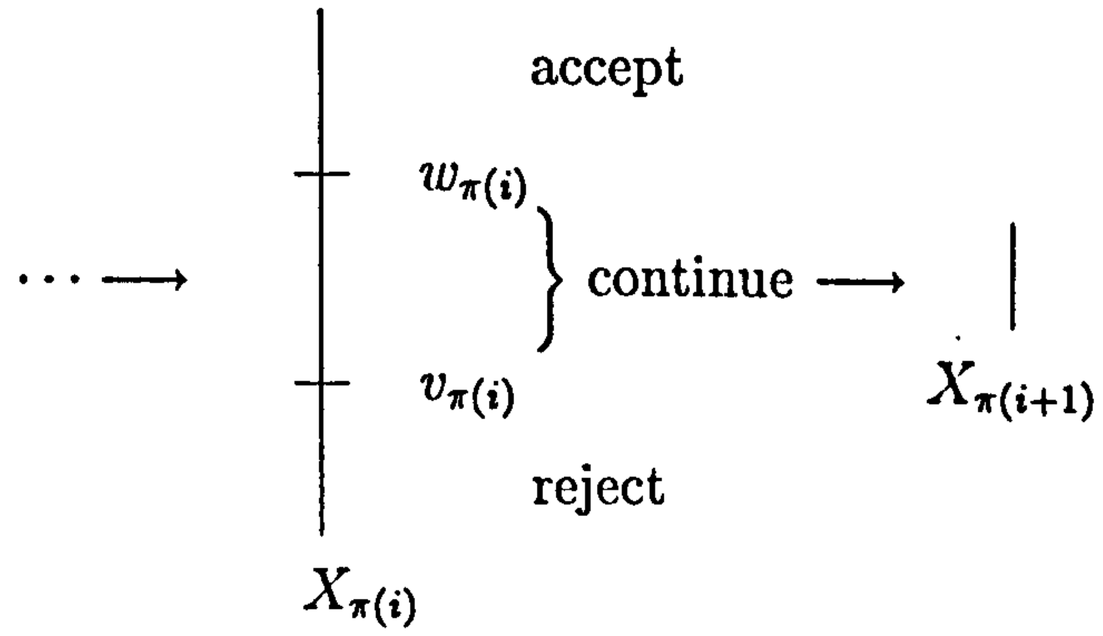
Consider a situation in which the screening variable $\underline{X} = (X_1, X_2, \ldots, X_p)$ is multivariate and T is binary. We wish to consider the problem of how best to design a sequence of screens, one for each X_i. Let π be a permutation of $\{1, 2, \ldots, p\}$ corresponding to an ordering of the screens, that is, $X_{\pi(1)}, X_{\pi(2)}, \ldots, X_{\pi(p)}$ is the sequence of screens corresponding to permutation π. We seek a Bayes sequential design of the form:

(i) an item that is not sentenced by the first $(i-1)$ screens based on $X_{\pi(1)}, X_{\pi(2)}, \ldots, X_{\pi(i-1)}$ respectively, is passed on to the $X_{\pi(i)}$-screen;

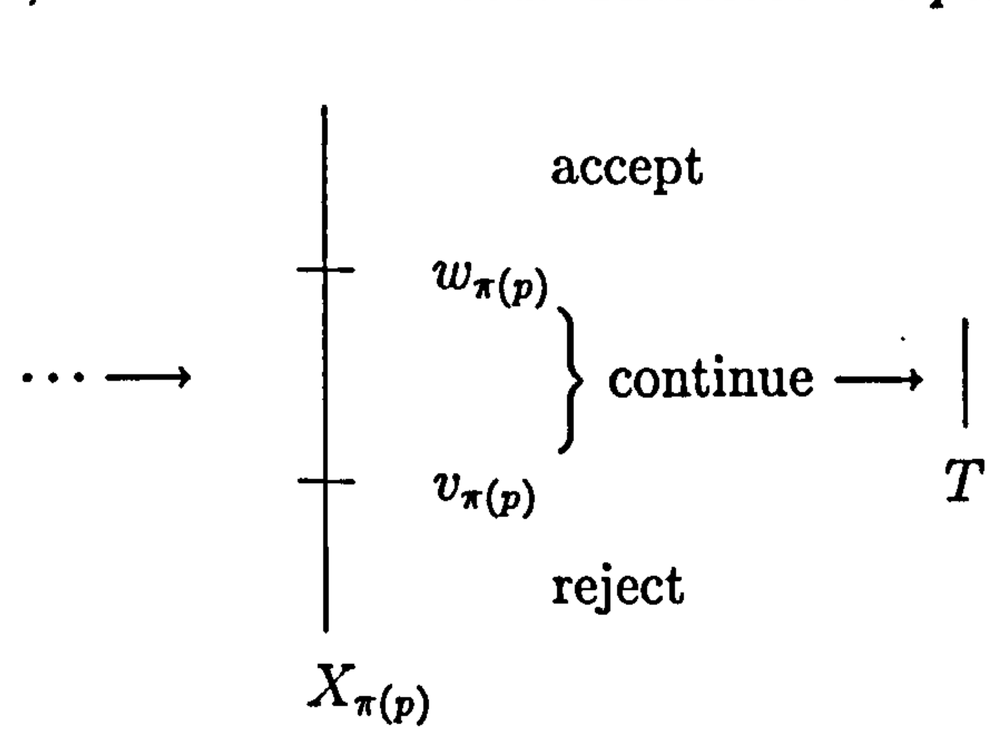
(ii) an item that is not sentenced by any of the p screens is passed on to the $(p+1)$th stage where T is measured.

We assume, for each X_i, $i = 1, 2, \ldots, p$, that large values of X_i tend to indicate that $T = 1$. In other words, each component of the screening variable $\underline{X}$ can be thought of as a larger–the–better screening variable when considered individually. We further assume that this assumption will hold when conditioning on the value of other components in $\underline{X}$, that is, $X_i|\underline{X}_{[i]} = \underline{x}_{[i]}$, $i = 1, 2, \ldots, p$, is a larger–the-better screening variable for all values of $\underline{x}_{[i]}$, where $\underline{X}_{[i]}$ is a vector of some or all of the components of $\underline{X}$ excluding X_i. If an item has not been sentenced on the basis of the first $i-1$ screens for $X_{\pi(1)}, X_{\pi(2)}, \ldots, X_{\pi(i-1)}$ respectively then we assume the

natural form for the ith stage of the sequential screen,



for suitably chosen $v_{\pi(i)} \leq w_{\pi(i)}$, $i = 1, 2, \ldots, p$. In other words, for $v_{\pi(i)} < w_{\pi(i)}$, if $X_{\pi(i)} \geq w_{\pi(i)}$ then the item is accepted, if $X_{\pi(i)} \leq v_{\pi(i)}$ then the item is rejected, and if $v_{\pi(i)} < X_{\pi(i)} < w_{\pi(i)}$ then the item is passed on to screen $i + 1$ for further testing; when $v_{\pi(i)} = w_{\pi(i)}$ we adopt the convention that if $X_{\pi(i)} \geq w_{\pi(i)}$ then the item is accepted, otherwise it is rejected. Should sentence not have been passed on the item at any of the p screens, then T is measured. Therefore the pth screen is assumed to have the form:



Chapter 2 concerns the case $p = 1$ and in Chapter 3 we have shown this form of screen to be optimal for the $p = 1$ case under conditions on the parameters of a probit model for $T|\underline{X}$.

There are two design questions. The first concerns the best ordering π of the p screens (see Bergman & Gittins (1985)) and the second, the optimum choice of $(\underline{v}_\pi, \underline{w}_\pi) = (v_{\pi(1)}, v_{\pi(2)}, \ldots, v_{\pi(p)}, w_{\pi(1)}, w_{\pi(2)}, \ldots, w_{\pi(p)})$ for a given screen order. We (in the main) consider the second of these questions only, noting that in many contexts there will in any event be strong constraints on the orderings allowed. The sequencing problem is nevertheless part of ongoing work in the area. In the remainder of this

chapter, π will be taken to be the identity permutation, that is, without loss of generality, we assume that the ordering of screens is $X_1, X_2, \ldots, X_p$ and the parameters that characterise a sequential screen are denoted $(\underline{v}, \underline{w}) = (v_1, v_2, \ldots, v_p, w_1, w_2, \ldots, w_p)$.

In section 4.2, we develop a heuristic design by (iteratively) using an approximate dynamic programming approach which performs successive backward passes through the problem, that is, redesign screen p, then $p-1$, and so on. At each stage, within each iteration, a delta method approximation is used to reduce each update to a version of the two–stage model of Chapter 2. Hence at each stage of the iterative procedure, and most importantly, at the final stage, the design of each screen is based on the simple, robust methods employed for the two–stage designs. Section 4.3 discusses the $p = 2$ case in detail and section 4.4 describes computational work that assesses the performance of our heuristic design and its sensitivity to changes in model parameters. In section 4.5 we design a sequential screen for the Conn's syndrome data and section 4.6 gives a couple of suggestions for improvements or alterations to our approach.

4.2 A heuristic multi–stage screen

In this section we develop a heuristic approach to the design of a sequential screen based on a dynamic programming methodology. In order to do that we first introduce the following vector notation: suppose that $\underline{r} = (r_1, r_2, \ldots, r_p)$ is a p-vector. We write

$$
\begin{aligned}
{}^{i}\underline{r} &= (r_1, r_2, \ldots, r_i) \text{ for the vector of the first } i \text{ elements of } \underline{r}, \\
\underline{r}^{i} &= (r_i, r_{i+1}, \ldots, r_p) \text{ for the vector of the last } p-i+1 \text{ elements of } \underline{r}, \\
\underline{r}_{(i)} &= (r_1, r_2, \ldots, r_{i-1}, r_{i+1}, \ldots, r_p) \text{ for the vector of the elements of } \underline{r} \text{ with } r_i \text{ deleted.}
\end{aligned}
$$

Also suppose that $\underline{q} = (q_1, q_2, \ldots, q_p)$ and $\underline{t} = (t_1, t_2, \ldots, t_p)$ are p-vectors, then we write

$$
\left({}^{(i-1)}\underline{q}, r_i, \underline{t}^{(i+1)} \right) = (q_1, q_2, \ldots, q_{i-1}, r_i, t_{i+1}, \ldots, t_p) .
$$

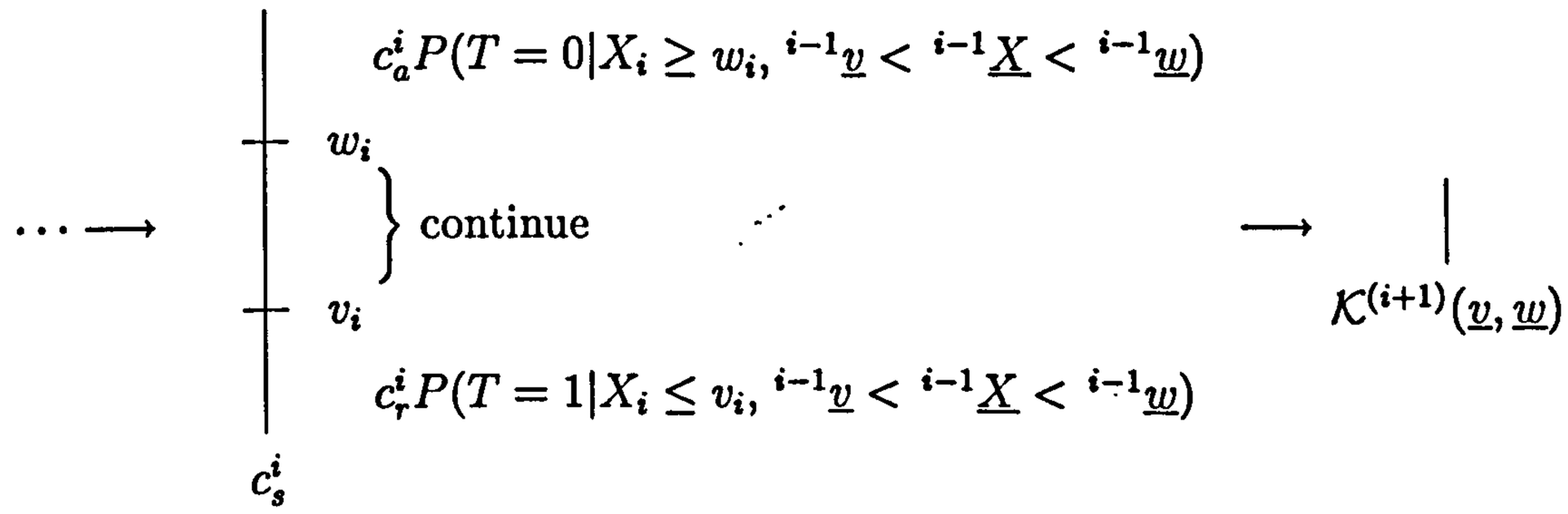
We assume a simple cost structure similar to that of the *standard case*, used

in Chapter 2. We denote as c_a^i and c_r^i, $i = 1, 2, \ldots, p$, the unit costs of wrongly accepting and wrongly rejecting an item at screen i (respectively) and we denote as c_s^i the screening cost per item at screen i. The unit cost of measuring the performance variable of an item is denoted c_m.

Write $\mathcal{K}^{(i)}(\underline{v}, \underline{w})$ for the conditional Bayes cost incurred from the implementation of the final $p - i + 1$ screens (i.e., screens i to p) for items unsentenced by the first $i - 1$. Then for $i = 1, 2, \ldots, p$

$$\begin{aligned}\mathcal{K}^{(i)}(\underline{v}, \underline{w}) = {} & c_r^i \, P(\text{good item rejected by } i\text{th screen} \mid \text{item reaches } i\text{th screen}) \\ & + c_a^i \, P(\text{bad item accepted by } i\text{th screen} \mid \text{item reaches } i\text{th screen}) \\ & + \mathcal{K}^{(i+1)}(\underline{v}, \underline{w}) P(\text{item passed to } (i+1)\text{th screen} \mid \text{item reaches } i\text{th screen}) \\ & + \text{cost of implementing } i\text{th screen},\end{aligned} \tag{4.1}$$

where the cost of implementing the ith screen is c_s^i or zero. This cost will be zero when the strategy at stage i is to reject or accept all items or pass all items on to screen $i + 1$. The following diagram illustrates the conditional Bayes costs that will be paid given each of the three possible outcomes at screen i and given that the item has reached screen i unsentenced:



The expression ${}^{i-1}\underline{v} < {}^{i-1}\underline{X} < {}^{i-1}\underline{w}$ is short-hand for the event

$$\bigcap_{j=1}^{i-1} \{v_j < X_j < w_j\},$$

namely that an item has passed the first $i-1$ screens in the design $(\underline{v}, \underline{w})$ unsentenced.

Hence the conditional Bayes cost in equation (4.1) is

$$\begin{aligned}\mathcal{K}^{(i)}(\underline{v},\underline{w}) = c_r^i\, P\left(T=1, X_i \le v_i \middle|\, {}^{i-1}\underline{v} < {}^{i-1}\underline{X} < {}^{i-1}\underline{w}\right) \\ + c_a^i\, P\left(T=0, X_i \ge w_i \middle|\, {}^{i-1}\underline{v} < {}^{i-1}\underline{X} < {}^{i-1}\underline{w}\right) \\ + \mathcal{K}^{(i+1)}(\underline{v},\underline{w}) P\left(v_i < X_i < w_i \middle|\, {}^{i-1}\underline{v} < {}^{i-1}\underline{X} < {}^{i-1}\underline{w}\right) \\ + c_s^i I(v_i, w_i)\end{aligned} \tag{4.2}$$

where $\mathcal{K}^{(p+1)} \equiv c_m$, the cost of the final stage of measuring T, and $\mathcal{K}^{(1)}(\underline{v},\underline{w}) = \mathcal{K}(\underline{v},\underline{w})$ is the Bayes cost for all p screens. The screening cost c_s^i is not paid at the ith stage when v_i and w_i both take infinite values, that is when $I(v_i, w_i) = 0$, where

$$I(v_i, w_i) = \begin{cases} 1, & \text{if either } v_i \text{ or } w_i \text{ are finite,} \\ 0, & \text{otherwise.} \end{cases} \tag{4.3}$$

Definition 4.1 $(\underline{v}^{opt}, \underline{w}^{opt})$ is a *Bayes sequential design* if

$$\mathcal{K}(\underline{v}^{opt}, \underline{w}^{opt}) = \inf_{\underline{v},\underline{w}} \mathcal{K}(\underline{v},\underline{w}), \tag{4.4}$$

where the infimum in (4.4) is over all $(\underline{v},\underline{w})$ satisfying $-\infty \le v_i \le w_i \le \infty$, $i = 1, 2, \ldots, p$.

The following result describes an optimality principle for the problem of obtaining a Bayes sequential design. It formally states that the optimal design parameters for screen i are those that minimise the conditional Bayes cost $\mathcal{K}^{(i)}$ given the optimal design parameters of the other $p-1$ $\underline{X}$—screens.

Lemma 4.1 If $(\underline{v}^{opt}, \underline{w}^{opt})$ is a Bayes sequential design such that

$$P\left(v_j^{opt} < X_j < w_j^{opt} \middle|\, {}^{j-1}\underline{v}^{opt} < {}^{j-1}\underline{X} < {}^{j-1}\underline{w}^{opt}\right) > 0, \qquad j = 1, 2, \ldots, p \tag{4.5}$$

then, for $i = 1, 2, \ldots, p$

$$\mathcal{K}^{(i)}(\underline{v}^{opt}, \underline{w}^{opt}) = \inf_{\underline{v}^{(i)},\underline{w}^{(i)}} \mathcal{K}^{(i)}\left\{\left({}^{(i-1)}\underline{v}^{opt}, \underline{v}^{(i)}\right), \left({}^{(i-1)}\underline{w}^{opt}, \underline{w}^{(i)}\right)\right\} \tag{4.6}$$

$$= \inf_{v_i,w_i} \mathcal{K}^{(i)}\left\{\left({}^{(i-1)}\underline{v}^{opt}, v_i, \underline{v}^{opt(i+1)}\right), \left({}^{(i-1)}\underline{w}^{opt}, w_i, \underline{w}^{opt(i+1)}\right)\right\} \tag{4.7}$$

Proof

Suppose condition (4.6) fails for $i = I$, say. This implies the existence of $(\underline{\bar{v}}^{(I)}, \underline{\bar{w}}^{(I)})$ such that

$$\mathcal{K}^{(I)}\left\{\left({}^{(I-1)}\underline{v}^{opt}, \underline{\bar{v}}^{(I)}\right), \left({}^{(I-1)}\underline{w}^{opt}, \underline{\bar{w}}^{(I)}\right)\right\} < \mathcal{K}^{(I)}\left(\underline{v}^{opt}, \underline{w}^{opt}\right).$$

Use of recursion (4.2) for $i = 1, 2, \ldots, I-1$, together with inequality (4.5) yields

$$\mathcal{K}^{(1)}\left\{\left({}^{(I-1)}\underline{v}^{opt}, \underline{\bar{v}}^{(I)}\right), \left({}^{(I-1)}\underline{w}^{opt}, \underline{\bar{w}}^{(I)}\right)\right\} < \mathcal{K}^{(1)}\left(\underline{v}^{opt}, \underline{w}^{opt}\right),$$

which contradicts the optimality of $(\underline{v}^{opt}, \underline{w}^{opt})$. This proves (4.6). Equation (4.7) formulates a weaker condition and follows immediately. □

Note that condition (4.5) simply guarantees that each screen makes a contribution to the overall Bayes cost $\mathcal{K}(\underline{v}^{opt}, \underline{w}^{opt})$. If (4.5) fails for the first time at $j = J$, then no items will proceed to screen $J+1$ and in an appropriate reformulation we recover (4.6) and (4.7) for $i = 1, 2, \ldots, J$. Note also that if the infimum in (4.7) is attained in the limit as $v_i \to -\infty$ and $w_i \to \infty$ then the X_i screen is omitted from the optimal design.

Relations (4.2) and (4.7) do not yield a dynamic programming approach to the determination of Bayes sequential designs based on conventional backwards induction because the cost $\mathcal{K}^{(i)}$ depends upon the entire design $(\underline{v}, \underline{w})$ and not only upon the design of the final $p - i + 1$ screens. However, these relations are strongly suggestive of a heuristic approach which is iterative. In this iterative scheme, the current design $(\underline{v}', \underline{w}')$ is updated to the new design $(\underline{v}'', \underline{w}'')$ by using a backwards induction as follows: perform a backward sweep through the problem, recalculating the designs in the order $p, p-1, \ldots, 1$. Suppose that the updates (v_j'', w_j''), $j = i+1, i+2, \ldots, p$ have been obtained. We consider the problem of calculating (v_i'', w_i''). The r.h.s. of (4.7) suggests that we need to consider the minimisation of a suitable approximation of

$$\mathcal{K}^{(i)}\left\{\left({}^{(i-1)}\underline{v}', v_i, \underline{v}''^{(i+1)}\right), \left({}^{(i-1)}\underline{w}', w_i, \underline{w}''^{(i+1)}\right)\right\}$$

with respect to (v_i, w_i). The approximating cost $\hat{\mathcal{K}}^{(i)}$ is derived from $\mathcal{K}^{(i)}$ by applying a delta-type approximation to the conditional distribution of $X_i | {}^{i-1}\underline{v} < {}^{i-1}\underline{X} <$

$^{i-1}\underline{w}$, that is, we approximate the distribution of $\underline{X}_{(i)}|\,^{i-1}\underline{v} < {}^{i-1}\underline{X} < {}^{i-1}\underline{w}$ by its expected value. A further approximation we use is to evaluate the cost $\mathcal{K}^{(i+1)}$ at $\left({}^{(i)}\underline{v}', \underline{v}''^{(i+1)}\right), \left({}^{(i)}\underline{w}', \underline{w}''^{(i+1)}\right)$ and so, in the updating procedure for screen i, the cost $\mathcal{K}^{(i+1)}$ is not a function of v_i and w_i. Following (4.2) we have

$$\begin{aligned}
&\hat{\mathcal{K}}^{(i)}\left\{\left({}^{(i-1)}\underline{v}', v_i, \underline{v}''^{(i+1)}\right), \left({}^{(i-1)}\underline{w}', w_i, \underline{w}''^{(i+1)}\right)\right\} \\
&= c_r^i\, P\left(T=1, X_i \le v_i | \underline{X}_{(i)} = \underline{\mu}_{(i)}\right) \\
&\quad + c_a^i\, P\left(T=0, X_i \ge w_i | \underline{X}_{(i)} = \underline{\mu}_{(i)}\right) \\
&\quad + c_s^i\, I(v_i, w_i) \\
&\quad + P\left(v_i < X_i < w_i | \underline{X}_{(i)} = \underline{\mu}_{(i)}\right) \\
&\qquad \times \mathcal{K}^{(i+1)}\left\{\left({}^{(i)}\underline{v}', \underline{v}''^{(i+1)}\right), \left({}^{(i)}\underline{w}', \underline{w}''^{(i+1)}\right)\right\},
\end{aligned} \tag{4.8}$$

where

$$\underline{\mu}_{(i)} = E\left[\underline{X}_{(i)}|\,^{i-1}\underline{v}' < {}^{i-1}\underline{X} < {}^{i-1}\underline{w}'\right]$$

and (v_i'', w_i'') are chosen to satisfy the relation

$$\begin{aligned}
&\hat{\mathcal{K}}^{(i)}\left\{\left({}^{(i-1)}\underline{v}', \underline{v}''^{(i)}\right), \left({}^{(i-1)}\underline{w}', \underline{w}''^{(i)}\right)\right\} \\
&= \inf_{(v_i, w_i)} \hat{\mathcal{K}}^{(i)}\left\{\left({}^{(i-1)}\underline{v}', v_i, \underline{v}''^{(i+1)}\right), \left({}^{(i-1)}\underline{w}', w_i, \underline{w}''^{(i+1)}\right)\right\}
\end{aligned} \tag{4.9}$$

Notice that, taking $c_m = \mathcal{K}^{(i+1)}$, the Bayes cost in (4.8) is of the same form as the Bayes cost given in Chapter 2, equation (2.1) and the relation (4.9) is of a similar nature to that given in Definition 2.1. Hence, one effect of this approximative dynamic programming approach is that when $T|\underline{X}$ is modelled by a probit regression then each optimisation (4.9) has the form of the simple two-stage problems discussed in Chapter 2. The designs obtained by this route thus inherit all the advantages of simplicity and interpretability of the solutions discussed there. The next section describes the updating procedure more thoroughly.

A full backward sweep produces an updated sequential design $(\underline{v}'', \underline{w}'')$. We continue updating until some convergence criterion defined by Bayes cost $\mathcal{K}(\underline{v}, \underline{w})$ is satisfied. If $(\hat{\underline{v}}, \hat{\underline{w}})$ is the limit design (assuming it exists and is unique), we have from

(4.9) that, for $i = 1, 2, \ldots, p$

$$\hat{\mathcal{K}}^{(i)}(\underline{\hat{v}}, \underline{\hat{w}}) = \inf_{(v_i, w_i)} \hat{\mathcal{K}}^{(i)} \left\{ \left({}^{(i-1)}\underline{\hat{v}}, v_i, \underline{\hat{v}}^{(i+1)} \right), \left({}^{(i-1)}\underline{\hat{w}}, w_i, \underline{\hat{w}}^{(i+1)} \right) \right\} \tag{4.10}$$

which has the form of the optimality equation (4.7) but for the approximate Bayes cost $\hat{\mathcal{K}}^{(i)}$.

This completes an outline of our heuristic approach to designing a sequential screen. Section 4.3 details the case $p = 2$ and in Section 4.4 we present a computational study that compares the heuristic designs based upon this approximative approach with designs that are Bayes optimal.

4.3 A heuristic three-stage screen

We now give in full detail, the solution procedure for the simplest, interesting case, namely $p = 2$. We develop the model for this case as follows. Screening variables X_1 and X_2 have joint probability density function $\psi(x_1, x_2)$. We assume a model for $T|\underline{X}$ expressed via a probit link function, that is,

$$P(T = 1|\underline{x}, \underline{\xi}) = \Phi(\underline{\xi}^T \underline{x}_0)$$

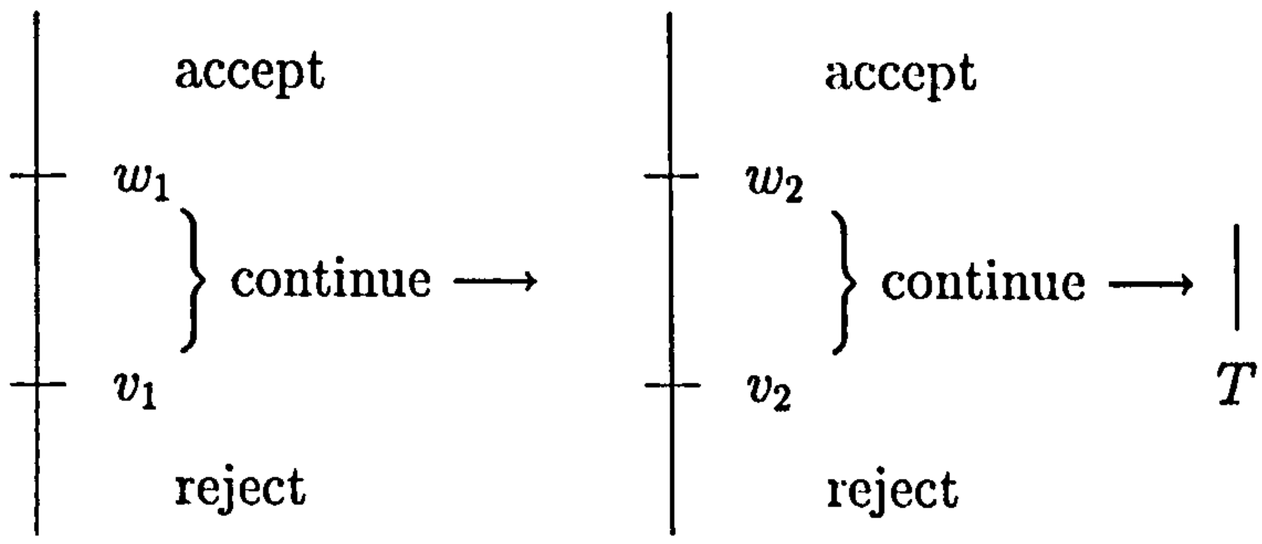
where $\underline{\xi} = (\xi_0, \xi_1, \xi_2)^T$ and $\underline{x}_0 = (1, x_1, x_2)^T$. We write the first and second moments of $\underline{\xi}$ as

$$E(\xi_i) = m_i, \quad var(\xi_i) = s_i^2, \quad corr(\xi_i, \xi_j) = r_{ij}, \quad i, j = 0, 1, 2 \quad i \neq j.$$

Also, we write the moments of the conditional distributions of $X_1|X_2 = x_2$ and $X_2|X_1 = x_1$ as

$$\begin{aligned} E(X_1|X_2 = x_2) &= \mu_{1|2}(x_2), \qquad var(X_1|X_2 = x_2) = \sigma^2_{1|2}(x_2), \\ E(X_2|X_1 = x_1) &= \mu_{2|1}(x_1), \qquad var(X_2|X_1 = x_1) = \sigma^2_{2|1}(x_1). \end{aligned} \tag{4.11}$$

Under the assumption that X_1 and X_2 are larger-the-better screening variables the natural form of the design is:



That is we accept an item only if

(i) $X_1 \geq w_1$,

(ii) $v_1 < X_1 < w_1$ and $X_2 \geq w_2$, or

(iii) $v_1 < X_1 < w_1$, $v_2 < X_2 < w_2$ and $T = 1$.

The Bayes cost of the above set–up is

$$\begin{aligned} \mathcal{K}(\underline{v}, \underline{w}) = \mathcal{K}^{(1)}(\underline{v}, \underline{w}) &= c_r^1 P(T = 1, X_1 \leq v_1) + c_a^1 P(T = 0, X_1 \geq w_1) \\ &\quad + \mathcal{K}^{(2)}(\underline{v}, \underline{w}) P(v_1 < X_1 < w_1) + c_s^1 I(v_1, w_1), \end{aligned} \tag{4.12}$$

where

$$\begin{aligned} \mathcal{K}^{(2)}(\underline{v}, \underline{w}) &= c_r^2 P(T = 1, X_2 \leq v_2 | v_1 < X_1 < w_1) \\ &\quad + c_a^2 P(T = 0, X_2 \geq w_2 | v_1 < X_1 < w_1) \\ &\quad + c_m P(v_2 < X_2 < w_2 | v_1 < X_1 < w_1) + c_s^2 I(v_2, w_2), \end{aligned} \tag{4.13}$$

is the conditional Bayes cost incurred from the implementation of the second screen for items unsentenced by the first. The indicator function $I(v_i, w_i)$ is defined in equation (4.3). Here $I(v_1, w_1)$ is zero when both v_1 and w_1 take infinite values, that is, when the first screening variable is not used in the design and hence screening cost c_s^1 is not paid. Otherwise $I(v_1, w_1) = 1$. Similarly, $I(v_2, w_2) = 1$ if the second screen

is implemented and zero otherwise. Here we use the heuristic method outlined in section 4.2 to find sub-optimal values of the characterising parameters (v_1, w_1, v_2, w_2).

In the iterative scheme of section 4.2 we denote the current solution by $(\underline{v}', \underline{w}') = (v_1', v_2', w_1', w_2')$ and compute updated values $(\underline{v}'', \underline{w}'')$. Of the two stages that will form an iteration, the first stage is to find updates (v_2'', w_2'') of the parameters characterising the second screen. Hence we consider $\mathcal{K}^{(2)}(v_1', v_2, w_1', w_2)$ above and write

$$\mu_1' = E(X_1 | v_1' < X_1 < w_1'). \tag{4.14}$$

Note that this conditional expectation is in fact $\mu_{(2)}$ defined in (4.8). We then have

$$\begin{aligned} \hat{\mathcal{K}}^{(2)}(v_1', v_2, w_1', w_2) = c_r^2 \, & P(T = 1, X_2 \leq v_2 | X_1 = \mu_1') \\ & + c_a^2 \, P(T = 0, X_2 \geq w_2 | X_1 = \mu_1') \\ & + c_m P(v_2 < X_2 < w_2 | X_1 = \mu_1') + c_s^2 \, I(v_2, w_2) \end{aligned} \tag{4.15}$$

as our approximation to Bayes cost $\mathcal{K}^{(2)}(v_1', v_2, w_1', w_2)$. Following (4.9), the updated values (v_2'', w_2'') are solutions to the minimisation problem

$$\hat{\mathcal{K}}^{(2)}(v_1', v_2'', w_1', w_2'') = \inf_{(v_2, w_2)} \hat{\mathcal{K}}^{(2)}(v_1', v_2, w_1', w_2). \tag{4.16}$$

However (4.16) can be recast as a single screen ($p = 1$) problem of the kind discussed in Chapter 2 as follows:

(a) the key cost parameters are c_a^2, c_r^2, c_s^2 and c_m;

(b) the probit regression of T on X_2 given that $X_1 = \mu_1'$ has parameters $\underline{\xi}^* = (\xi_0^*, \xi_1^*)^T$, where $\xi_0^* = \xi_0 + \mu_1' \xi_1$ and $\xi_1^* = \xi_2$.

Part (b) follows because, under the probit regression model,

$$\begin{aligned} P(T = 1 | \underline{\xi}, X_1 = \mu_1', x_2) &= \Phi(\xi_0 + \xi_1 \mu_1' + \xi_2 x_2) \\ &= \Phi[\{\xi_0 + \xi_1 \mu_1'\} + \xi_2 x_2] \\ &= \Phi(\xi_0^* + \xi_1^* x_2) \\ &= P(T = 1 | \underline{\xi}^*, x_2). \end{aligned}$$

The mean of the regression parameters $\underline{\xi}^*$ is

$$E(\underline{\xi}^*) \equiv \begin{pmatrix} m_0^* \\ m_1^* \end{pmatrix} = \begin{pmatrix} E(\xi_0) + \mu_1' E(\xi_1) \\ E(\xi_2) \end{pmatrix} = \begin{pmatrix} m_0 + \mu_1' m_1 \\ m_2 \end{pmatrix}$$

and the covariance structure of $\underline{\xi}^*$ is

$$\begin{aligned} Var(\xi_0^*) \equiv s_0^{*^2} &= Var(\xi_0) + \mu_1'^2 Var(\xi_1) + 2\mu_1' Cov(\xi_0, \xi_1), \\ &= s_0^2 + \mu_1'^2 s_1^2 + 2\mu_1' r_{01} s_0 s_1, \\ Var(\xi_1^*) \equiv s_1^{*^2} &= Var(\xi_2), \\ &= s_2^2, \\ Cov(\xi_0^*, \xi_1^*) \equiv r^* s_0^* s_1^* &= Cov(\xi_0, \xi_2) + \mu_1' Cov(\xi_1, \xi_2) \\ &= r_{02} s_0 s_2 + \mu_1' r_{12} s_1 s_2. \end{aligned}$$

Since in general we do not have $m_0^* = 0$ and $s_1^{*^2} = 1$ then a change of variable to $(\xi_0^\dagger, \xi_1^\dagger)$, where $\xi_0^\dagger = \xi_0^* - m_0^* \xi_1^* / m_1^*$ and $\xi_1^\dagger = \xi_1^* / s_2^*$, is needed before application of Theorem 2.1. Following section 2.3.8 this results in the following mean and covariance structure for the new variables:

$$\begin{aligned} &E(\xi_0^\dagger) \equiv m_0^\dagger = 0, && E(\xi_1^\dagger) \equiv m_1^\dagger = m_1^* / s_1^*, \\ &Var(\xi_0^\dagger) \equiv s_0^{\dagger^2} = s_0^{*^2} + m_0^{*^2} s_1^{*^2} / m_1^{*^2} - 2 m_0^* r^* s_0^* s_1^* / m_1^*, && Var(\xi_1^\dagger) \equiv s_1^{\dagger^2} = 1, \\ &Cov(\xi_0^\dagger, \xi_1^\dagger) \equiv r^\dagger s_0^\dagger s_1^\dagger = r^* s_0^* - m_0^* s_1^* / m_1^*. \end{aligned}$$

With this change of variable, design parameters for the second screen are now derived from Theorem 2.1 and denoted $(v_2^\dagger, w_2^\dagger)$.

Having obtained $(v_2^\dagger, w_2^\dagger)$, reversing the change of variable (see equations (2.10)) gives the updated screen for $X_2 | X_1 = \mu_1'$, that is, (v_2'', w_2'') in (4.16) as:

$$\begin{aligned} v_2'' &= v_2^\dagger / s_1^* - m_0^* / m_1^* = v_2^\dagger / s_2 - (m_0 + m_1 \mu_1') / m_2 \\ w_2'' &= w_2^\dagger / s_1^* - m_0^* / m_1^* = w_2^\dagger / s_2 - (m_0 + m_1 \mu_1') / m_2. \end{aligned} \tag{4.17}$$

This completes the update of the second screen parameters and we now use (v_2'', w_2'') to derive new iterates (v_1'', w_1'') for the design of the X_1-screen.

We write

$$\mu_2' = E(X_2) \tag{4.18}$$

where μ_2' is $\mu_{(1)}$ defined in (4.8). We now have

$$\begin{aligned}\hat{\mathcal{K}}^{(1)}(v_1, v_2'', w_1, w_2'') &= c_r^1 P(T=1, X_1 \leq v_1 | X_2 = \mu_2') \\ &\quad + c_a^1 P(T=0, X_1 \geq w_1 | X_2 = \mu_2') \\ &\quad + c_s^1 I(v_1, w_1) \\ &\quad + P(v_1 < X_1 < w_1 | X_2 = \mu_2') \, \mathcal{K}^{(2)}(v_1', v_2'', w_1', w_2'')\end{aligned} \tag{4.19}$$

as our approximation to the Bayes cost $\mathcal{K}^{(1)}(v_1, v_2'', w_1, w_2'')$. Following (4.9) again, we choose (v_1'', w_1'') to satisfy

$$\hat{\mathcal{K}}^{(1)}(v_1'', v_2'', w_1'', w_2'') = \inf_{(v_1, w_1)} \hat{\mathcal{K}}^{(1)}(v_1, v_2'', w_1, w_2''). \tag{4.20}$$

The minimisation (4.20) may be recast as a single screen ($p = 1$) problem of the kind discussed in Chapter 2 as follows:

(a) the key cost parameters are c_a^1, c_r^1, c_s^1 and $\mathcal{K}^{(2)}(v_1', v_2'', w_1', w_2'')$;

(b) the probit regression of T on X_1 given that $X_2 = \mu_2'$ has parameters $\underset{\sim}{\tilde{\xi}} = (\tilde{\xi}_0, \tilde{\xi}_1)^T$, where $\tilde{\xi}_0 = \xi_0 + \mu_2' \xi_2$ and $\tilde{\xi}_1 = \xi_1$.

Part (b) follows because, under the probit regression model,

$$\begin{aligned}P(T=1 | \underset{\sim}{\xi}, x_1, X_2 = \mu_2') &= \Phi(\xi_0 + \xi_1 x_1 + \xi_2 \mu_2') \\ &= \Phi[\{\xi_0 + \xi_2 \mu_2'\} + \xi_1 x_1] \\ &= \Phi\left(\tilde{\xi}_0 + \tilde{\xi}_1 x_1\right) \\ &= P(T=1 | \underset{\sim}{\tilde{\xi}}, x_1).\end{aligned}$$

The mean of the regression parameters $\underset{\sim}{\xi}$ is

$$\begin{aligned}E(\underset{\sim}{\tilde{\xi}}) \equiv \begin{pmatrix} \tilde{m}_0 \\ \tilde{m}_1 \end{pmatrix} &= \begin{pmatrix} E(\xi_0) + \mu_2' E(\xi_2) \\ E(\xi_1) \end{pmatrix} \\ &= \begin{pmatrix} m_0 + \mu_2' m_2 \\ m_1 \end{pmatrix}\end{aligned}$$

and the covariance structure of $\underline{\tilde{\xi}}$ is

$$\begin{aligned}
Var(\tilde{\xi}_0) \equiv \tilde{s}_0^2 &= Var(\xi_0) + \mu_2'^2 Var(\xi_2) + 2\mu_2' Cov(\xi_0, \xi_2), \\
&= s_0^2 + \mu_2'^2 s_2^2 + 2\mu_2' r_{02} s_0 s_2, \\
Var(\tilde{\xi}_1) \equiv \tilde{s}_1^2 &= Var(\xi_1), \\
&= s_1^2, \\
Cov(\tilde{\xi}_0, \tilde{\xi}_1) \equiv \tilde{r}\tilde{s}_0\tilde{s}_1 &= Cov(\xi_0, \xi_1) + \mu_2' Cov(\xi_1, \xi_2) \\
&= r_{01} s_0 s_1 + \mu_2' r_{12} s_1 s_2.
\end{aligned}$$

Since we do not in general have $\tilde{m}_0 = 0$ and $\tilde{s}_1^2 = 1$ then a change of variable to $(\xi_0^\dagger, \xi_1^\dagger) = (\tilde{\xi}_0 - \tilde{m}_0\tilde{\xi}_1/\tilde{m}_1,\ \tilde{\xi}_1/\tilde{s}_1)$ is needed before application of Theorem 2.1. For completeness we give the mean and covariance structure of $(\xi_0^\dagger, \xi_1^\dagger)$ which follows from section 2.3.8:

$$\begin{aligned}
&E(\xi_0^\dagger) \equiv m_0^\dagger = 0, && E(\xi_1^\dagger) \equiv m_1^\dagger = \tilde{m}_1/\tilde{s}_1, \\
&Var(\xi_0^\dagger) \equiv s_0^{\dagger 2} = \tilde{s}_0^2 + \tilde{m}_0^2\tilde{s}_1^2/\tilde{m}_1^2 - 2\tilde{m}_0\tilde{r}\tilde{s}_0\tilde{s}_1/\tilde{m}_1, && Var(\xi_1^\dagger) \equiv s_1^{\dagger 2} = 1, \\
&Cov(\xi_0^\dagger, \xi_1^\dagger) \equiv r^\dagger s_0^\dagger s_1^\dagger = \tilde{r}\tilde{s}_0 - \tilde{m}_0\tilde{s}_1/\tilde{m}_1.
\end{aligned}$$

With this change of variable, the optimum design for the first screen is now derived from Theorem 2.1 and denoted $(v_1^\dagger, w_1^\dagger)$.

Reversal of the transformation gives the updates of the first screen as cut-off points for the variable $X_1|X_2 = 0$, that is, (v_1'', w_1'') in (4.20) are given by

$$\begin{aligned}
v_1'' &= v_1^\dagger/\tilde{s}_1 - \tilde{m}_0/\tilde{m}_1 = v_1^\dagger/s_1 - (m_0 + m_2\mu_2')/m_1 \\
w_1'' &= w_1^\dagger/\tilde{s}_1 - \tilde{m}_0/\tilde{m}_1 = w_2^\dagger/s_1 - (m_0 + m_2\mu_2')/m_1.
\end{aligned} \tag{4.21}$$

This now completes a single iteration for the heuristic solution. We continue to iterate until an appropriate convergence criterion defined with respect to Bayes cost $\mathcal{K}(v, w)$ is satisfied. This concludes discussion of the $p = 2$ case.

Comments

1. Not surprisingly in such a complicated set-up, we have not been able to prove convergence of the above scheme to a unique limit theoretically. Considerable

computational experience (which is reported in the next section) suggests that convergence is rapid.

2. The design $(\hat{v}_1, \hat{v}_2, \hat{w}_1, \hat{w}_2)$ obtained at convergence in the $p = 2$ case above can be characterised via equations (4.17) and (4.21) since iterating from the limiting design will leave it unchanged.

3. Simplifications to the scheme arise when $\underline{X}$ is assumed to be multivariate normal because of the special forms of the conditional means and variances used in the computations, see the next section.

4. Since this heuristic solution is based on approximations using robust two-stage screens, it is plain that the method will inherit properties of robustness in performance to departures from model assumptions.

5. Note that initial values of v_1 and w_1 are needed to start the iterative procedure. Computational evidence (given later) suggests that the choice of starting values has little effect on the design obtained at convergence but has some small effect on the speed of convergence.

6. In Chapter 2 we make a normality assumption about the regression parameters in the probit model. In our solution to the sequential problem we make the same assumption about the parameters ξ_0, ξ_1 and ξ_2 of the probit regression model on $T|X_1, X_2$. Note that throughout section 4.3 all changes of regression variable involve a linear combination of normal variables and hence the new variables will also be normal.

7. Note that knowledge of the distribution of (X_1, X_2) is needed to compute our heuristic design. This information is needed to calculate: (a) $\mathcal{K}^{(2)}$, the conditional Bayes cost of the second screen only, which is required to calculate the design for screen one, and (b) $\mathcal{K} = \mathcal{K}^{(1)}$, the full Bayes cost of the design, which is required to assess whether the cost of the design has converged.

4.4 Performance of the heuristic solution

Here we present a simulation study that compares the performance of the heuristic design developed in section 4.3, with that of a fully optimal 2–screen sequential design. We vary the possible inputs to the design and look at the effect on the performance of the heuristic.

Section 4.4.1 describes the inputs required to find both the fully optimal and heuristic designs of the sequential screen. This section also gives the range of values chosen for the inputs to the simulation study. These were chosen to make the study manageable while still considering many possible scenarios. The algorithm for computing the heuristic screen is given in section 4.4.2, together with a note on computing the fully optimal design. Finally, section 4.4.3 gives the results of the simulation study.

4.4.1 Inputs

Here we list the information needed to compute the heuristic design and the fully optimal Bayes sequential design for the 2–screen case. To make the subsequent simulation study more tractable we reduce the range of possible values for inputs by making some assumptions.

(i) **Initial values.** We will look at the effect of a selection of starting designs for the first screen, denoted (v_1^0, w_1^0) say, for the iterative process that is used to compute the heuristic design. It seems sensible to consider values of v_1^0 and w_1^0 that correspond to reasonably likely values of the variable X_1. We will assume that X_1 is standard normal and so we allow v_1^0 and w_1^0 to take values in the range $(-3, 3)$.

(ii) **Cost parameters.** Recall that only the relative size of the cost parameters is important and so we set the cost of measuring the performance variable $c_m = 1$ and vary c_r^1, c_r^2, c_a^1 and c_a^2. We further assume that the screening costs c_s^1 and

c_s^2 will be negligible and so set $c_s^1 = c_s^2 = 0$. As discussed elsewhere, this will often be the case in practice.

Recall that in Chapter 2, equation (2.9) gives a requirement in terms of cost parameters for a non-trivial solution of the one–screen problem to exist. In the heuristic design of our two–screen procedure in which $c_m = 1$ this requirement becomes

$$\frac{1}{c_a^2} + \frac{1}{c_r^2} \leq 1, \tag{4.22}$$

when designing the second screen and

$$\frac{1}{c_a^1} + \frac{1}{c_r^1} \leq \frac{1}{\mathcal{K}^{(2)}} \tag{4.23}$$

when designing the first screen. Plainly $\mathcal{K}^{(2)}$ in (4.23) will be unknown prior to screen design but will be at most 1 (when it is not optimal to use the second screen and $\mathcal{K}^{(2)} = c_m$). Hence to ensure that both (4.22) and (4.23) are satisfied we put a lower bound of 2.5 on misclassification costs. Also in Chapter 2, the expressions (2.6) and (2.7) refer to cost functions k_1 and k_2 which take values between 0 and 1, so that $c_1 = \Phi^{-1}(k_1)$ and $c_2 = \Phi^{-1}(k_2)$. To avoid taking values of c_i in the tail we only consider values of $k_i \in (0.0667, 0.9333)$, that is $c_i \in (-1.501, 1.501)$. From the definition of k_1 and k_2 this corresponds to setting an upper bound of 15 on misclassification costs. Hence we consider values of $(c_r^1, c_a^1, c_r^2, c_a^2)$ between 2.5 and 15. (In initial trials a bound of 20 on misclassification costs resulted in too many scenarios in which it was not optimal to base the screen on both covariates.)

(iii) **Distribution of $\underline{X}$.** We assume throughout the study that (X_1, X_2) follows a bivariate standard normal distribution with correlation coefficient r_x. Under such a normality assumption, the conditional moments in (4.11) are $\mu_{1|2}(x) = \mu_{2|1}(x) = r_x x$ and $\sigma_{1|2}^2(x) = \sigma_{2|1}^2(x) = 1 - r_x^2$. Also μ_1' in (4.14) can be computed by noting that:

$$\mu_1' = E(X_1 | v_1' < X_1 < w_1') = \frac{\int_{v_1'}^{w_1'} x_1 \phi(x_1) dx_1}{P(v_1' < X_1 < w_1')} = \frac{\phi(v_1') - \phi(w_1')}{\Phi(w_1') - \Phi(v_1')}, \tag{4.24}$$

where $\phi(\cdot)$ is the standard normal density function and $\Phi(\cdot)$ is the standard normal cumulative distribution function.

As we assume that both X_1 large and X_2 large imply that $T = 1$ it seems natural to only consider cases in which X_1 and X_2 are positively correlated, that is $r_x \geq 0$. We also only allow the correlation coefficient between screening variables to be moderately high, that is we allow $r_x \in [0, 0.75]$. This is to avoid cases in which the measurement of one covariate almost determines the second covariate.

(iv) **Probit model parameters.** Information about the probit regression parameters $\underline{\xi} = (\xi_0, \xi_1, \xi_2)^T$ is given in terms of the first and second moments of $\underline{\xi}$ denoted by

$$\underline{m} = \begin{pmatrix} m_0 \\ m_1 \\ m_2 \end{pmatrix} \quad \text{and} \quad \boldsymbol{S} = \begin{pmatrix} s_0^2 & r_{01}s_0s_1 & r_{02}s_0s_2 \\ r_{01}s_0s_1 & s_1^2 & r_{12}s_1s_2 \\ r_{02}s_0s_2 & r_{12}s_1s_2 & s_2^2 \end{pmatrix}.$$

Assuming a large sample on $(T, X_1, X_2)^T$, the posterior distribution of $\underline{\xi}$ is $N_3(\underline{m}, \boldsymbol{S})$, where $\underline{m}$ is the maximum likelihood estimate of $\underline{\xi}$ and $\boldsymbol{S}$ is the inverse of Fisher's information matrix evaluated at $\underline{m}$. In practice $\underline{m}$ and $\boldsymbol{S}$ can be calculated using a method similar to that given in section 2.3.7.

In the simulation study we allow $\underline{m}$ to vary but calculate the covariance structure of $\underline{\xi}$ as the inverse of the expected value of Fisher's information matrix assuming a linear probit regression with parameters $(m_0, m_1, m_2)^T$, the expectation being taken with respect to the distribution of $(X_1, X_2)^T$. Thus the covariance structure is typical of those obtained when $\underline{\xi}$ has a posterior distribution based on a reasonably large sample of data. Following (2.13), under a probit regression model the (j,k)th element of Fisher's information matrix from one observation $(t, x_1, x_2)^T$ on $(T, X_1, X_2)^T$ is, $j, k = 0, 1, 2$:

$$i_{jk}(\underline{\xi}, x_1, x_2) = \frac{x_j x_k \phi(\xi_0 + \xi_1 x_1 + \xi_2 x_2)^2}{\Phi(\xi_0 + \xi_1 x_1 + \xi_2 x_2)\{1 - \Phi(\xi_0 + \xi_1 x_1 + \xi_2 x_2)\}}, \tag{4.25}$$

where $x_0 = 1$. Hence, assuming a standard bivariate normal distribution for $(X_1, X_2)^T$ and that the probit regression parameters are $(m_0, m_1, m_2)^T$, the expected value of Fisher's information matrix for one observation on $(T, X_1, X_2)^T$ has elements

$$E_{X_1,X_2}\left\{\imath_{jk}(\underline{m}, X_1, X_2)\right\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \imath_{jk}(\underline{m}, x_1, x_2)\phi_2(x_1, x_2|r_x)\, dx_1\, dx_2, \tag{4.26}$$

where $j, k = 0, 1, 2$ and $\phi_2(x_1, x_2|r_x)$ is the bivariate standard normal density function with correlation r_x. Therefore, for a sample of n observations, the expected value of Fisher's information matrix, $E_{X_1,X_2}\left\{I(\underline{m}, X_1, X_2)\right\}$, has (j,k)th element

$$\begin{aligned} E_{X_1,X_2}\left\{I_{jk}(\underline{m}, X_1, X_2)\right\} &= \sum_{i=1}^{n} E_{X_1,X_2}\left\{\imath_{jk}(\underline{m}, X_1, X_2)\right\} \\ &= nE_{X_1,X_2}\left\{\imath_{jk}(\underline{m}, X_1, X_2)\right\}. \end{aligned}$$

In oui study we suppose that we base our screen on a sample of size $n = 30$.

Under a linear probit regression model

$$P(T=1) = E_{\underline{X},\underline{\xi}}\left[P(T=1|\underline{X}, \underline{\xi})\right] = E_{\underline{X},\underline{\xi}}\left[\Phi(\xi_0 + \xi_1 X_1 + \xi_2 X_2)\right].$$

Making a delta-type approximation gives

$$P(T=1) \simeq P\left[T=1|\underline{X} = E(\underline{X}), E(\underline{\xi})\right] = \Phi(m_0) \tag{4.27}$$

since $E(\underline{X}) = \underline{0}$ and $E(\underline{\xi}) = \underline{m}$. Here we only consider cases in which the value of m_0 gives only weak or moderate inferences about T and restrict values of m_0 to the range $(-1, 1)$. In the approximation (4.27), this corresponds to the range $(0.1587, 0.8413)$ for $P(T=1)$. We assume that X_1 large and X_2 large imply that $T = 1$, in the probit model above we then have $\xi_1 > 0$ and $\xi_2 > 0$ and hence we only consider cases in which m_1 and m_2 are positive.

4.4.2 Algorithms

Given the inputs to the design described in section 4.4.1 above, we now outline the method for computing both the parameters that characterise our heuristic design and

those of the fully optimal design. We also discuss computing Bayes costs with which we can compare the two designs.

Covariance structure

For each set of inputs, the first computational problem to be solved is that of calculating the covariance structure as in section 4.4.1 (iv) above. The information matrix is symmetric so that the (j,k)th element and the (k,j)th element are equal. Therefore to calculate the expectation of Fisher's information matrix we need to perform 6 integrals of the form of equation (4.26). Simple quadrature routines for performing integration over unbounded intervals transform a problem to integration over some bounded interval, such as $(0,1)$, by making a suitable change of variables. Experimentation has shown that such routines typically do badly for numerical computation of these six two dimensional integrals. However, as we shall see, a method called Gauss–Hermite quadrature will work well for this problem.

For integration over $\mathbb{R}^p$, a Gauss–Hermite product rule that evaluates the function at n_i different values of x_i, will integrate exactly functions of the form

$$p_1(\underline{x}) \exp\left\{-\frac{\underline{x}^T\underline{x}}{2}\right\}, \tag{4.28}$$

where the function $p_1(\underline{x})$ is a polynomial of order $2n_i - 1$ in x_i, $i = 1, 2, \ldots, p$. Hence, for integrating a function that is loosely approximated by a $N_p(\underline{0}, \boldsymbol{I})$ density ($\boldsymbol{I}$ is the identity matrix) multiplied by a polynomial, such rules will provide a reasonably accurate answer provided the function is evaluated at a suitable number of design points. Now suppose that we have a function that is loosely approximated by

$$p_2(\underline{x}) \times \phi_p\left(\underline{x}|\underline{\mu}, \boldsymbol{\Sigma}\right), \tag{4.29}$$

where $p_2(\underline{x})$ is a polynomial in each of the components of $\underline{x}$ and $\phi_p\left(\underline{x}|\underline{\mu}, \boldsymbol{\Sigma}\right)$ is the multivariate Normal density function with mean $\underline{\mu}$ and variance matrix $\boldsymbol{\Sigma}$. If we make a change of variables to

$$\underline{z} = \boldsymbol{B}^{-1}\left(\underline{x} - \underline{\mu}\right)$$

those of the fully optimal design. We also discuss computing Bayes costs with which we can compare the two designs.

Covariance structure

For each set of inputs, the first computational problem to be solved is that of calculating the covariance structure as in section 4.4.1 (iv) above. The information matrix is symmetric so that the (j,k)th element and the (k,j)th element are equal. Therefore to calculate the expectation of Fisher's information matrix we need to perform 6 integrals of the form of equation (4.26). Simple quadrature routines for performing integration over unbounded intervals transform a problem to integration over some bounded interval, such as $(0,1)$, by making a suitable change of variables. Experimentation has shown that such routines typically do badly for numerical computation of these six two dimensional integrals. However, as we shall see, a method called Gauss–Hermite quadrature will work well for this problem.

For integration over $\mathbb{R}^p$, a Gauss–Hermite product rule that evaluates the function at n_i different values of x_i, will integrate exactly functions of the form

$$p_1(\underline{x}) \exp\left\{-\frac{\underline{x}^T\underline{x}}{2}\right\}, \tag{4.28}$$

where the function $p_1(\underline{x})$ is a polynomial of order $2n_i - 1$ in x_i, $i = 1, 2, \ldots, p$. Hence, for integrating a function that is loosely approximated by a $N_p(\underline{0}, \boldsymbol{I})$ density ($\boldsymbol{I}$ is the identity matrix) multiplied by a polynomial, such rules will provide a reasonably accurate answer provided the function is evaluated at a suitable number of design points. Now suppose that we have a function that is loosely approximated by

$$p_2(\underline{x}) \times \phi_p\left(\underline{x}|\underline{\mu}, \boldsymbol{\Sigma}\right), \tag{4.29}$$

where $p_2(\underline{x})$ is a polynomial in each of the components of $\underline{x}$ and $\phi_p\left(\underline{x}|\underline{\mu}, \boldsymbol{\Sigma}\right)$ is the multivariate Normal density function with mean $\underline{\mu}$ and variance matrix $\boldsymbol{\Sigma}$. If we make a change of variables to

$$\underline{z} = \boldsymbol{B}^{-1}\left(\underline{x} - \underline{\mu}\right)$$

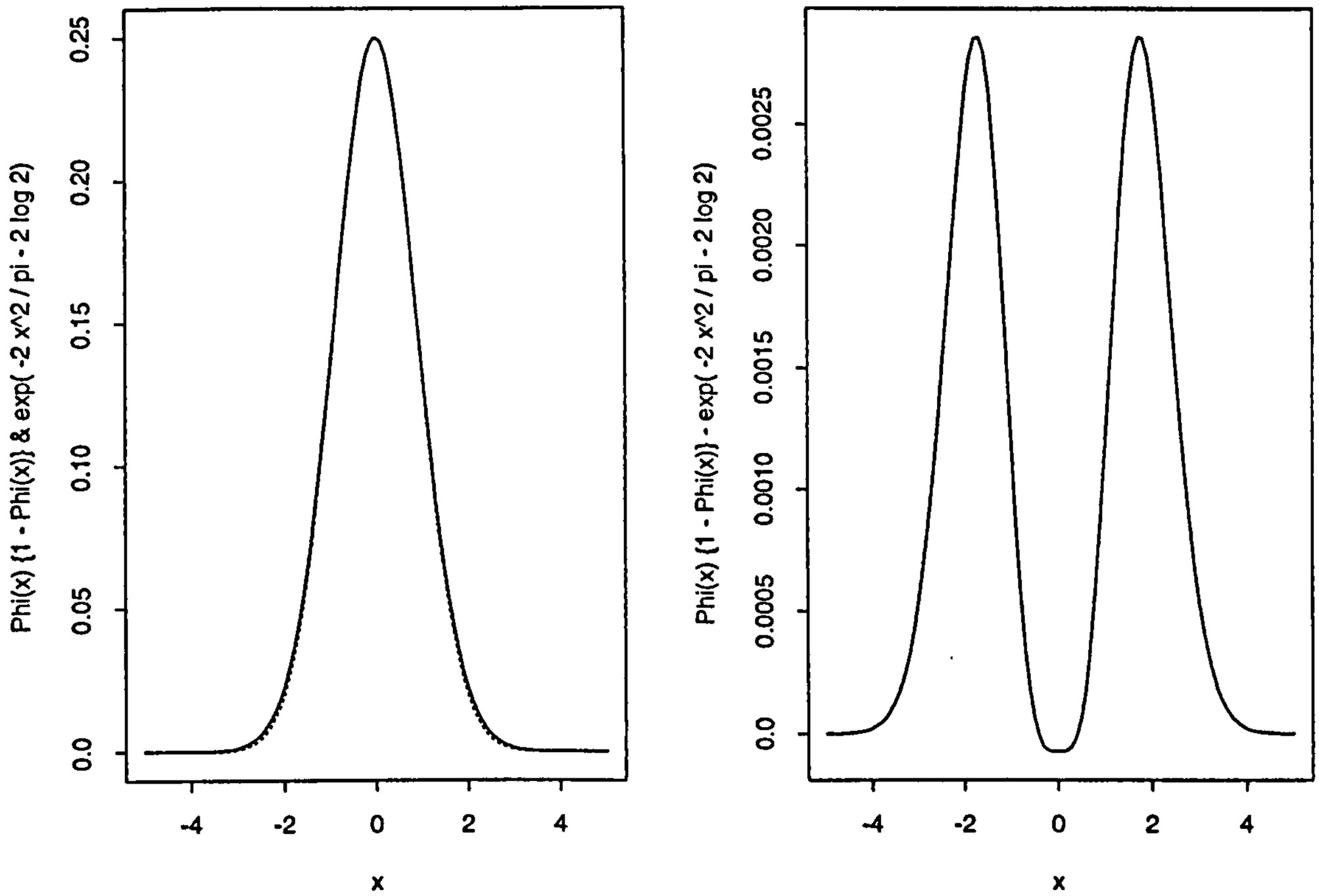


Figure 4.1: Plots of the approximation of $\Phi(x)\{1-\Phi(x)\}$ by $\exp(-2x^2/\pi - 2\log 2)$

then the function in (4.29) has the form

$$|\boldsymbol{B}|p_2(\boldsymbol{B}\underline{z}+\underline{\mu}) \times \phi_p(\underline{z}|\underline{0}, \boldsymbol{I}).$$

Therefore, as $p_2(\cdot)$ is now a polynomial in each of the components of $\underline{z}$, the function is approximately of the form (4.28) and can be accurately integrated using a Gauss–Hermite rule. In the above, $\boldsymbol{B}$ is a square root of the variance matrix $\boldsymbol{\Sigma}$, such as the Cholesky square root, and so $\boldsymbol{B}^T\boldsymbol{B} = \boldsymbol{\Sigma}$.

We now show that our integration problem is approximately of the form in (4.29) and tackle the problem of choosing $\underline{\mu}$ and $\boldsymbol{\Sigma}$.

First notice that the denominator in (4.25) is $\Phi(\cdot)\{1-\Phi(\cdot)\}$, given by the solid line in the left plot of Figure 4.1. Clearly the function resembles a normal likelihood and analysis of $\Phi(\cdot)\{1-\Phi(\cdot)\}$ shows that the function has only one turning point,

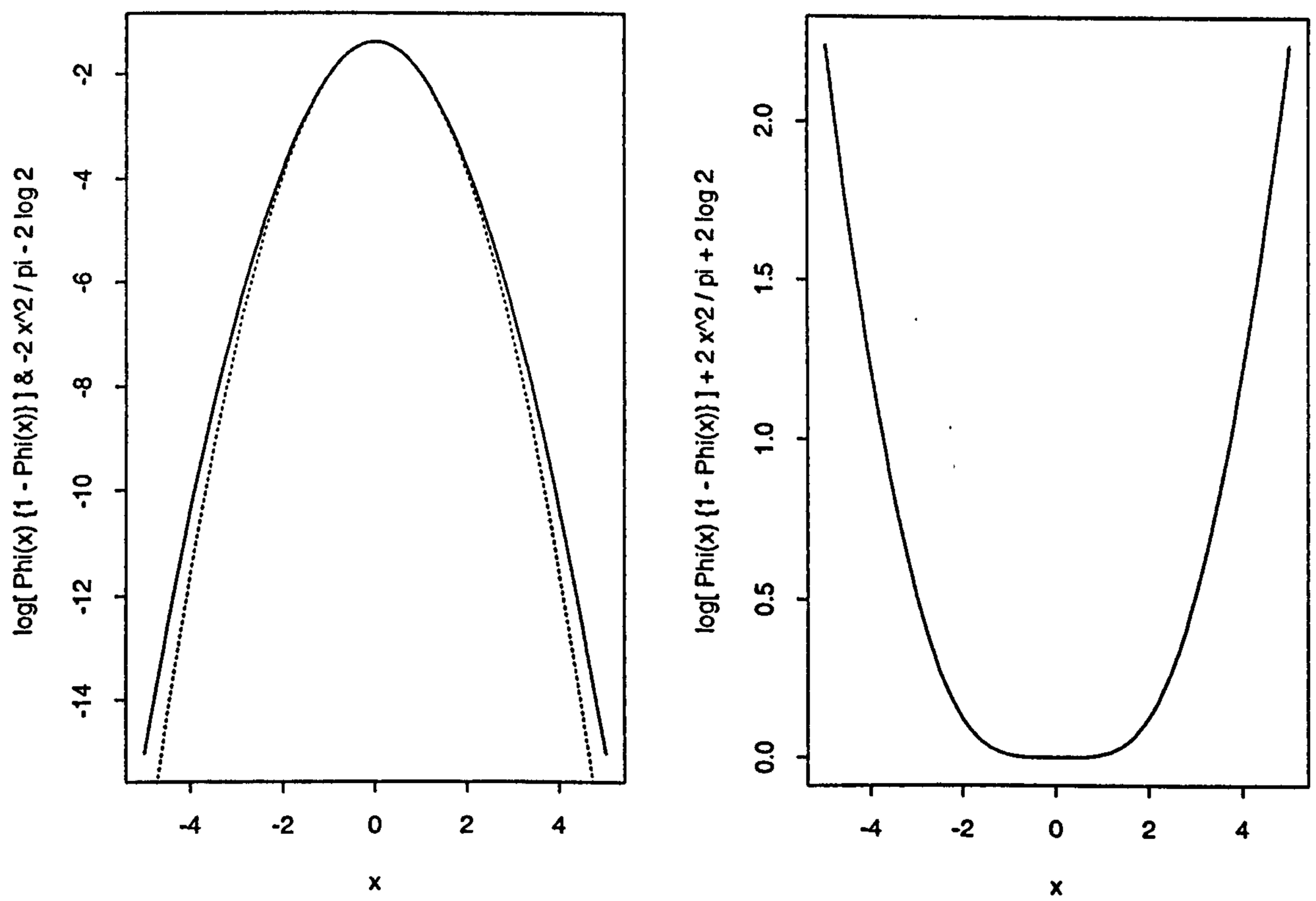


Figure 4.2: Plots of the approximation of $\log[\Phi(x)\{1-\Phi(x)\}]$ by $-2x^2/\pi - 2\log 2$

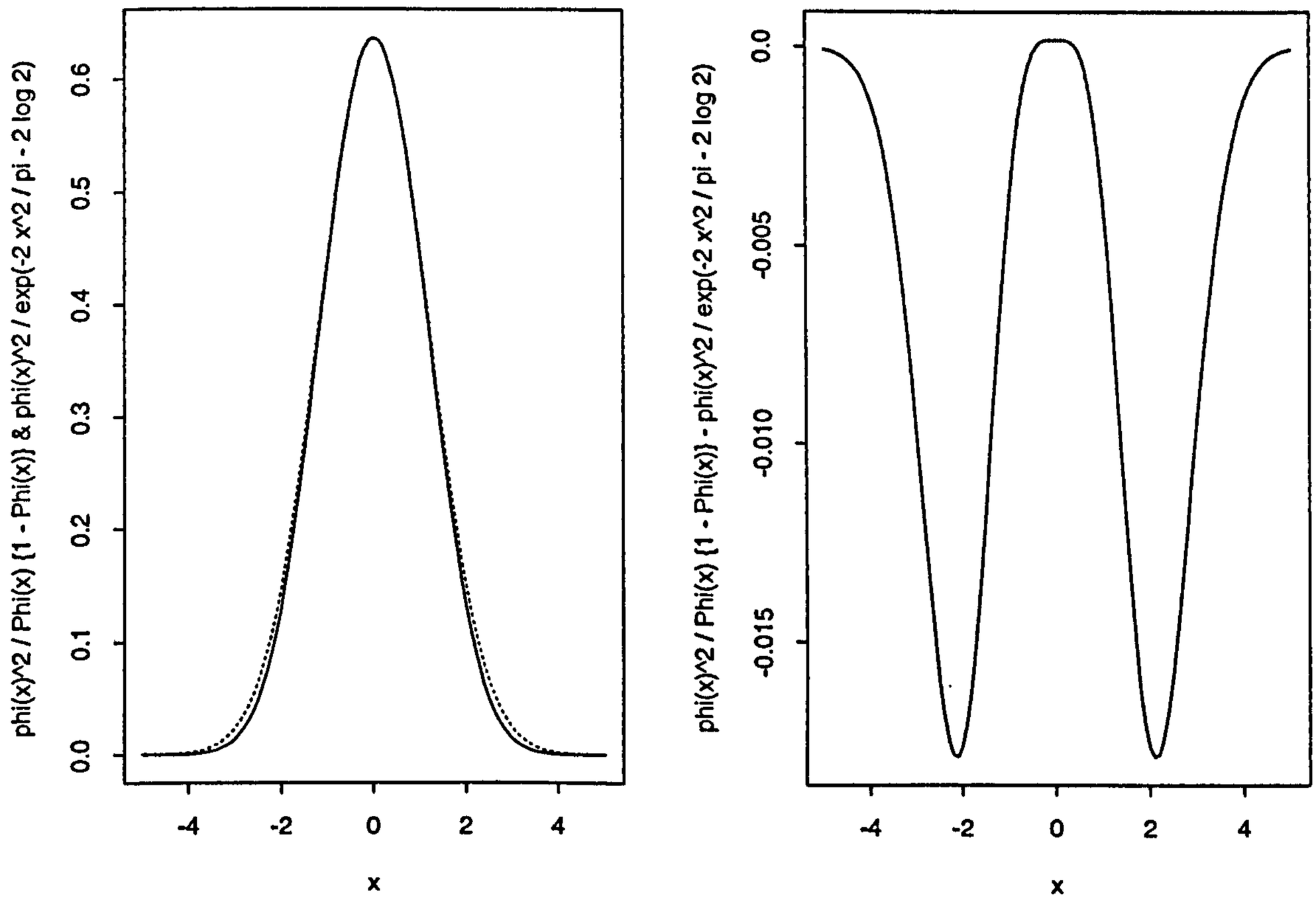


Figure 4.3: Plots of the approximation of $\phi(x)^2/\left[\Phi(x)\{1-\Phi(x)\}\right]$ by $\phi(x)^2/\exp\left(-2x^2/\pi - 2\log 2\right)$

the maximum at zero, and $\lim_{x\to\pm\infty}\Phi(x)\{1-\Phi(x)\}=0$. To find a normal likelihood that might approximate $\Phi(\cdot)\{1-\Phi(\cdot)\}$ one idea is to look at the first few terms in a Taylor series expansion of $\log[\Phi(x)\{1-\Phi(x)\}]$ about the origin. Such an expansion only has even terms and the first few are given by,

$$-2\log 2-\frac{2}{\pi}x^2+\left(\frac{2}{\pi^2}-\frac{2}{3\pi}\right)x^4+\cdots \tag{4.30}$$

Unfortunately we have not been able to discover whether the remainder term of this series tends to zero. However, Figure 4.2 shows that the first two terms of the series give a loose approximation to the function. In Figure 4.2, the solid line is given by $\log[\Phi(x)\{1-\Phi(x)\}]$ and the dotted line denotes the first two terms in (4.30). The discrepancy between the two functions is shown in the right hand plot. It is now natural to propose that $\Phi(x)\{1-\Phi(x)\}$ might be loosely approximated by the function

$$\exp\left(-\frac{2}{\pi}x^2-2\log 2\right).$$

In the left hand plot of Figure 4.1 the solid line is $\Phi(x)\{1-\Phi(x)\}$ and the dotted line is given by $\exp\left(-\frac{2}{\pi}x^2-2\log 2\right)$. The right hand plot shows the difference between the two functions. Notice that here the approximation appears to perform better. Finally, we assess whether $\phi(x)^2/[\Phi(x)\{1-\Phi(x)\}]$ may be approximated by

$$\frac{1}{2\pi}\exp\left\{-\left(1-\frac{2}{\pi}\right)x^2+2\log 2\right\}.$$

The left plot in Figure 4.3 displays the two functions and the right plot gives the disparity between them. The approximation appears reasonable and hence we take $\phi(x)^2/[\Phi(x)\{1-\Phi(x)\}]$ as approximately proportional to a $N(0,\pi/\{2(\pi-2)\})$ density. Note that here we are not intending to use the approximating functions as a substitute for the actual functions in any calculations, our purpose is to find a normal likelihood that looks something like the integrand in (4.26). The likelihood will then be used as a starting point for the scaling parameters of a Gauss–Hermite quadrature scheme. A full analytical examination of the functions discussed above may need to be carried out if the approximating functions are to be used for other purposes.

Returning to the integrand in (4.26) we can now see that it is approximately proportional to a polynomial in x_1 and x_2 multiplied by

$$q(\underline{x}) = \exp\left\{-\frac{1}{2a}\left(m_0 + \underline{x}^T \underline{m}_{(0)}\right)^2 - \frac{1}{2}\underline{x}^T \Lambda^{-1} \underline{x}\right\}$$

where

$$\begin{aligned} a &= \frac{\pi}{2(\pi-2)} \\ \underline{x}^T &= (x_1, x_2) \\ \underline{m}_{(0)}^T &= (m_1, m_2) \\ \Lambda^{-1} &= \frac{1}{(1-r_x^2)}\begin{pmatrix} 1 & -r_x \\ -r_x & 1 \end{pmatrix}. \end{aligned}$$

Plainly

$$\begin{aligned} q(\underline{x}) &\propto \exp\left[-\frac{1}{2}\left\{\underline{x}^T\left(\frac{1}{a}\underline{m}_{(0)}\,\underline{m}_{(0)}^T + \Lambda^{-1}\right)\underline{x} + \frac{2m_0}{a}\underline{x}^T\underline{m}_{(0)}\right\}\right] \\ &\propto \exp\left\{-\frac{1}{2}\left(\underline{x}-\underline{\mu}\right)^T \Sigma^{-1}\left(\underline{x}-\underline{\mu}\right)\right\}, \end{aligned}$$

where

$$\begin{aligned} \Sigma^{-1} &= \frac{1}{a}\underline{m}_{(0)}\,\underline{m}_{(0)}^T + \Lambda^{-1} \\ \underline{\mu} &= -\frac{m_0}{a}\Sigma\underline{m}_{(0)}. \end{aligned}$$

In order to find Σ and $\underline{\mu}$ we write out the matrix Σ^{-1}

$$\Sigma^{-1} = \frac{1}{a(1-r_x^2)}\begin{pmatrix} m_1^2(1-r_x^2)+a & m_1m_2(1-r_x^2)-r_xa \\ m_1m_2(1-r_x^2)-r_xa & m_2^2(1-r_x^2)+a \end{pmatrix}$$

and so

$$\Sigma = \frac{1}{m_1^2+m_2^2+2m_1m_2r_x+a}\begin{pmatrix} m_2^2(1-r_x^2)+a & r_xa-m_1m_2(1-r_x^2) \\ r_xa-m_1m_2(1-r_x^2) & m_1^2(1-r_x^2)+a \end{pmatrix} \tag{4.31}$$

and

$$\underline{\mu} = \frac{-m_0}{m_1^2+m_2^2+2m_1m_2r_x+a}\begin{pmatrix} m_1+r_xm_2 \\ r_xm_1+m_2 \end{pmatrix}. \tag{4.32}$$

It now follows that the integrand in (4.26) is approximately proportional to a polynomial multiplied by a bivariate Normal density function with mean $\underline{\mu}$ given in (4.32) and variance matrix Σ given by (4.31) and is approximately of the form in (4.29). The approximations in the above are imprecise but a Gauss–Hermite rule should perform well if it evaluates the function at enough points.

So far we have shown that Gauss–Hermite quadrature should do well for our integration problem by showing that the integrand is of an appropriate form and we have given a reasoned choice for the scaling parameters $\underline{\mu}$ and Σ. However, we have only shown that our choice of $\underline{\mu}$ and Σ corresponds to a loose approximation of the integrand and a better choice may exist. So, in our algorithm, we use the values of $\underline{\mu}$ and Σ in (4.32) and (4.31) as a starting point for the iterative rescaling method of Naylor & Smith (1982). Using their method, given current values of the scaling parameters, $\hat{\underline{\mu}}^{(\ell)}$ and $\hat{\Sigma}^{(\ell)}$ say, we suppose that, for $j = k = 0$, the integrand in (4.26) is a density function of two random variables, Y_1 and Y_2 say, up to a constant of proportionality. We then obtain new values of the scaling parameters by estimating the mean and variance matrix of the random variables. That is, we use Gauss–Hermite quadrature based on the current values $\hat{\underline{\mu}}^{(\ell)}$ and $\hat{\Sigma}^{(\ell)}$ to compute estimates, $\hat{c}$, $\widehat{E(Y_j)}$ and $\widehat{E(Y_jY_k)}$ respectively, of

$$\begin{aligned}
c &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \imath_{00}(\underline{m}, y_1, y_2)\phi_2(y_1, y_2|r_x)\,dy_1\,dy_2 \\
E(Y_j) &= \frac{1}{c}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y_j\imath_{00}(\underline{m}, y_1, y_2)\phi_2(y_1, y_2|r_x)\,dy_1\,dy_2 \\
E(Y_jY_k) &= \frac{1}{c}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y_jy_k\imath_{00}(\underline{m}, y_1, y_2)\phi_2(y_1, y_2|r_x)\,dy_1\,dy_2,
\end{aligned}$$

for $j, k = 1, 2$. The new values of the scaling parameters are then found by setting

$$\hat{\mu}_j^{(\ell+1)} = \widehat{E(Y_j)}$$

and

$$\hat{\Sigma}_{jk}^{(\ell+1)} = \widehat{E(Y_jY_k)} - \widehat{E(Y_j)}\widehat{E(Y_k)},$$

for $j, k = 1, 2$. The procedure is repeated until the estimates satisfy some convergence criteria (typically that any change in the value of the estimates between iterations is

below some threshold value). At the outset the scheme starts with a small number of design points, say 5, in each direction and then, when convergence is reached, the number of grid points in each direction is increased by one. The final estimates are taken when an increase in grid size also results in a negligible change in the estimates.

The integrals in (4.26) can easily be recovered from the final estimates obtained as

$$E_{X_1,X_2}\{\imath_{jk}(\underline{m}, X_1, X_2)\} = cE(Y_jY_k)$$

where $Y_0 = 1$ and for $j, k = 0, 1, 2$.

The result of each integration (4.26) is multiplied by 30 to give the expectation of Fisher's information matrix for 30 observations. The covariance structure is then computed by inverting this matrix using a routine based on Gaussian elimination.

Heuristic screen design

The heuristic design of the three–stage screen is computed by making a series of iterative steps that terminate when a convergence criteria in terms of Bayes cost is met. The algorithm to perform one iteration of this process is as follows. Recall that the parameters of the current design are denoted by (v_1', w_1', v_2', w_2') and those of the new design by $(v_1'', w_1'', v_2'', w_2'')$. When computing integrals in the following algorithm we approximate infinite bounds by ± 5, as appropriate, computing one dimensional integrals using the N.A.G. routine D01AHF and two dimensional integrals using the N.A.G. routine D01DAF, see N.A.G. (1990). Outside of the range ± 5 each integrand has negligible value.

(a) Second screen design

1. Calculate $\mu_1' = E(X_1 | v_1' < X_1 < w_1')$ using the formula given in equation (4.24). The standard normal distribution function $\Phi(\cdot)$ can be calculated using the N.A.G. routine G01EAF, (N.A.G. (1990)).

2. Compute the mean and covariance structure of the probit regression parameters $\underline{\xi}^*$ for the regression on $X_2 | X_1 = \mu_1'$. Formulae for this calculation

are given in section 4.3 in terms of the mean $\underline{m}$ (an input to the algorithm) and the covariance structure $\boldsymbol{S}$ (calculated above) of the probit regression model of T on X_1 and X_2.

3. Further, compute the mean and covariance structure of $\underline{\xi}^{\dagger}$, the regression parameters of a model that has moments with the property that $m_0^{\dagger} = 0$ and $s_1^{\dagger} = 1$. Section 4.3 contains formulae for computing these values from the moments of the regression parameters of $\underline{\xi}^{*}$ computed above.

4. Now use Theorem 2.1, with the values $m_1^{\dagger}$, $s_0^{\dagger}$ and $r_0^{\dagger}$ computed above and the cost parameters c_r^2, c_a^2 and c_m, to compute values $v_2^{\dagger}$ and $w_2^{\dagger}$. The values $v_2^{\dagger}$ and $w_2^{\dagger}$ are the parameter values for an optimal design of a single screen based on the regression parameters $\underline{\xi}^{\dagger}$.

5. The design parameters of the second screen, v_2'' and w_2'' can now be computed from equation (4.17).

6. Note that when obtaining the values $v_2^{\dagger}$ and $w_2^{\dagger}$ and hence v'' and w'', we must remember to ensure that they specify a global minimum of the one screen problem. That is, if we find that one or both of v_2'' and w_2'' represent a local minimum via Theorem 2.1, we should check whether the minimum is a global minimum by comparing the Bayes cost of the design with the cost of designs in which one or both of the design parameters take limiting values ($\pm\infty$). To check whether this is the case we must calculate the approximating Bayes cost $\hat{\mathcal{K}}^{(2)}$ in (4.15). Under our modelling assumptions, the approximate Bayes cost in (4.15), for finite v_2'' and w_2'', is

$$\begin{aligned}\hat{\mathcal{K}}^{(2)}(v_1', v_2'', w_1', w_2'') \;=\; & c_r^2 \int_{-\infty}^{v_2''} \Phi\left\{f(\mu_1', x_2)\right\} \phi(x_2 | r_x \mu_1', 1 - r_x^2) dx_2 \\ & + c_a^2 \int_{w_2''}^{\infty} \left[1 - \Phi\left\{f(\mu_1', x_2)\right\}\right] \phi(x_2 | r_x \mu_1', 1 - r_x^2) dx_2 \quad (4.33) \\ & + c_m \left[\Phi\left\{\frac{w_2'' - r_x \mu_1'}{(1 - r_x^2)^{1/2}}\right\} - \Phi\left\{\frac{v_2'' - r_x \mu_1'}{(1 - r_x^2)^{1/2}}\right\}\right] + c_s^2.\end{aligned}$$

where

$$f(x_1, x_2) = \frac{m_0 + m_1 x_1 + m_2 x_2}{1 + s_0^2 + x_1^2 s_1^2 + x_2^2 s_2^2 + 2 r_{01} s_0 s_1 x_1 + 2 r_{02} s_0 s_2 x_2 + 2 r_{12} s_1 s_2 x_1 x_2}.$$

If, for example, $v_2^\dagger$ (and hence v_2'') represents a local minimum then we compare (4.33) with the cost of $\hat{\mathcal{K}}^{(2)}(v_1', v_2, w_1', w_2'')$ when v_2 takes limiting values — the formula follows easily from equation (4.33). We proceed similarly if there is a local minimum at $w_2^\dagger$. We compare the Bayes cost of designs with the parameters of the finite (local) minima, with the Bayes cost for limiting values of the design parameters to assess where the true global minima lie. Note that throughout the simulation study we assume that $c_s^2 = 0$ and hence there is no need to compare the approximate Bayes cost $\hat{\mathcal{K}}^{(2)}$ with the cost of scenarios in which the screening cost is not paid, c.f. comment 4, section 2.2.

(b) **First screen design**

1. Work out the full Bayes cost, $\mathcal{K}^{(2)}(v_1', v_2'', w_1', w_2'')$ given in equation (4.13), of the second screen under the current design (the updated second screen design and the yet to be updated first screen design). Under the assumptions of our model, for the design (v_1, v_2, w_1, w_2), in which all parameters are finite,

$$\begin{aligned}\mathcal{K}^{(2)}(v_1, v_2, w_1, w_2) = \\ \Big(c_r^2 \int_{v_1}^{w_1}\int_{-\infty}^{v_2} \Phi\{f(x_1,x_2)\}\,\phi_2(x_1,x_2|r_x)dx_2dx_1 \\ + c_a^2 \int_{v_1}^{w_1}\int_{w_2}^{\infty} [1-\Phi\{f(x_1,x_2)\}]\,\phi_2(x_1,x_2|r_x)dx_2dx_1 \\ + c_m \int_{v_1}^{w_1}\int_{v_2}^{w_2} \phi_2(x_1,x_2|r_x)dx_2dx_1\Big)\{\Phi(w_1)-\Phi(v_1)\}^{-1} + c_s^2,\end{aligned} \tag{4.34}$$

where $f(x_1, x_2)$ is as given in (4.33) above.

2. Throughout the study we assume that X_2 is standardised with mean zero and so μ_2' in (4.18) is zero. Hence $\tilde{\xi}_0 = \xi_0$ and $\tilde{\xi}_1 = \xi_1$ and the moments of $\underline{\tilde{\xi}}$ follow directly.

3. Compute the mean and covariance structure of $\underline{\xi}^\dagger$, regression parameters that have moments so that $E(\xi_0^\dagger) = m_0^\dagger = 0$ and $Var(\xi_1^\dagger) = s_1^\dagger = 1$. Section 4.3 contains formulae for computing these moments from the moments of $\underline{\tilde{\xi}}$.

4. Follow Theorem 2.1 to find a solution $v_1^\dagger$ and $w_1^\dagger$, using values $m_1^\dagger$, $s_0^\dagger$ and $r_0^\dagger$ computed above and cost parameters c_r^1, c_a^1 and $\mathcal{K}^{(2)}(v_1', v_2'', w_1', w_2'')$, the latter having been computed in step 1. The parameters $v_1^\dagger$ and $w_1^\dagger$ describe an optimal design of a single screen based on a probit regression with parameters $\underline{\xi}^\dagger$.

5. The updated parameters for the first screen, v_1'' and w_1'', can now be obtained from equations (4.21).

6. Again we must ensure that the values $v_1^\dagger$ and $w_1^\dagger$ and hence v_1'' and w_1'' constitute a global minimum of this one screen problem. Hence we must calculate the approximate Bayes cost $\hat{\mathcal{K}}^{(1)}$ in (4.19) for v_1'' and w_1'' and compare the result with the Bayes cost when v_1 and/or w_1 take infinite values. Under our usual assumptions, for finite v_1'' and w_1'', the approximate Bayes cost $\hat{\mathcal{K}}^{(1)}$ in (4.19) is

$$\begin{aligned}\hat{\mathcal{K}}^{(1)}(v_1'', v_2'', w_1'', w_2'') &= c_r^1 \int_{-\infty}^{v_1''} \Phi\{f(x_1, 0)\}\, \phi(x_1|0, 1-r_x^2)dx_1 \\ &+ c_a^1 \int_{w_1''}^{\infty} [1 - \Phi\{f(x_1, 0)\}]\, \phi(x_1|0, 1-r_x^2)dx_1 \qquad (4.35) \\ &+ \mathcal{K}^{(2)}(v_1', v_2'', w_1', w_2'') \left[\Phi\left\{\frac{w_1''}{(1-r_x^2)^{1/2}}\right\} - \Phi\left\{\frac{v_1''}{(1-r_x^2)^{1/2}}\right\}\right] + c_s^1.\end{aligned}$$

Formulae for computing $\hat{\mathcal{K}}^{(1)}$ when v_1 and/or w_1 take limiting values follow easily from (4.35). The screening cost c_s^1 is assumed to be zero throughout our simulation study so we have no need to consider situations in which it is not paid (again refer to comment 4, section 2.2).

Convergence

The iterative procedure is terminated when $\mathcal{K}(v_1'', v_2'', w_1'', w_2'')$, the Bayes cost of the updated design, is sufficiently close to the Bayes cost of the current design $\mathcal{K}(v_1', v_2', w_1', w_2')$. That is, the initial convergence criteria devised for our scheme is that

$$|\mathcal{K}(v_1'', v_2'', w_1'', w_2'') - \mathcal{K}(v_1', v_2', w_1', w_2')|$$

is small enough. However, to ensure that the convergence criteria remains reasonably consistent whatever the Bayes costs we suppose that convergence is reached if

$$\frac{|\mathcal{K}(v_1'', v_2'', w_1'', w_2'') - \mathcal{K}(v_1', v_2', w_1', w_2')|}{\mathcal{K}(v_1', v_2', w_1', w_2')} < \delta_c \tag{4.36}$$

for some small δ_c. Here the absolute difference between the costs is assessed relative to $\mathcal{K}(v_1', v_2', w_1', w_2')$, the cost of the current design. Throughout our study we fixed $\delta_c = 10^{-5}$. To assess whether the convergence criteria has been met we compute the Bayes cost $\mathcal{K}(v_1, v_2, w_1, w_2)$ for the updated and current design. Under our modelling assumptions, from (4.13) the full Bayes cost $\mathcal{K}$ of a design with finite parameters $(\underline{v}, \underline{w})$ is given by

$$\begin{aligned}\mathcal{K}(v_1, v_2, w_1, w_2) &= c_r^1 \int_{-\infty}^{v_1} \int_{-\infty}^{\infty} \Phi\{f(x_1, x_2)\} \phi_2(x_1, x_2 | r_x) dx_2 dx_1 \\ &+ c_a^1 \int_{w_1}^{\infty} \int_{-\infty}^{\infty} [1 - \Phi\{f(x_1, x_2)\}] \phi_2(x_1, x_2 | r_x) dx_2 dx_1 \\ &+ \mathcal{K}^{(2)}(v_1, v_2, w_1, w_2) \{\Phi(w_1) - \Phi(v_1)\} + c_s^1,\end{aligned} \tag{4.37}$$

where $\mathcal{K}^{(2)}(v_1, v_2, w_1, w_2)$ and $f(x_1, x_2)$ are as given in (4.34) above. We denote the design reached at convergence as $(\hat{v}_1, \hat{v}_2, \hat{w}_1, \hat{w}_2)$.

The optimal design

So far we have described how to compute a heuristic design, $(\hat{v}_1, \hat{v}_2, \hat{w}_1, \hat{w}_2)$, for the three–stage sequential screen and the Bayes cost of this design. To assess whether our heuristic scheme is performing well we compare the heuristic design with the fully optimal design that minimises Bayes cost $\mathcal{K}$. The fully optimal design can be found using numerical techniques and the routine we use is due to Davidon–Fletcher–Powell (see Chambers (1973)). This routine finds the minimum of a function with respect to one or more variables given first derivatives of the function and starting values for each variable. Here the function we need to minimise is $\mathcal{K}$ in (4.37) above, with respect to the parameters of the screen design (v_1, v_2, w_1, w_2). We set $(\underline{v}, \underline{w}) = (-1, -1, 1, 1)$ as starting values for finding the fully optimal design and, for reference, we give below

the derivatives of Bayes cost $\mathcal{K}$ with respect to each of v_1, v_2, w_1 and w_2.

$$\begin{aligned}\frac{(1-r_x^2)^{1/2}}{\phi(v_1)}\frac{\partial \mathcal{K}}{\partial v_1} &= c_r^1 \int_{-\infty}^{\infty} \Phi\{f(v_1,x_2)\}\,\phi\left\{\frac{x_2-r_x v_1}{(1-r_x^2)^{1/2}}\right\}dx_2 \\ &\quad - c_r^2 \int_{-\infty}^{v_2} \Phi\{f(v_1,x_2)\}\,\phi\left\{\frac{x_2-r_x v_1}{(1-r_x^2)^{1/2}}\right\}dx_2 \\ &\quad - c_a^2 \int_{w_2}^{\infty} [1-\Phi\{f(v_1,x_2)\}]\,\phi\left\{\frac{x_2-r_x v_1}{(1-r_x^2)^{1/2}}\right\}dx_2 \\ &\quad - c_m(1-r_x^2)^{1/2}\left[\Phi\left\{\frac{w_2-r_x v_1}{(1-r_x^2)^{1/2}}\right\}-\Phi\left\{\frac{v_2-r_x v_1}{(1-r_x^2)^{1/2}}\right\}\right] \\ &\quad - c_s^2(1-r_x^2)^{1/2},\end{aligned}$$

$$\begin{aligned}\frac{(1-r_x^2)^{1/2}}{\phi(w_1)}\frac{\partial \mathcal{K}}{\partial w_1} &= -c_a^1 \int_{-\infty}^{\infty} [1-\Phi\{f(w_1,x_2)\}]\,\phi\left\{\frac{x_2-r_x w_1}{(1-r_x^2)^{1/2}}\right\}dx_2 \\ &\quad + c_r^2 \int_{-\infty}^{v_2} \Phi\{f(w_1,x_2)\}\,\phi\left\{\frac{x_2-r_x w_1}{(1-r_x^2)^{1/2}}\right\}dx_2 \\ &\quad + c_a^2 \int_{w_2}^{\infty} [1-\Phi\{f(w_1,x_2)\}]\,\phi\left\{\frac{x_2-r_x w_1}{(1-r_x^2)^{1/2}}\right\}dx_2 \\ &\quad + c_m(1-r_x^2)^{1/2}\left[\Phi\left\{\frac{w_2-r_x w_1}{(1-r_x^2)^{1/2}}\right\}-\Phi\left\{\frac{v_2-r_x w_1}{(1-r_x^2)^{1/2}}\right\}\right] \\ &\quad + c_s^2(1-r_x^2)^{1/2},\end{aligned}$$

$$\begin{aligned}\frac{(1-r_x^2)^{1/2}}{\phi(v_2)}\frac{\partial \mathcal{K}}{\partial v_2} &= c_r^2 \int_{v_1}^{w_1} \Phi\{f(x_1,v_2)\}\,\phi\left\{\frac{x_1-r_x v_2}{(1-r_x^2)^{1/2}}\right\}dx_1 \\ &\quad - c_m(1-r_x^2)^{1/2}\left[\Phi\left\{\frac{w_1-r_x v_2}{(1-r_x^2)^{1/2}}\right\}-\Phi\left\{\frac{v_1-r_x v_2}{(1-r_x^2)^{1/2}}\right\}\right],\end{aligned}$$

$$\begin{aligned}\frac{(1-r_x^2)^{1/2}}{\phi(w_2)}\frac{\partial \mathcal{K}}{\partial w_2} &= -c_a^2 \int_{v_1}^{w_1} [1-\Phi\{f(x_1,w_2)\}]\,\phi\left\{\frac{x_1-r_x w_2}{(1-r_x^2)^{1/2}}\right\}dx_1 \\ &\quad + c_m(1-r_x^2)^{1/2}\left[\Phi\left\{\frac{w_1-r_x w_2}{(1-r_x^2)^{1/2}}\right\}-\Phi\left\{\frac{v_1-r_x w_2}{(1-r_x^2)^{1/2}}\right\}\right],\end{aligned}$$

Again, integration is carried out using the N.A.G.routine D01AHF, with infinite limits approximated by ± 5, as appropriate. The design computed using this minimisation routine specifies a Bayes sequential design (see Definition 4.1) of the screen, with Bayes cost $\mathcal{K}(v_1^{opt}, w_1^{opt}, v_2^{opt}, w_2^{opt})$.

4.4.3 Results

The sensitivity of the performance of the heuristic design procedure to input parameters can now be assessed and we do so by a series of simulation studies. As the heuristic scheme is only appropriate for designing a sequential screen which has at least two $\underline{X}$–stages we discard any problems in which it is optimal to screen using 0 or 1 covariates. This may bias our results slightly but we argue that we should avoid making claims about the performance of the heuristic in cases in which its use is not appropriate.

(i) **Initial values.** The first study deals with the initial start-up values, (v_1^0, w_1^0), of the heuristic procedure. We investigate whether the heuristic design alters under different start-up values and hence we compare both the designs at convergence $(\hat{v}_1, \hat{v}_2, \hat{w}_1, \hat{w}_2)$ and the Bayes costs of such designs, under varying values of (v_1^0, w_1^0).

In the study we generated 1000 design problems, each problem taking values of the input parameters generated from independent uniform distributions on the following intervals: $c_r^1 \in (2.5, 15)$, $c_a^1 \in (2.5, 15)$, $c_r^2 \in (2.5, 15)$, $c_a^2 \in (2.5, 15)$, $m_0 \in (-1, 1)$, $m_1 \in (0.1, 3.1)$, $m_2 \in (0.1, 3.1)$ and $r_x \in (0.0, 0.75)$. For each problem we further generated ten sets of initial value parameters (v_1^0, w_1^0) where both v_1^0 and w_1^0 were generated from independent uniform distributions on $(-3, 3)$.

In each of the one thousand problems, for all ten pairs of (v_1^0, w_1^0) the iterative procedure converged to the same Bayes cost to within the desired accuracy specified by the choice of δ_c in (4.36). This achievement is despite the fact that for many cases the starting values were such that $w_1^0 < v_1^0$. In all cases the screening parameters at convergence $(\hat{v}_1, \hat{v}_2, \hat{w}_1, \hat{w}_2)$ were seen to be the same to an accuracy of at least five decimal places. The smallest number of steps taken to convergence was 2 and the largest number was 8. In each problem the number of iterations until convergence varied by at most 2 steps.

As it is apparent that the choice of initial values has little effect on the design reached at convergence, throughout the remaining studies we fix $v_1^0 = -1$ and $w_1^0 = 1$.

In the remainder of the simulation studies we assess the performance of the heuristic design by (i) contrasting the design at convergence with the fully optimal design and (ii) comparing the Bayes costs of the two designs. To assess the difference between the screening parameters of the heuristic design and those of the optimal design for each stage of the screen we use a standard deviation measure given by,

$$\Delta_i = \left[\frac{1}{2}\left\{(\hat{v}_i - v_i^{opt})^2 + (\hat{w}_i - w_i^{opt})^2\right\}\right]^{1/2},$$

for $i = 1, 2$. As X_1 and X_2 are assumed standardised normal variates, cases in which the absolute value of any screening parameter is greater than or equal to 5 are effectively cases in which at least one of the screening parameters is given by an infinite limit. To prevent such cases from distorting the results of the study we do not calculate Δ_i for those problems in which one of $|\hat{v}_i|$, $|v_i^{opt}|$ $|\hat{w}_i|$ or $|w_i^{opt}|$ is found to be greater than 5.

The difference in Bayes costs is measured by the percentage discrepancy between the Bayes cost of the heuristic design and the Bayes cost of the fully optimal design, given by

$$\Delta\mathcal{K} = \frac{\mathcal{K}(\hat{v}_1, \hat{v}_2, \hat{w}_1, \hat{w}_2) - \mathcal{K}(v_1^{opt}, w_1^{opt}, v_2^{opt}, w_2^{opt})}{\mathcal{K}(v_1^{opt}, w_1^{opt}, v_2^{opt}, w_2^{opt})} \times 100.$$

In the main we present our results in the form of box—and—whisker plots. As there is a variety of styles for such plots it seems appropriate to quickly describe the style used here. The edges of the box are formed at the lower and upper quartiles respectively with the median drawn as a line that divides the box. The whiskers are dashed lines that extend from the lower and upper quartiles to the most extreme value not beyond a standard span from the quartiles, where a standard span is defined as 1.5×Interquartile Range. Those data values more extreme than the whiskers are marked separately. The width of the box is proportional to the square root of the number of data points used in the plot.

We performed two further studies, each under a different cost set–up. In the first we assumed that the costs of wrongly accepting and wrongly rejecting an item remain fixed through both stages of the screen, with $c_r = c_r^1 = c_r^2$ and $c_a = c_a^1 = c_a^2$. The second study supposes that there may be different costs of misclassification at the two different $\underline{X}$–stages of the screen but at each stage the penalty for wrongly sentencing an item is the same whether retaining a bad item or discarding a good item. Here we write $c^1 = c_r^1 = c_a^1$ and $c^2 = c_r^2 = c_a^2$. Hence we first vary costs within stages and secondly, between stages. Both of these cost scenarios are likely to occur in practice.

Study 1

We generated 5000 design problems with input parameters drawn from independent uniform distributions on the following intervals: $c_r \in (2.5, 15)$, $c_a \in (2.5, 15)$, $m_0 \in (-1, 1)$, $m_1 \in (0.1, 3.1)$, $m_2 \in (0.1, 3.1)$ and $r_x \in (0.0, 0.75)$. The parameters m_1 and m_2 were permuted to create a re–ordering of the stages of the screen and the heuristic and fully optimal designs, together with their associated Bayes costs, were obtained for both orderings. Hence for each set of inputs we have two values for m_1 and m_2 and two observations on each of Δ_1, Δ_2 and $\Delta\mathcal{K}$. We now assess the effect of the inputs to the algorithm on the performance of the heuristic.

(ii.i) Cost parameters.

Effects on design. Figure 4.4 and Figure 4.5 display boxplots of Δ_1 and Δ_2 against (a) c_r, (b) c_a, both grouped in intervals of 1.25, (c) $c_r - c_a$ and (d) $c_r + c_a$, both grouped in intervals of 2.5. The effect of the magnitudes of c_r and c_a on the difference between the heuristic and optimal designs seems minimal. However, it seems possible that the difference between the designs is likely to be slightly larger when c_r and/or c_a are larger. Also, the plots which display Δ_i against $c_r + c_a$ indicate more strongly that performance in design terms is likely to be worse for cases in which the sum of the costs is large. The effect of the difference $c_r - c_a$ on Δ_i also seems minimal with slightly worse performances when the differences are small. However, care should be taken when making

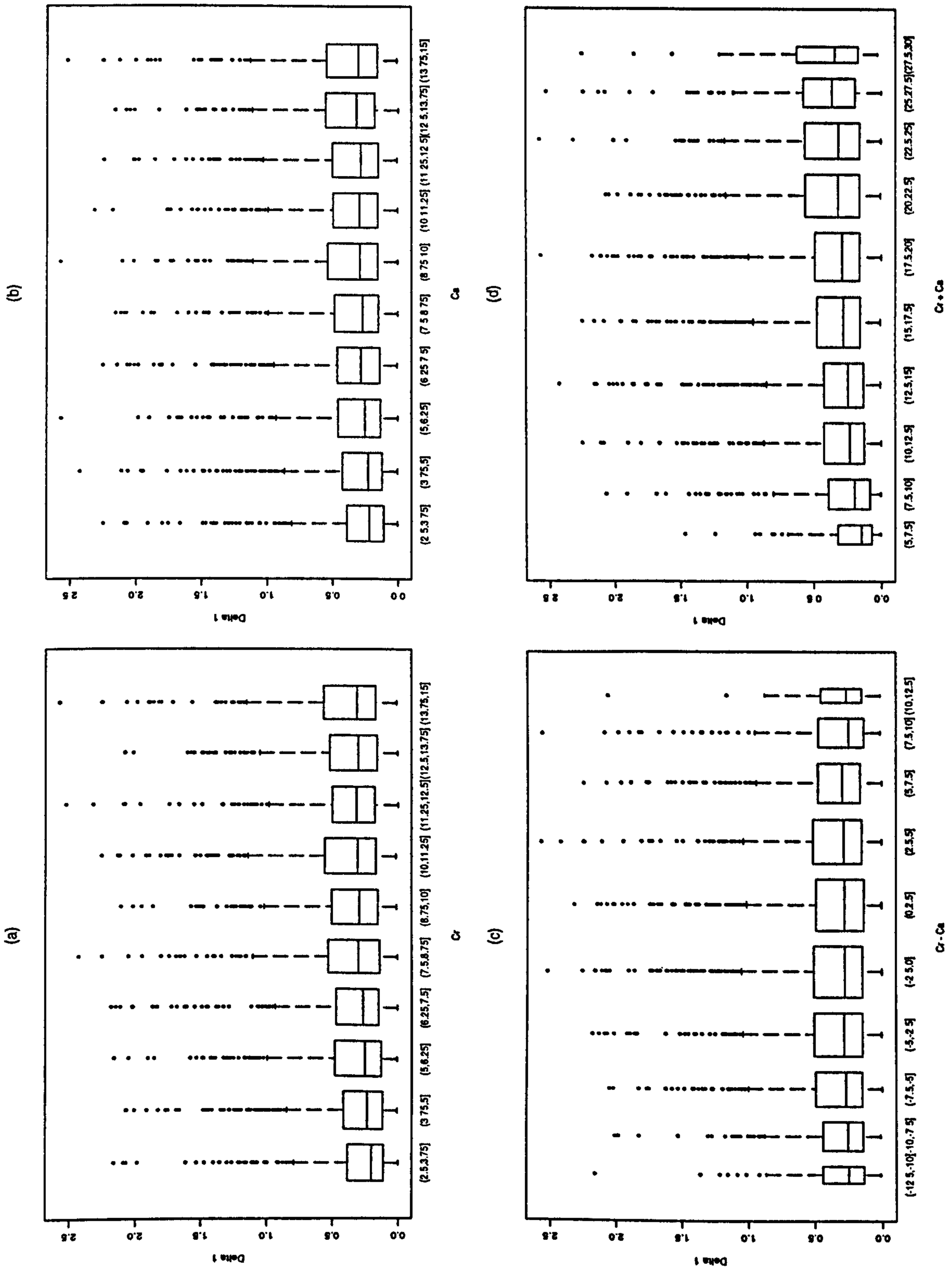


Figure 4.4: Boxplots of Δ_1 for simulated values of (a) c_r, (b) c_a, (c) $c_r - c_a$, and (d) $c_r + c_a$.

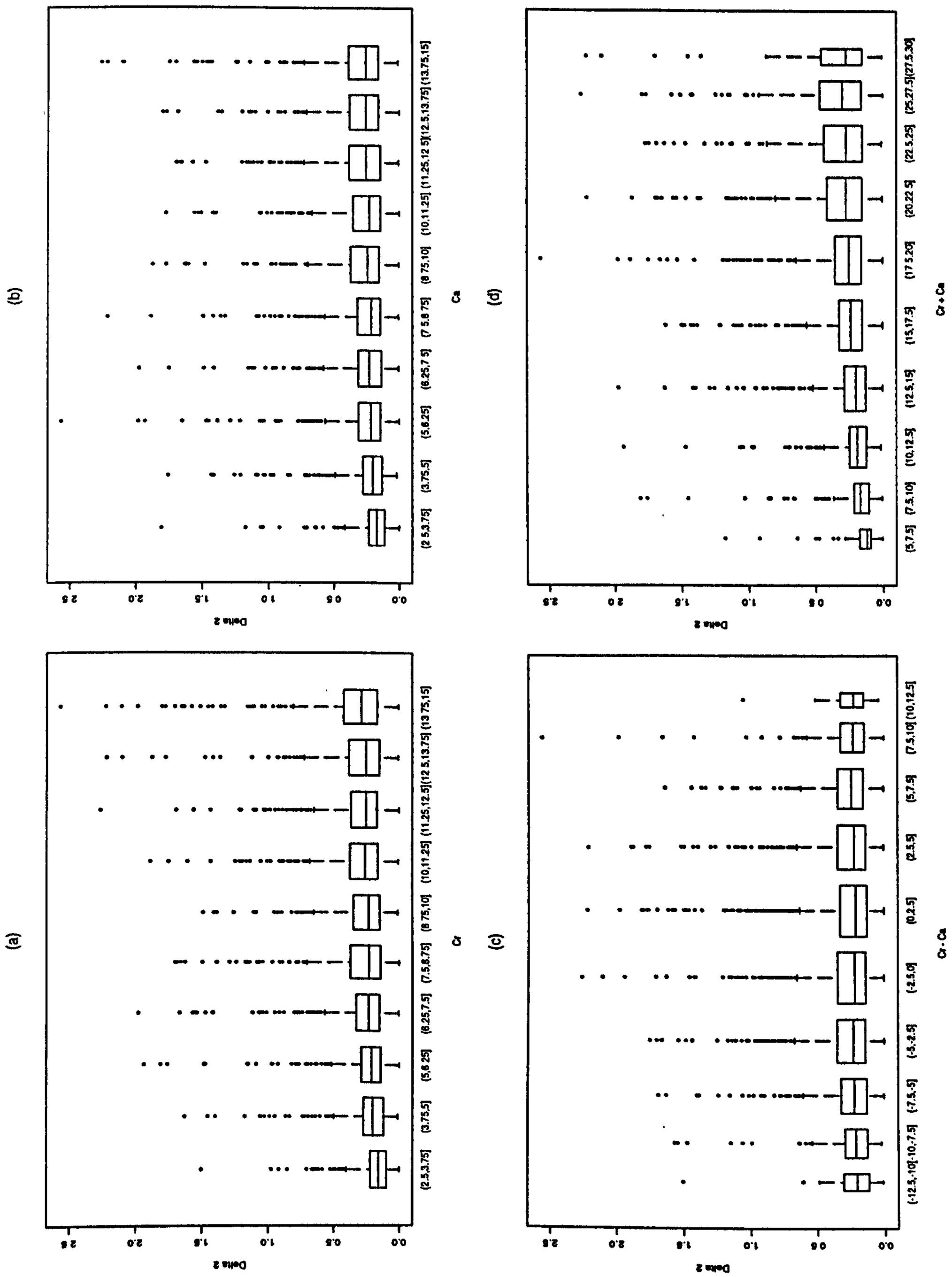


Figure 4.5: Boxplots of Δ_2 for simulated values of (a) c_r, (b) c_a, (c) $c_r - c_a$, and (d) $c_r + c_a$.

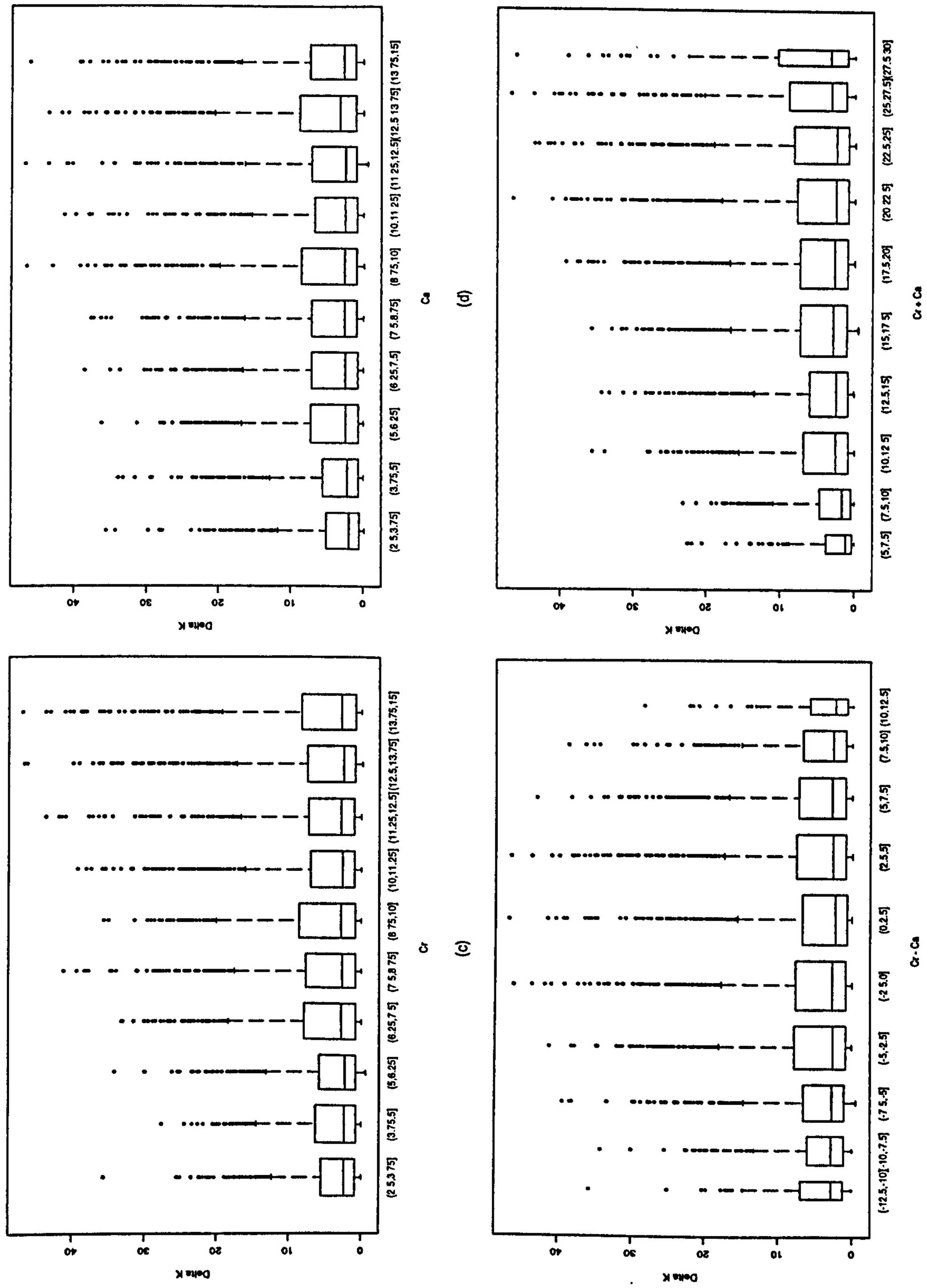


Figure 4.6: Boxplots of $\Delta\mathcal{K}$ for simulated values of (a) c_r, (b) c_a, (c) $c_r - c_a$, and (d) $c_r + c_a$.

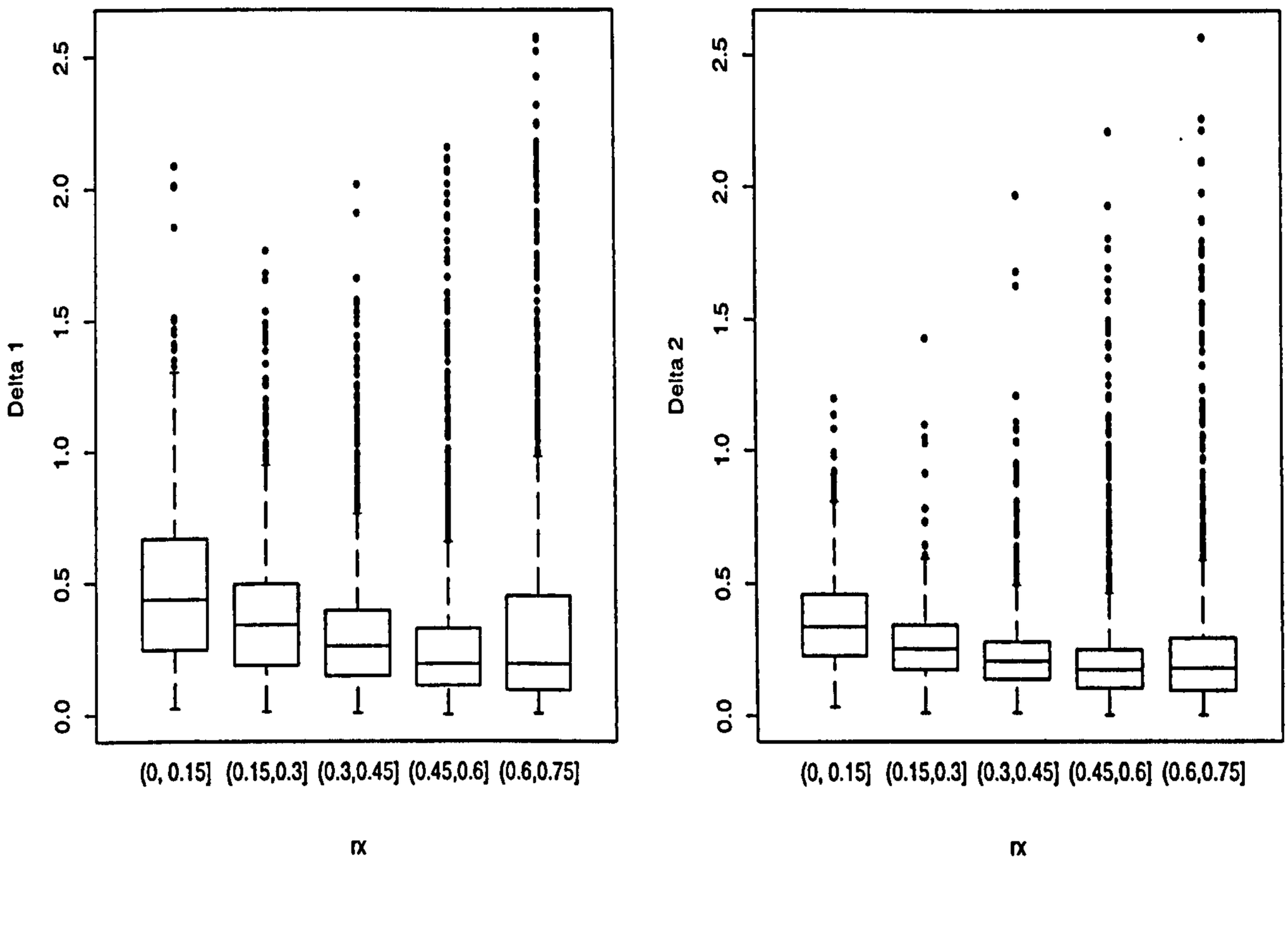


Figure 4.7: Boxplots of (a) Δ_1 and (b) Δ_2 for simulated values of r_x.

inferences from the plots for $c_r - c_a$ and $c_r + c_a$. It is difficult to distinguish effects as due to $c_r - c_a$ or $c_r + c_a$ as when $c_r + c_a$ is very large or very small then $|c_r - c_a|$ must be small. Similarly when $|c_r - c_a|$ is large then $c_r + c_a$ must take a value in the middle of its range.

Effects on cost. The degree of suboptimality of the heuristic in terms of Bayes cost is shown in Figure 4.6, with boxplots of $\Delta\mathcal{K}$ versus (a) c_r, (b) c_a, (c) $c_r - c_a$ and (d) $c_r + c_a$, grouped as before. Again the effect of cost parameters on performance seems minimal and the only observation worthy of note is that there is some indication that $\Delta\mathcal{K}$ is more likely to be larger when the sum of cost parameters $c_r + c_a$ is larger.

(iii.i) **Distribution of $\underline{X}$.**

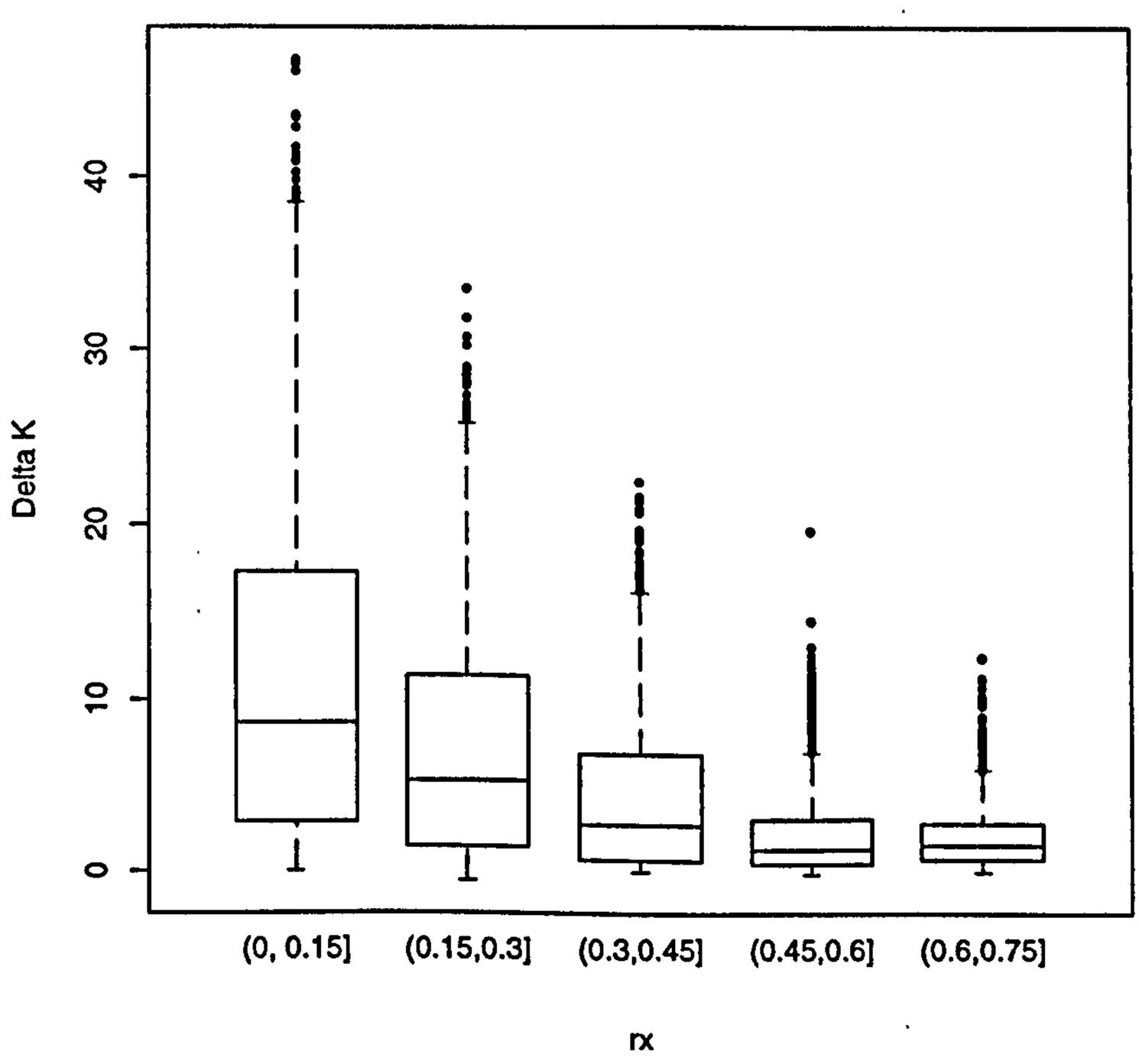


Figure 4.8: Boxplots of $\Delta\mathcal{K}$ for simulated values of r_x.

Effects on design. The boxplots in Figure 4.7 show Δ_1 and Δ_2, respectively, against the correlation coefficient r_x, grouped in intervals of 0.15. In both cases, 'typical' values for Δ_i are more likely to be small when r_x is large, except for r_x very large. However, the length of the tail, which denotes large observations also increases as r_x increases.

Effects on cost. Figure 4.8 plots percentage discrepancy in Bayes cost against r_x, again grouped in intervals of 0.15. It is very clear from this plot that the heuristic performs more consistently well when r_x is large.

(iv.i) Probit regression parameters.

Effects on design. The influence of the mean of the probit regression parameters on the heuristic design can be assessed by Figures 4.9 & 4.10. Here Δ_1 and Δ_2 are plotted against (a) m_0, grouped in intervals of 0.25, (b) m_1 and (c) m_2, both grouped in intervals of 0.3. The difference in the heuristic and optimal parameters of the first stage can be seen to be typically larger when $|m_0|$ is larger, m_1 is smaller or m_2 is larger. Δ_2 is also larger when $|m_0|$ is larger and, ignoring the first boxplot for m_1 which is based on relatively few observations, the effect of m_1 and m_2 on Δ_2 is the reverse of their effect on Δ_1.

Effects on cost. Figure 4.11 shows boxplots of $\Delta\mathcal{K}$ against (a) m_0, (b) m_1 and (c) m_2. Here it is clear that the heuristic is more likely to perform well when $|m_0|$, m_1 or m_2 is small. The m_2 effect is the most marked.

Study 2

The second study followed along the same lines as the first. Another 5000 problems were generated using a similar scheme for drawing input parameters as in the first study. Here the only difference was that the cost parameters were $c^1 = c_a^1 = c_r^1$ and $c^2 = c_a^2 = c_r^2$ both drawn from a uniform distribution on $(2.5, 15)$. That is, misclassification costs varied between X-screens rather than within X-screens. When permuting the covariate stages of the screen we swapped c^1 and c^2 as well as m_1 and

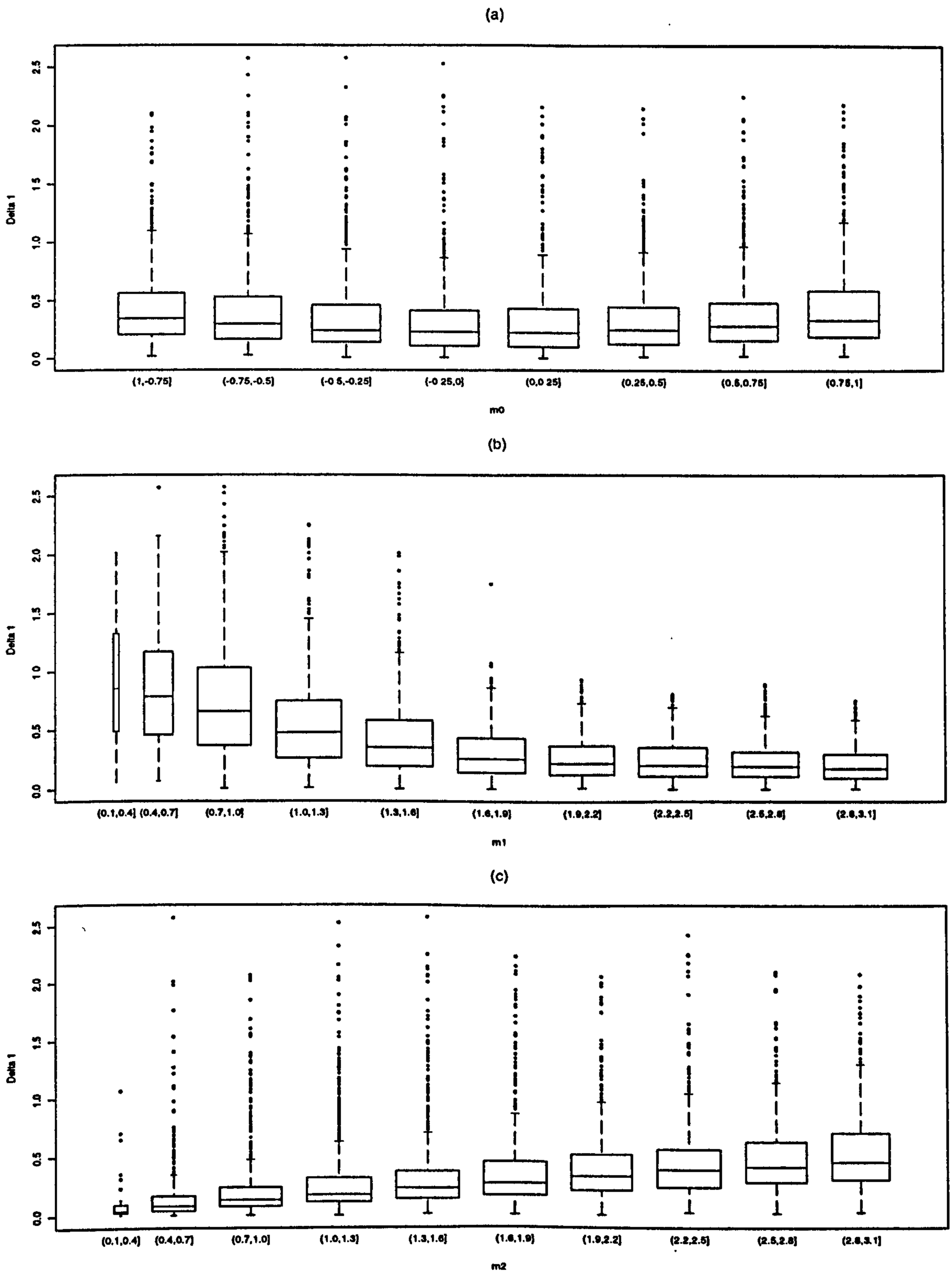


Figure 4.9: Boxplots of Δ_1 for simulated values of (a) m_0, (b) m_1, and (c) m_2.

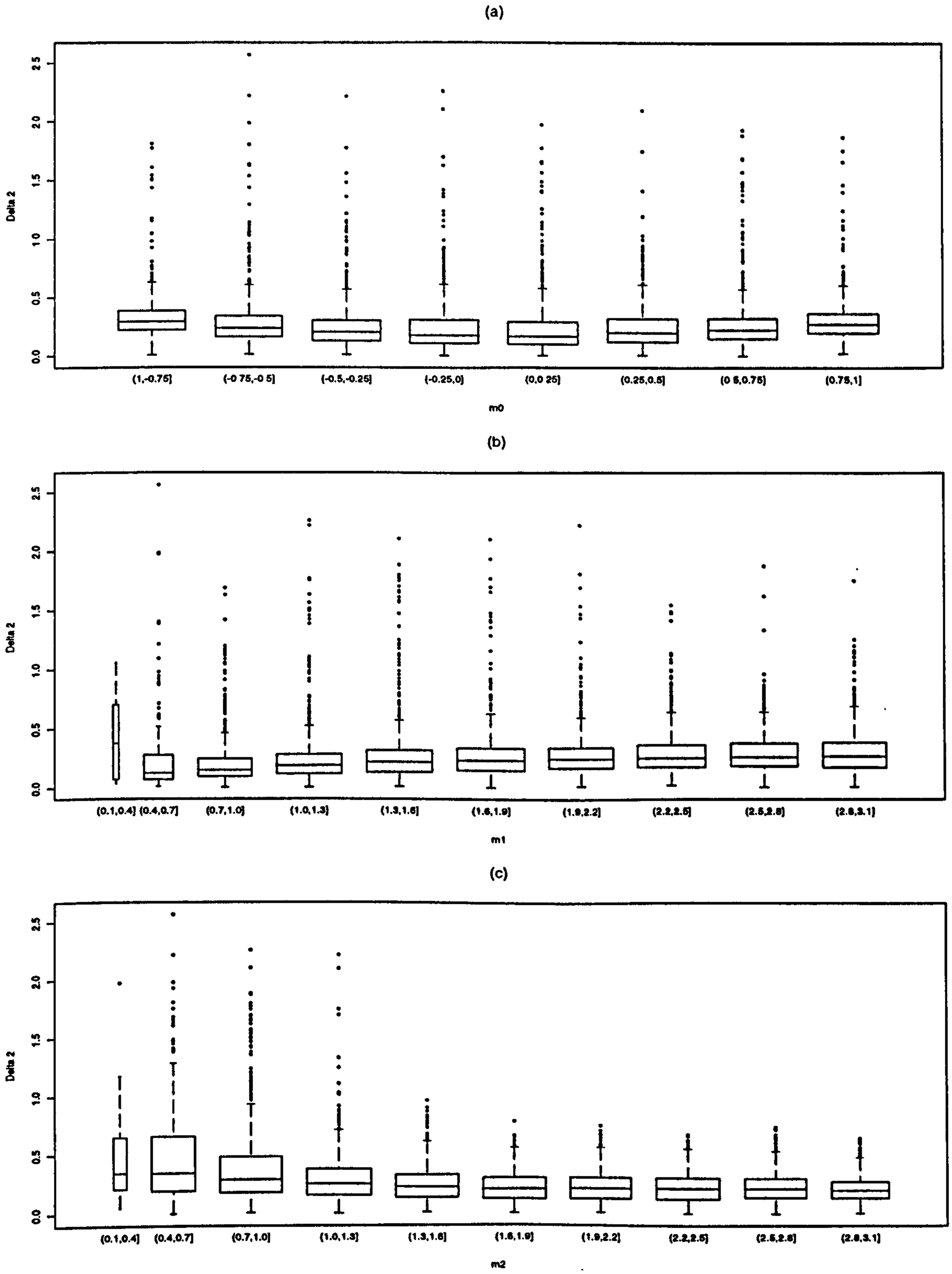


Figure 4.10: Boxplots of Δ_2 for simulated values of (a) m_0, (b) m_1, and (c) m_2.

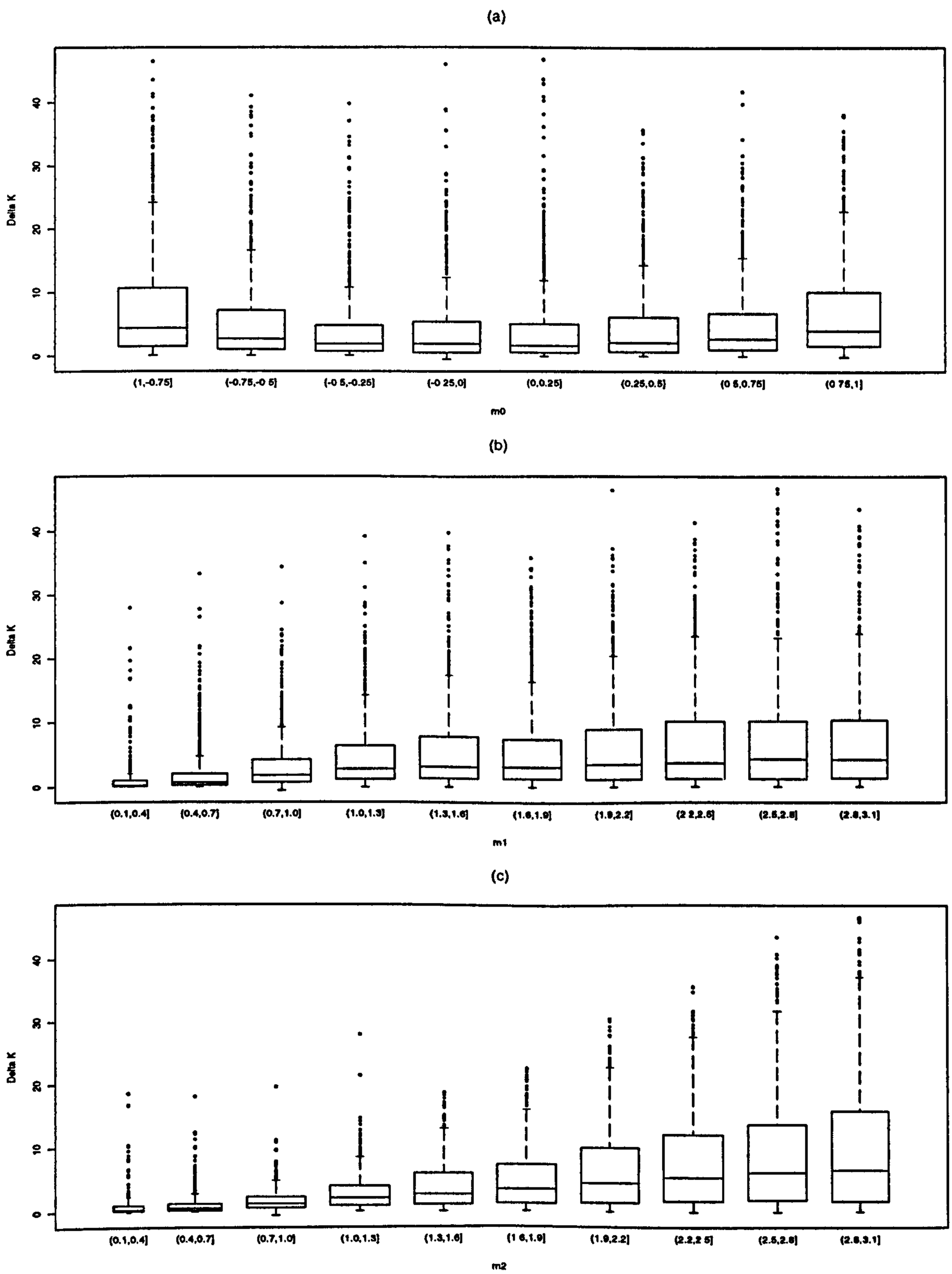


Figure 4.11: Boxplots of $\Delta\mathcal{K}$ for simulated values of (a) m_0, (b) m_1, and (c) m_2.

m_2.

(ii.ii) Cost parameters.

Effects on design. Figure 4.12 and Figure 4.13 display boxplots of Δ_1 and Δ_2 against (a) c^1, (b) c^2, both grouped in intervals of 1.25, (c) $c^1 - c^2$ and (d) $c^1 + c^2$, both grouped in intervals of 2.5. Again there is no clear correlation between the performance of the heuristic and the cost parameters, with the possibility that larger magnitudes of c^1 and c^2 give a slightly worse performance.

Effects on cost. A comparison of the heuristic and optimal designs in terms of Bayes cost is shown in Figure 4.14, with boxplots of $\Delta\mathcal{K}$ versus (a) c^1, (b) c^2, (c) $c^1 - c^2$ and (d) $c^1 + c^2$, grouped as before. As in Study 1 the only inference we can make is a tentative remark that the heuristic does worse when the cost parameters are larger.

(iii.ii) Distribution of $\underline{X}$.

Effects on design. Figure 4.15 shows boxplots of Δ_1 and Δ_2, respectively, against the correlation coefficient r_x, grouped in intervals of 0.15. Again the 'location' of the performance indicators Δ_1 and Δ_2 improve with r_x but their variability of also increases with r_x.

Effects on cost. Figure 4.16 plots percentage discrepancy in Bayes cost against r_x, again grouped in intervals of 0.15. The very clear relationship between the level of suboptimality and the correlation coefficient r_x is repeated.

(iv.ii) Probit regression parameters.

Effects on design. Figures 4.17 & 4.18 plot Δ_1 and Δ_2 against (a) m_0, grouped in intervals of 0.25, (b) m_1 and (c) m_2, both grouped in intervals of 0.3. The effects reported in (iv.i) above are replicated.

Effects on cost. Figure 4.19 shows boxplots of $\Delta\mathcal{K}$ against (a) m_0, (b) m_1 and (c) m_2. Again it is clear that the heuristic is more likely to perform well when $|m_0|$, m_1 or m_2 are small, with the m_2 effect the most distinct.

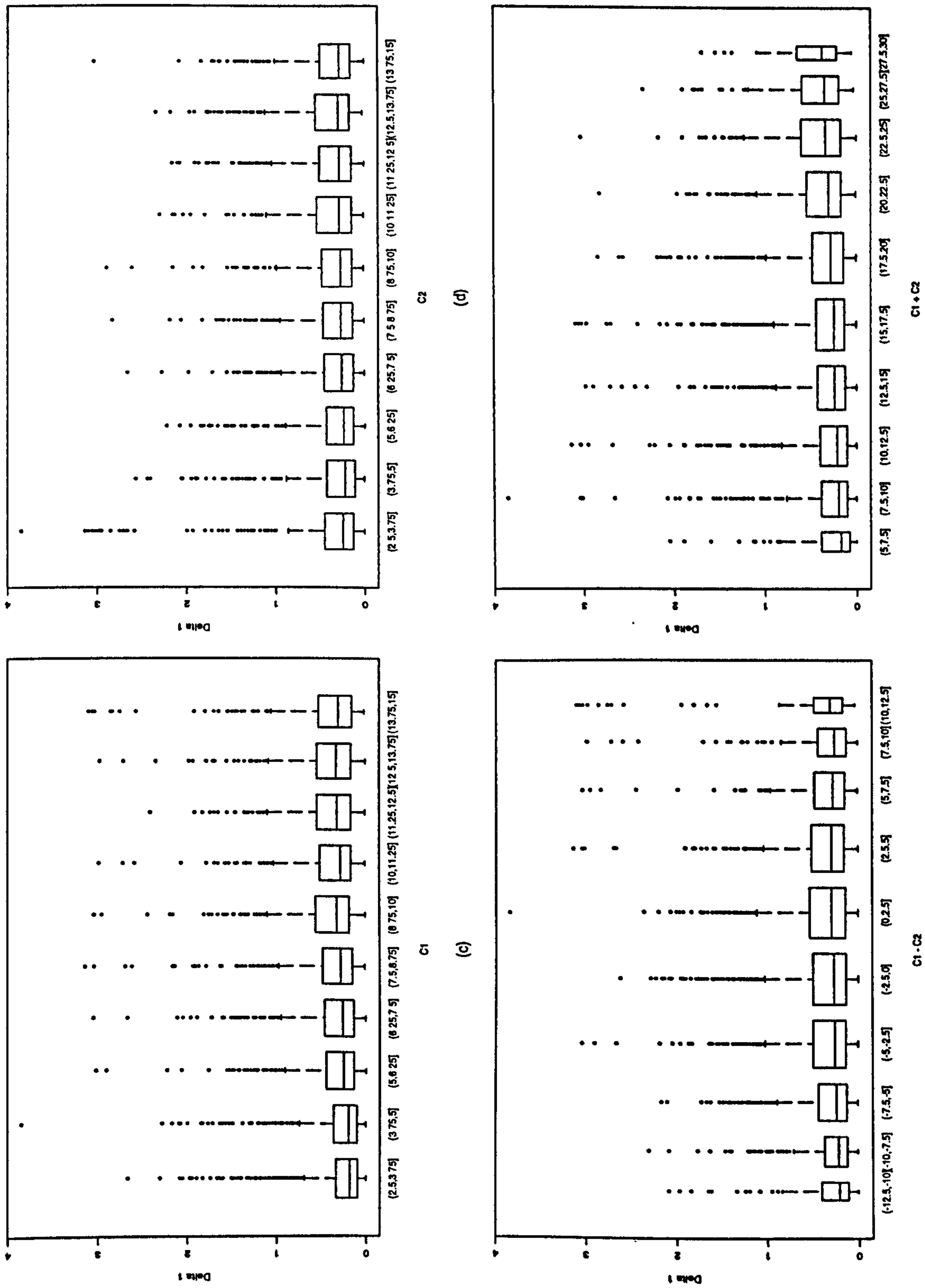


Figure 4.12: Boxplots of Δ_1 for simulated values of (a) c^1, (b) c^2, (c) $c^1 - c^2$, and (d) $c^1 + c^2$.

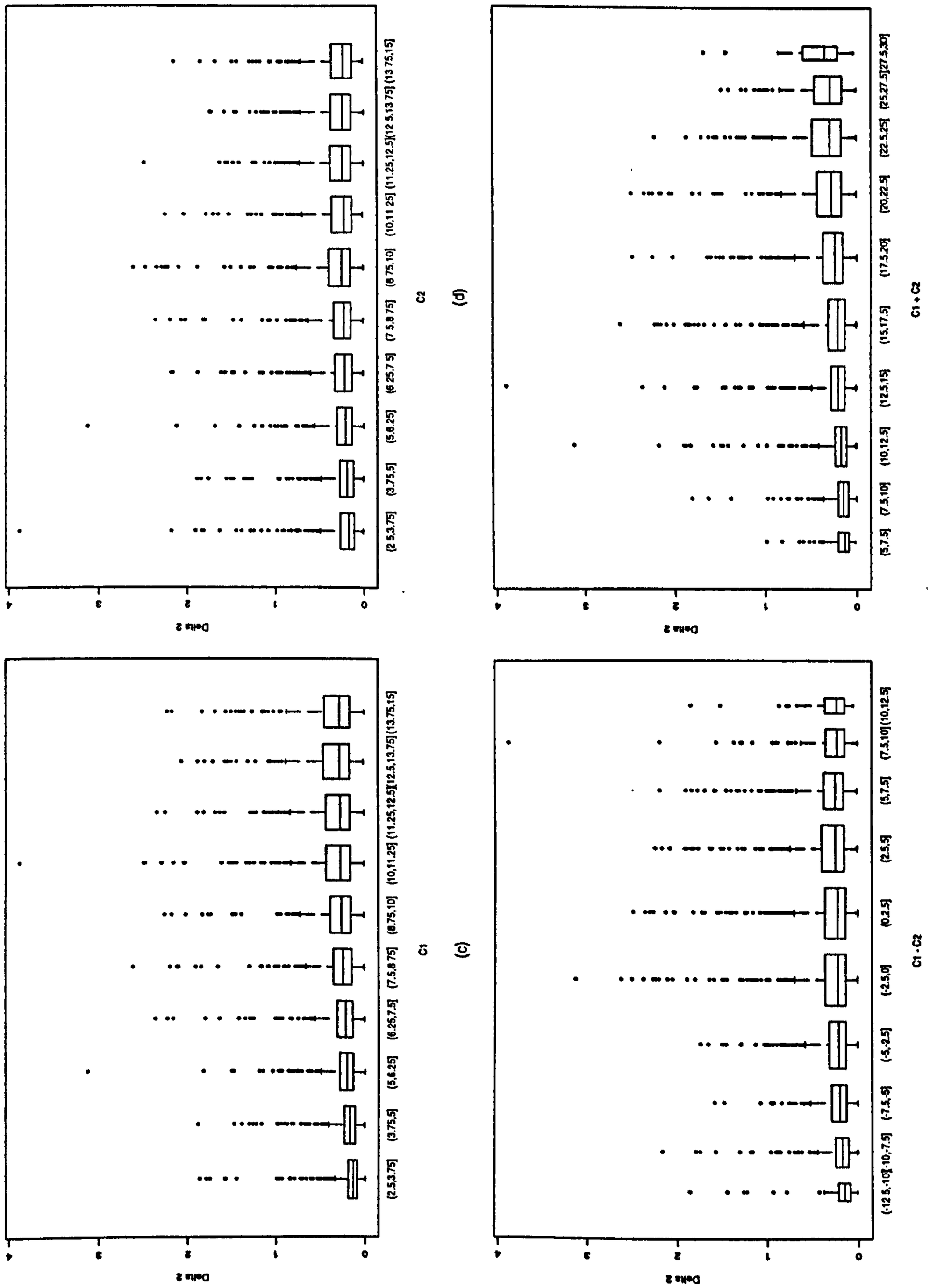


Figure 4.13: Boxplots of Δ_2 for simulated values of (a) c^1, (b) c^2, (c) $c^1 - c^2$, and (d) $c^1 + c^2$.

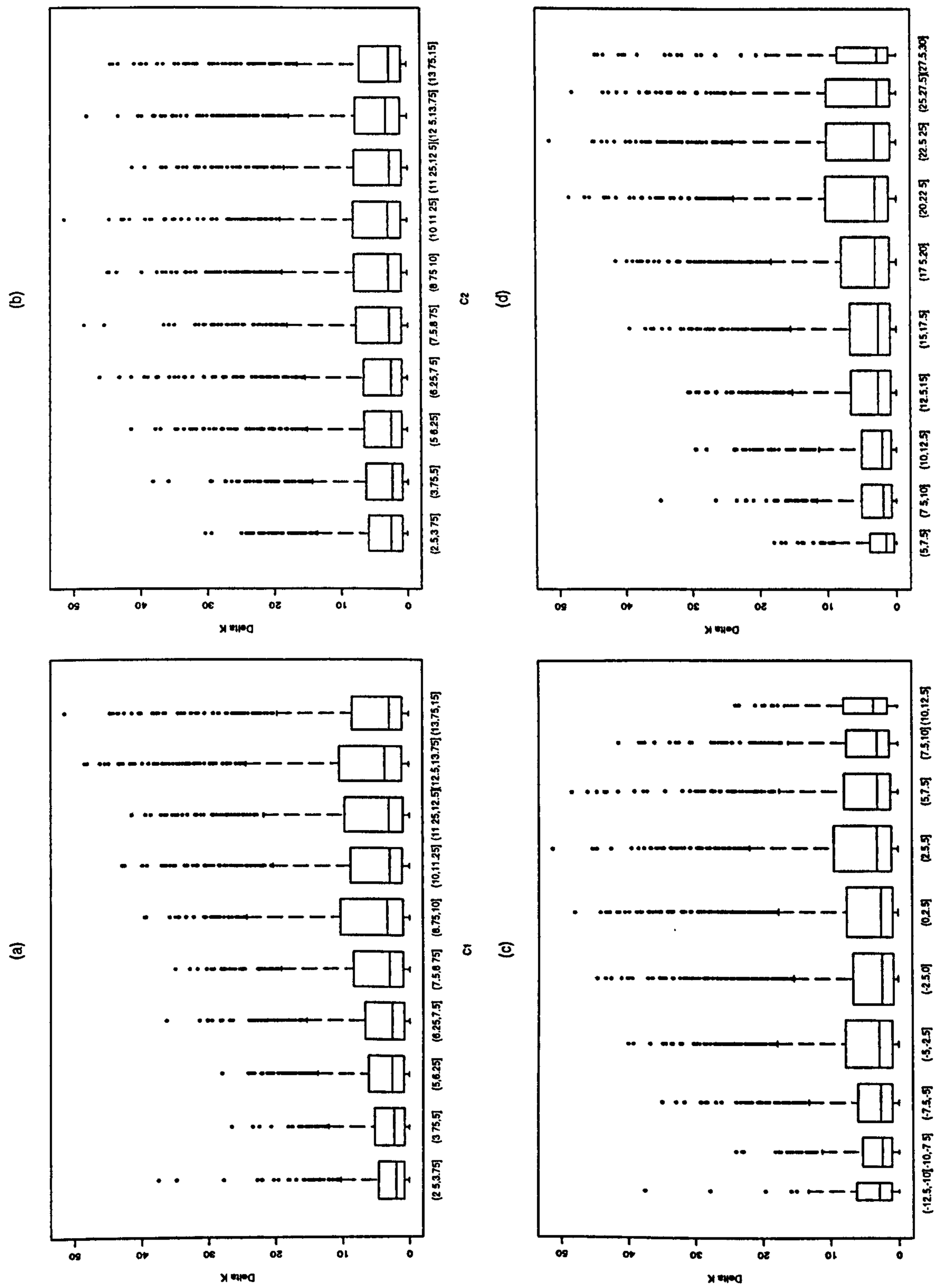


Figure 4.14: Boxplots of $\Delta\mathcal{K}$ for simulated values of (a) c^1, (b) c^2, (c) $c^1 - c^2$, and (d) $c^1 + c^2$.

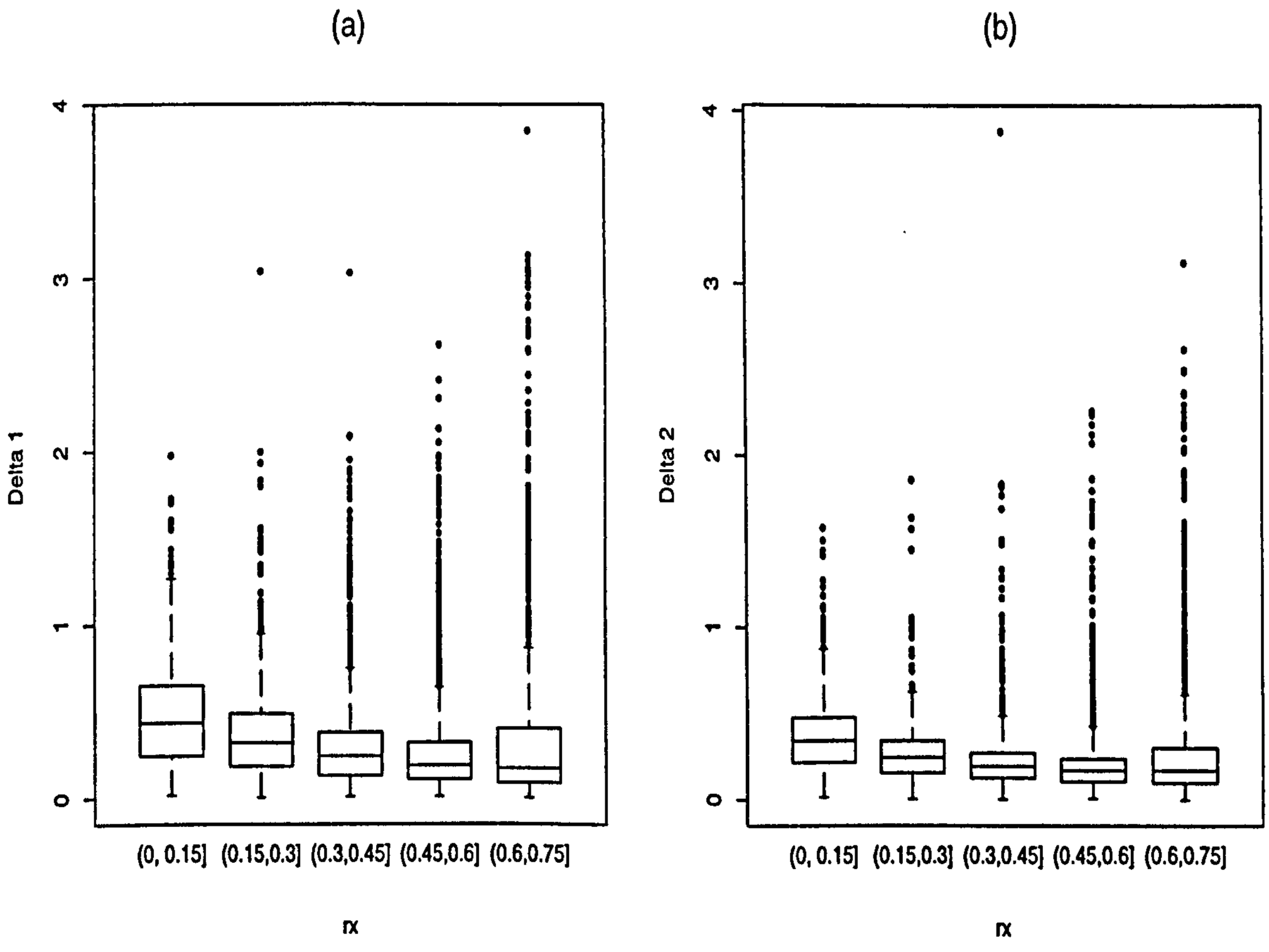


Figure 4.15: Boxplots of (a) Δ_1 and (b) Δ_2 for simulated values of r_x.

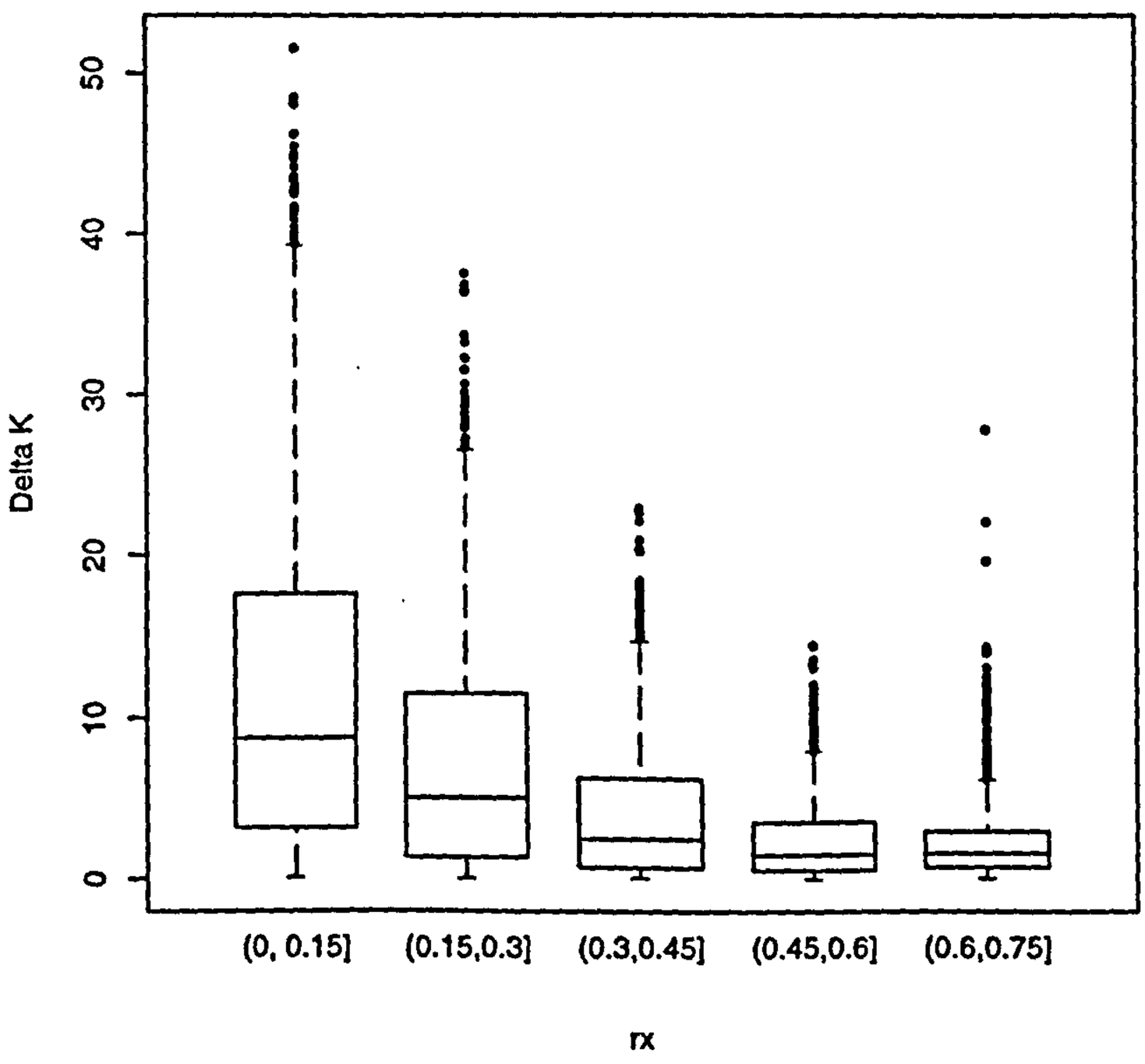


Figure 4.16: Boxplots of $\Delta\mathcal{K}$ for simulated values of r_x.

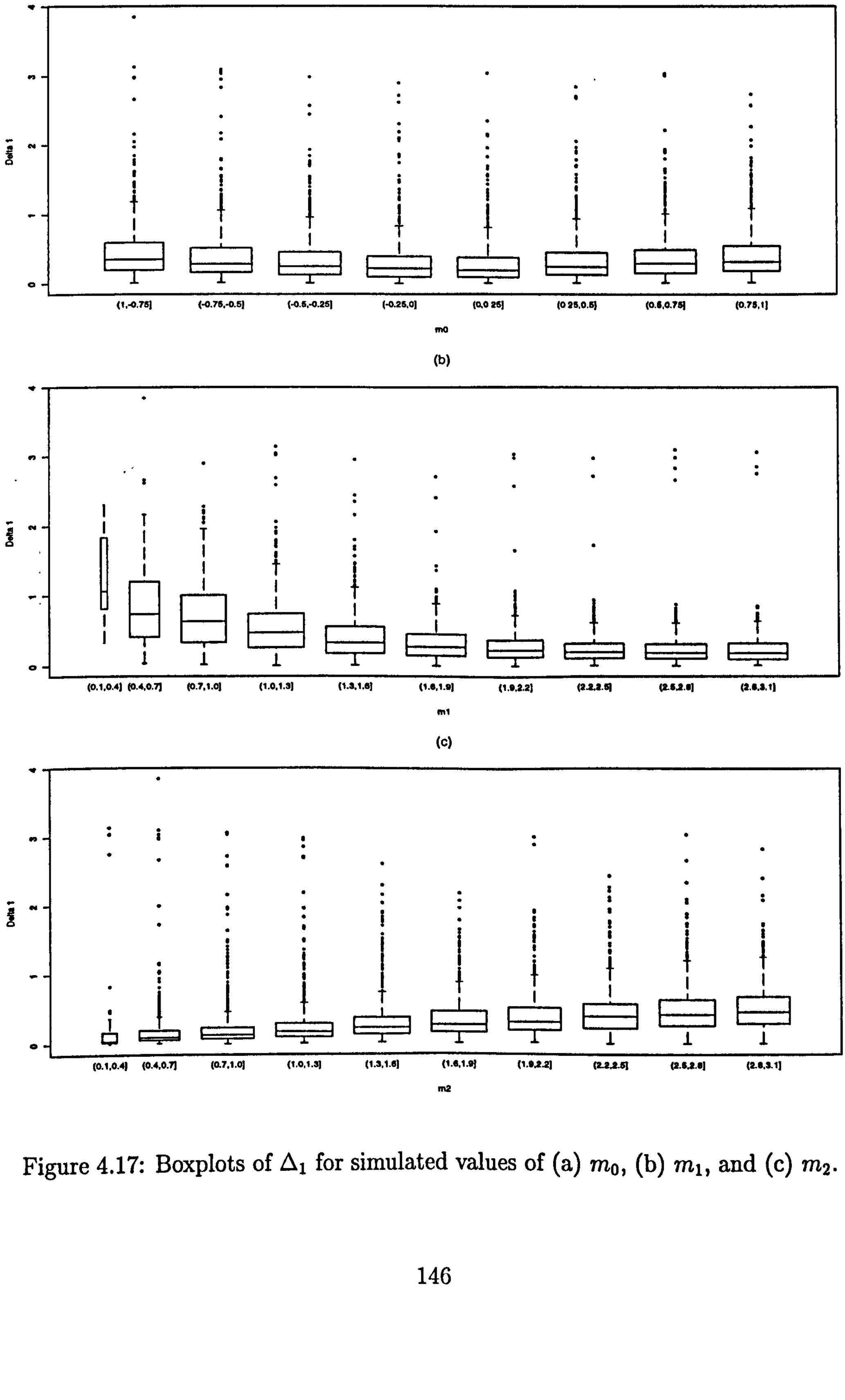


Figure 4.17: Boxplots of Δ_1 for simulated values of (a) m_0, (b) m_1, and (c) m_2.

Figure 4.18: Boxplots of Δ_2 for simulated values of (a) m_0, (b) m_1, and (c) m_2.

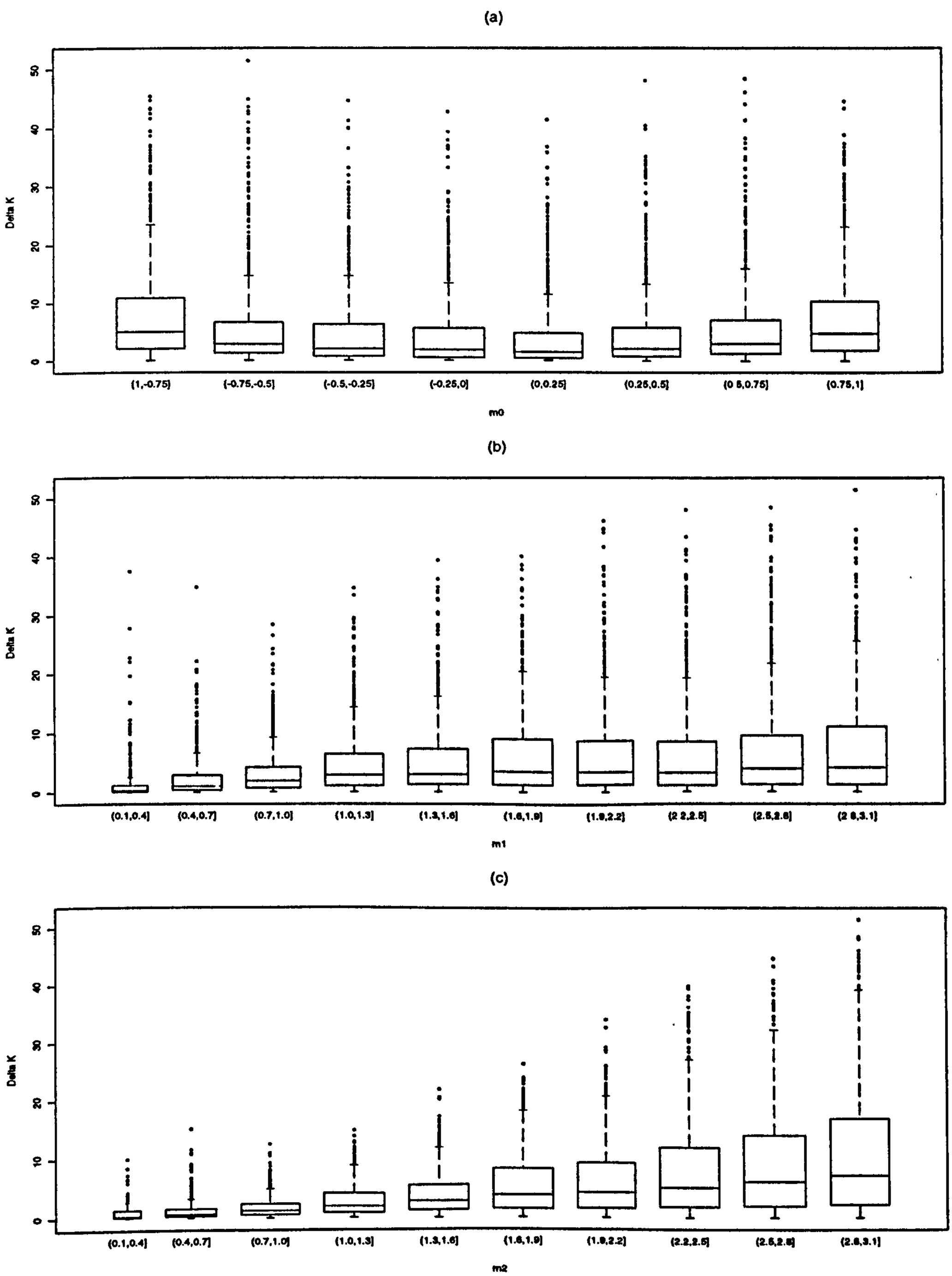


Figure 4.19: Boxplots of $\Delta\mathcal{K}$ for simulated values of (a) m_0, (b) m_1, and (c) m_2.

Comments

1. Clearly the performance of the heuristic is variable, with the discrepancy in Bayes cost over 40% for some problems. However, when the correlation coefficient r_x is above 0.45 the heuristic performs consistently well. As both X_1 and X_2 are, by necessity, strongly correlated to T then it is intuitive that in most cases X_1 will be reasonably highly correlated to X_2.

2. Note that we must take care in inferring from the plots any joint conditions on input parameters under which the heuristic performs well. The plots show the marginal effect of each parameter and it is not clear how these effects interact. However, it is clear that the heuristic performs worst when m_2 is large and r_x is small.

3. When the covariance structure of the regression parameters is as typical and misclassification costs are the same at each stage of the screen, then, intuitively, ordering the stages of the screen by decreasing m_i–value should be optimal. In such a scenario the most discriminative variables are measured first and so items are sentenced by the screen as early as possible in the sequence. In Study 1 the misclassification costs are assumed constant at both covariate stages of the screen and a comparison of the Bayes costs of the fully optimal designs under both permutations of m_1 and m_2 yields an optimal ordering with $m_1 > m_2$ on 93.8% of occasions.

4. The poor performance of the heuristic when the regression parameter m_2 is large is a natural result. It is sensible that the delta approximation of the distribution of X_2 by its expected value in (4.19) will be poorest when X_2 is highly correlated with T, that is when m_2 is larger.

5. Notice also that the differences between the optimal and heuristic design parameters for the second stage of the screen are more consistent and likely to be smaller than the differences at the first stage of the screen.

4.5 Conn's syndrome example

Returning to the numerical example described in section 2.4 and section 3.3.3, we consider the construction of a sequential screen to determine the cause of Conn's syndrome. Section 3.3.3 described the optimal two–stage screen for this example. Two screening variables, denoted X_1 and X_2, are measured at the first stage of that screen, with patients sentenced according to the value of both measurements. A sequential screen will have three stages, at each of the first two stages one of the covariates is measured and patients may be either, (i) said to have a benign tumour ($T = 1$) or a condition of the adrenal glands ($T = 0$), or (ii) passed on to the next stage of the screen. Any patients that remain unsentenced by the covariates are passed on to the third stage of the screen where an exploratory operation determines the cause of Conn's syndrome.

To obtain both the heuristic and fully-optimal sequential screen designs requires computation of Bayes costs, which, in turn, requires the distribution of the screening variable to be specified. In section 3.3.3 we have noted that a bivariate normality assumption is reasonable for $(X_1, X_2)^T$ as given in (3.35). Also X_1 and X_2 are standardised. However, from the data given in Table 3.1 and the subsequent analysis it is clear that X_1 is a smaller–the–better screening variable with $m_1 < 0$. Our method now requires that both covariates are larger–the–better screening variables and so we develop sequential screens based on

$$\begin{aligned} -X_1 &= -\text{standardised log concentration of potassium,} \\ X_2 &= \text{standardised log concentration of carbon dioxide.} \end{aligned}$$

The data estimates for the regression parameters, $\underline{\xi}^-$, under a probit model for $T|-X_1, X_2$ can easily be recovered from those for the $T|X_1, X_2$ model given in section 3.3.3, as

$$\begin{aligned} P(T = 1|-X_1, X_2, \underline{\xi}^-) &= \Phi(\xi_0^- - \xi_1^- X_1 + \xi_2^- X_2) \\ &= \Phi(\xi_0 + \xi_1 X_1 + \xi_2 X_2) = P(T = 1|X_1, X_2, \underline{\xi}), \end{aligned}$$

by setting $\xi_0^- = \xi_0$, $\xi_1^- = -\xi_1$ and $\xi_2^- = \xi_2$. Hence the moments of ξ_0^- and ξ_2^- will be the same as those of ξ_0 and ξ_2 and $E(\xi_1^-) = -E(\xi_1)$, $Var(\xi_1^-) = Var(\xi_1)$, $Cov(\xi_0, \xi_1^-) = -Cov(\xi_0, \xi_1)$ and $Cov(\xi_1^-, \xi_2) = -Cov(\xi_1, \xi_2)$. Hence, from (3.36), the data give

$$\underline{m}^- = \begin{pmatrix} 1.241 \\ 1.576 \\ 0.953 \end{pmatrix} \quad \text{and} \quad S^- = \begin{pmatrix} 0.3030 & 0.2202 & 0.1468 \\ 0.2202 & 0.4874 & -0.0254 \\ 0.1468 & -0.0254 & 0.3129 \end{pmatrix}.$$

Making a further simple adjustment to the model used in section 3.3.3, we assume that the distribution of $(-X_1, X_2)^T$ is bivariate normal with correlation coefficient given by the sample value, $r_x = 0.6783$. The cost structure is as before, with the costs of misclassification the same at both covariate stages of the screen, that is,

$$c_r^1 = c_r^2 = c_r, \qquad c_a^1 = c_a^2 = c_a = 3c_r/4, \qquad \text{and} \qquad c_m = 3c_r/40.$$

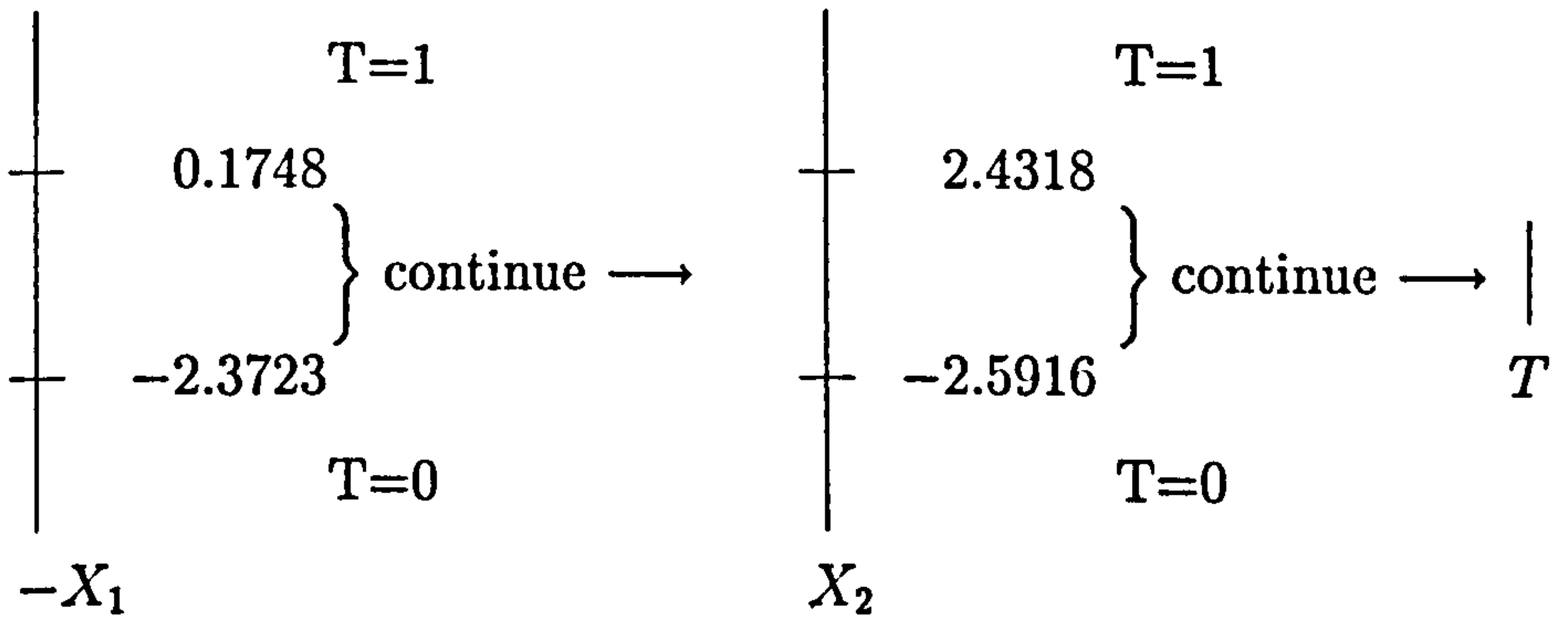
The costs of measuring the covariates are assumed negligible and so we set $c_s^1 = c_s^2 = 0$. Now we compare and contrast the heuristic and fully optimal sequential screens under the two orderings of the covariate stages of the screen.

Ordering 1

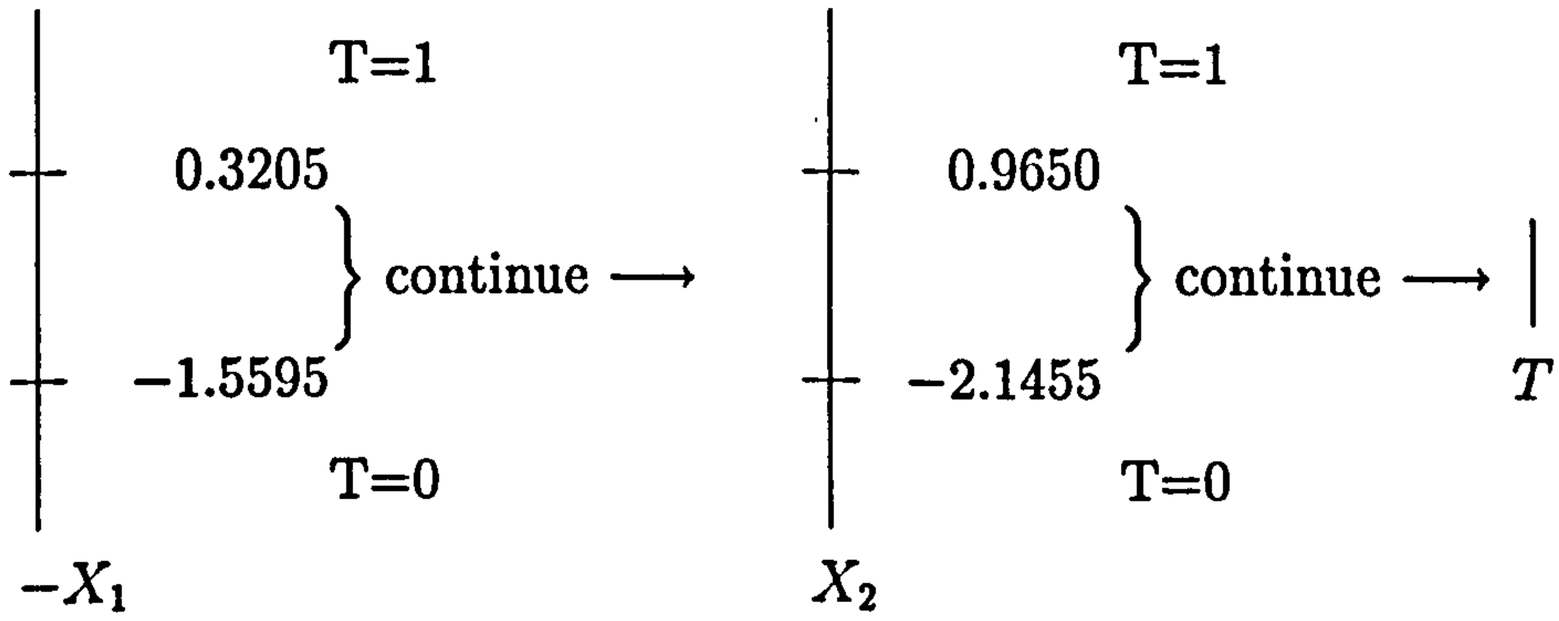
First we fix the ordering of the covariates so that the X_1-stage comes before the X_2-stage. Submitting the model parameters given above to the algorithm that determines the heuristic design of the screen, as given in section 4.4.2, the iterative procedure gives,

Iteration	Design parameters				Bayes cost
Number	v_2''	w_2''	v_1''	w_1''	$\mathcal{K}(\underline{v}'', \underline{w}'')$
1	$-\infty$	0.2997	−2.5051	0.2352	0.06050
2	−2.7080	2.3064	−2.3712	0.1743	0.05584
3	−2.5906	2.4329	−2.3723	0.1748	0.05579
4	−2.5916	2.4318	−2.3723	0.1748	0.05579

where the initial parameters for the first stage of the screen were specified as $v_1^0 = -1.0$ and $w_1^0 = 1.0$, and $-\infty$ denotes that $\hat{\mathcal{K}}^{(2)}$ in (4.15) was minimised in the limit as $v_2 \to -\infty$. The convergence criterion of the heuristic scheme was chosen to match that imposed throughout the simulation study of the previous section. The heuristic design at the final iteration is depicted by,



and has Bayes cost $\mathcal{K} = 0.05579c_r$. Using numerical methods, the fully optimal sequential screen under this ordering of covariates is



and has Bayes cost $\mathcal{K} = 0.05342c_r$. Comparing the two screens, the screening parameters $(\underline{v}, \underline{w})$ are appreciably different for both covariate stages of the screen, but the Bayes cost of the heuristic design is only an increase of 4.44% over the Bayes cost of the fully optimal design. Table 4.1 shows how the cases used in the design of the screen are classified by the covariates. Those cases remaining unclassified are passed on to the final stage of the screen. The fully optimal design performs better than the heuristic, correctly classifying two more patients. However, notice that at the first stage, the heuristic screen correctly classifies at least one of the cases with $T = 1$ that the fully–optimal screen fails to classify. No cases were misclassified.

Table 4.1: Screen classifications of the sequential screen – ordering 1

		Heuristic design				Optimal design			
		1st stage		2nd stage		1st stage		2nd stage	
Actual	Total	$T=1$	$T=0$	$T=1$	$T=0$	$T=1$	$T=0$	$T=1$	$T=0$
$T=1$	20	13	0	0	0	11	0	1	0
$T=0$	11	0	0	0	0	0	3	0	1

Ordering 2

Under the ordering of X_2-stage then X_1-stage, the iterates of the heuristic design are:

Iteration	Design parameters				Bayes cost
Number	v_2''	w_2''	v_1''	w_1''	$\mathcal{K}(\underline{v}'', \underline{w}'')$
1	−2.1006	0.1602	$-\infty$	$+\infty$	0.05490
2	−2.1006	0.1602	$-\infty$	$+\infty$	0.05490

where the initial parameters for the first stage of the screen are again $v_1^0 = -1.0$ and $w_1^0 = 1.0$, and $\pm\infty$ denotes a minimisation in the appropriate limit. Note that here v_1'' and w_1'' are parameters of the X_2-stage of the screen and v_2'' and w_2'' are parameters of the X_1-stage. The heuristic scheme reaches convergence immediately but it converges to a screen in which the covariate X_2 is never measured. Hence the two–stage screen below gives the heuristic design for this ordering of the screens,

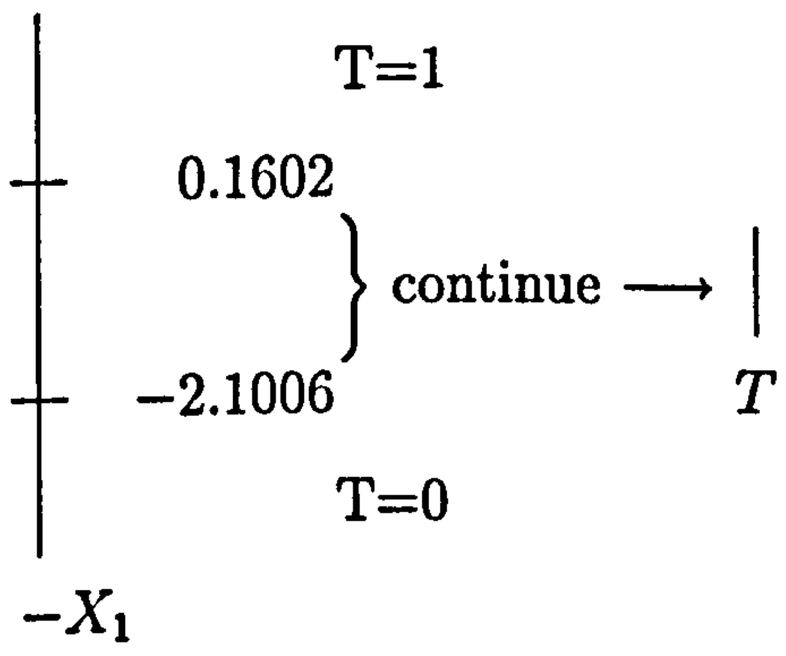
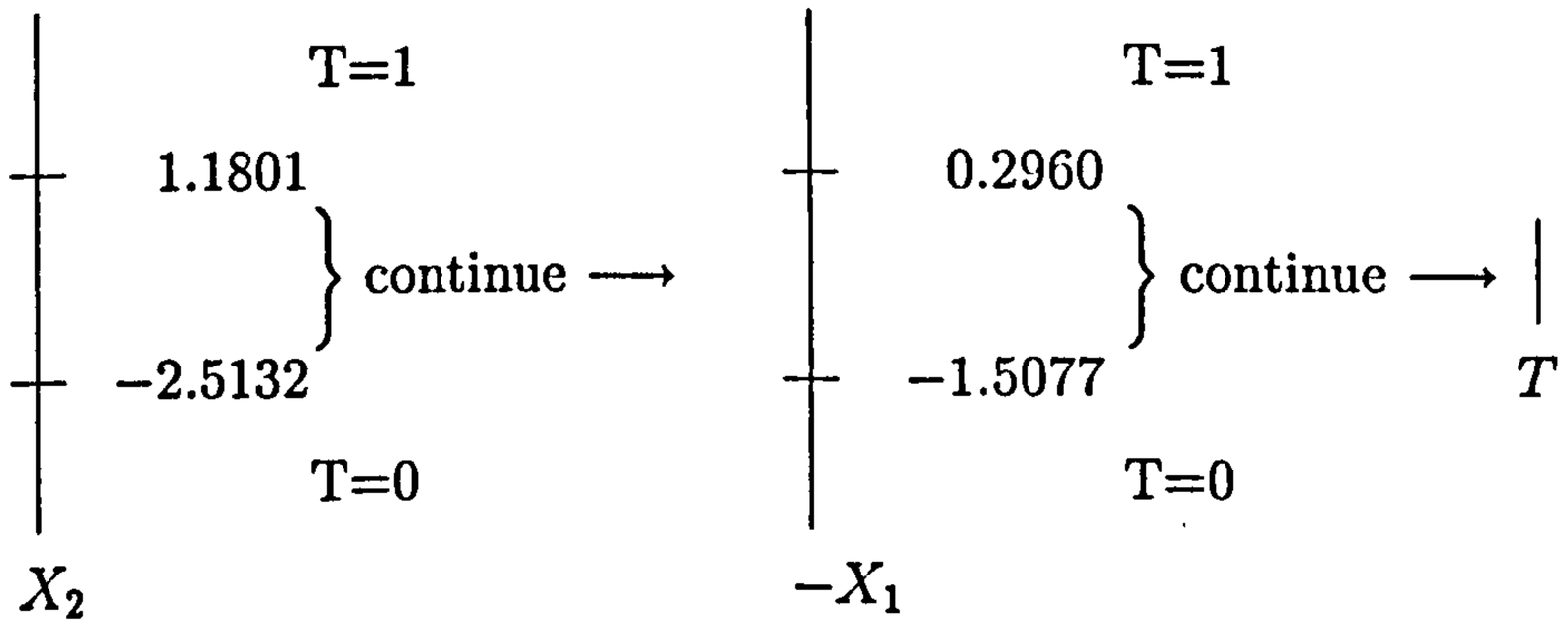


Table 4.2: Screen classifications of a sequential screen – ordering 2

		Heuristic design				Optimal design			
		1st stage		2nd stage		1st stage		2nd stage	
Actual	Total	$T=1$	$T=0$	$T=1$	$T=0$	$T=1$	$T=0$	$T=1$	$T=0$
$T=1$	20	–	–	13	0	6	0	5	0
$T=0$	11	–	–	0	0	0	0	0	3

This screen has Bayes cost $\mathcal{K} = 0.0549c_r$. The fully optimal sequential screen for this ordering is not degenerate in the same way as the heuristic. It is given by



and has Bayes cost $\mathcal{K} = 0.0528c_r$. Again, the parameters that characterise the screen are appreciably different for both covariate stages of the screen, in fact, the heuristic recommends that an X_2-stage should not come before an X_1-stage, contrary to the fully optimal design. However, comparing Bayes costs, the cost of the heuristic design is an increase of only 3.97% on the fully–optimal design. Table 4.2 shows the classification of the case data under both screens. Again the optimal design performs marginally better in terms of numbers correctly classified but the heuristic design correctly classifies at least two cases which the optimal fails to classify. No cases are wrongly categorized.

Further comparisons

Clearly, for both the heuristic and fully-optimal designs, the second ordering of the covariates is Bayes optimal. However, notice that the fully optimal design under the

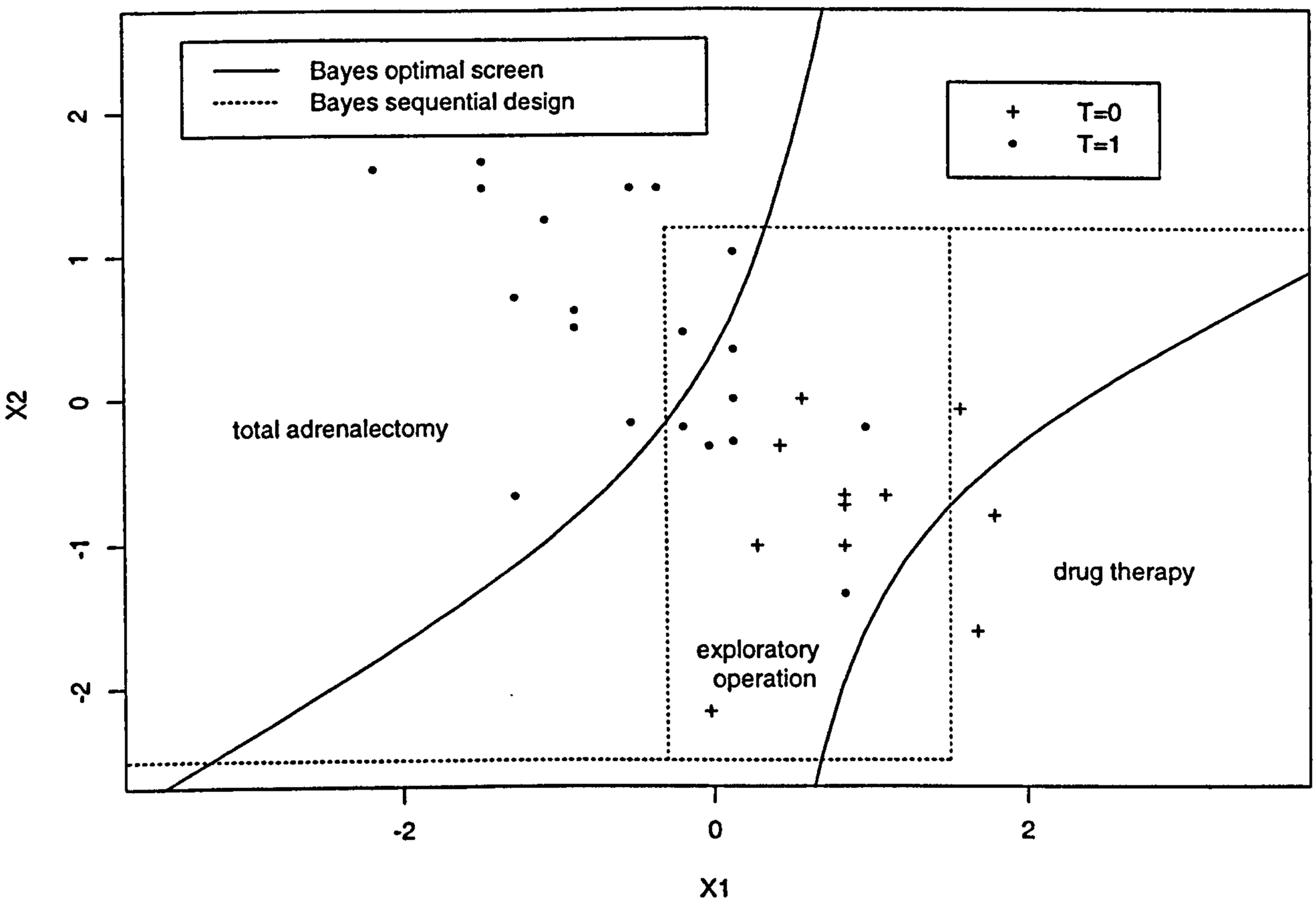


Figure 4.20: Plot of Bayes optimal batch and sequential screens

first ordering correctly classifies more of the sample cases than its counterpart when ordering is reversed.

Comparing the optimal sequential screen with a Bayes two–stage screen described in section 3.3.3, the Bayes cost of $0.0528c_r$ is an increase of 8.2% on the Bayes cost of the optimal two–stage screen. This seems intuitive, as in this chapter we impose a form to the regions of $\underline{X}$ that make up the screen and optimise over all regions of that form. In Chapter 3 we optimise assuming no fixed form for the screen. Figure 4.20 displays the optimal sequential screen and the Bayes two–stage screen. Here the cost of measuring covariates is assumed negligible and the optimal two–stage screen is preferable to a sequential screen. This may not be the case in general, a sequential screen will be more competitive when the cost of measuring covariates at

the later stages of the screen is substantial. In such cases, there will be some benefit to classifying items or patients as early as possible in the screen, and so measuring covariates in the latter part of the screen less often. In the next chapter we look at the effect of reducing the number of covariates in a one–stage screen.

4.6 Further Comments

Here we briefly discuss a few further observations but, in the main, refer the reader to the discussion in section 2.3 which covers many points relevant to this chapter.

1. **Interaction terms.** In section 4.3 we assume a model for $T|\underline{X}$ expressed via a probit link function which is linear in the screening variables $\underline{X}$ as well as linear in the regression parameters $\underline{\xi}$. This does not allow for the inclusion of any interaction term that may be important in the model for $T|\underline{X}$. When there are only two screening variables, as there are in section 4.3, including an interaction term results in the probit regression model

$$P\left(T=1|\underline{x},\underline{\xi}\right)=\Phi\left(\xi_0+\xi_1x_1+\xi_2x_2+\xi_3x_1x_2\right).$$

If such a term significantly improves the fit of the model then we can easily adapt our method accordingly.

In the heuristic of section 4.3 we design the second screen by making the approximation that $X_1=\mu_1'$ given in (4.14) and show that the minimisation problem can be recast as a single screen problem as in Chapter 2. To adapt to a model that includes an interaction term the only change needed here is that the probit regression of T on X_2 given that $X_1=\mu_1'$ now has parameters $\underline{\xi}^*=(\xi_0^*,\xi_1^*)$ where $\xi_0^*=\xi_0+\mu_1'\xi_1$ and $\xi_1^*=\xi_2+\mu_1'\xi_3$ as

$$\begin{aligned}P(T=1|\underline{\xi},X_1=\mu_1',x_2) &= \Phi\left(\xi_0+\xi_1\mu_1'+\xi_2x_2+\xi_3\mu_1'x_2\right)\\ &= \Phi\left\{\left(\xi_0+\xi_1\mu_1'\right)+\left(\xi_2+\xi_3\mu_1'\right)x_2\right\}\\ &= \Phi\left(\xi_0^*+\xi_1^*x_2\right)\\ &= P(T=1|\underline{\xi}^*,x_2).\end{aligned}$$

The moments of (ξ_0^*, ξ_1^*) can easily be calculated from those of $\underline{\xi}$.

Similarly, in the design of the first screen we make the approximation that $X_2 = \mu_2' = E(X_2)$ and reduce to a one screen problem. For the case in which X_2 is standardised with mean zero, the inclusion of an interaction term in the probit model for $T|X_1, X_2$ leaves the parameters of the model for T on X_1 given $X_2 = 0$ unchanged, since

$$P(T = 1|\underline{\xi}, x_1, X_2 = 0) = \Phi(\xi_0 + \xi_1 x_1 + \xi_2 \times 0 + \xi_3 x_1 \times 0) = \Phi(\xi_0 + \xi_1 x_1).$$

Further computational studies could be carried out to determine the importance of any interaction term and its effect on the performance of our heuristic.

2. In practice, when an item reaches the ith screen having been unsentenced by the first $i-1$ screens, the measurements of the first $i-1$ screening variables for that item are available. Hence, whenever it is practical to do so, the remaining $p-i+1$ screens can be redesigned conditional on these known values. While less practical than a scheme in which screen designs remain fixed, such a method would make use of all the available information at each stage.

Chapter 5

Dimensionality reduction in screen design

5.1 Introduction

Recall that in a typical screening scenario, a p-dimensional screening variable $\underline{X}$ is used to screen for attributes described by a performance variable $\underline{T}$ such that $\underline{T} \in C_{\underline{T}}$. So far we have largely ignored the possibility that, in cases when the cost of measuring covariates is not negligible, it may be optimal to reduce the number of covariates measured when screening for $\underline{T}$. In Chapter 3, for a two–stage screen, we assumed the case in which all (or none) of the covariates are measured in the screen and in Chapter 4, in the context of a sequential screen, we imposed a screening cost for measuring each covariate but assumed that such costs were negligible in the simulation study that looked at screcn designs. Here we impose a cost for performing a screen, and look at the problem of choosing

(a) the number of covariates to use in the screen, and

(b) which covariates to use in the screen,

in the context of the simple screening scenario of a one–stage screen. In a one–stage screen all items are sentenced as acceptable or unacceptable based on a measurement of the screening variable $\underline{X}$. Such a screen is appropriate when the performance variable $\underline{T}$ is measured via destructive testing or is considered too expensive to measure, for example, when $c_m > c_a c_r/(c_a + c_r)$ in Chapter 2 and Chapter 3. In the case when T is binary and $T|\underline{X}$ is modelled by a probit regression, we propose a variety of heuristic approaches for choosing screening components and assess the performance of the heuristics with a numerical study.

Under the standard case of costs, as defined in Chapter 3, section 5.2 recovers the Bayes optimal one–stage screen for a general model on $(\underline{T}, \underline{X})$. The result is taken from Turkman & Amaral Turkman (1989) and is a special case of Theorem 3.1. Turkman & Amaral Turkman give an example in which $(\underline{T}, \underline{X})$ is multivariate normal with a conjugate prior for the model parameters. Here we take the case in which T is binary and $T|\underline{X}$ is modelled by a probit regression model. Section 5.3 discusses the construction of screen designs that use $d(< p)$ components of the screening variable, and reviews the problem of choosing how many and which covariates should be used in the screen. In the context of the probit regression model, we suggest some heuristic approaches to choosing screening components in section 5.4. We suppose that the cost of measuring each covariate is the same and base our heuristics on model parameters. Section 5.5 investigates the performance of the heuristics in a numerical study.

5.2 Optimal screening regions

Suppose we have a performance variable $\underline{T}$ which defines whether an item is acceptable ($\underline{T} \in C_{\underline{T}}$) or unacceptable ($\underline{T} \notin C_{\underline{T}}$). Suppose also that a one–stage screen is to be based on a p-dimensional screening variable $\underline{X}^{(p)} = (X_1, X_2, \ldots, X_p)$, with sample space $\Omega_{\underline{X}^{(p)}}$, which advocates the acceptability of an item if $\underline{X}^{(p)} \in C_{\underline{X}^{(p)}}$. We shall assume, without loss of generality, that the screening variable has been standardised so that its components each have zero mean and unit variance. As the rationale is

to predict whether $\underline{T}$ lies in $C_{\underline{T}}$ using $\underline{X}^{(p)}$ it is sensible to focus on the relationship between the performance and screening variable through the conditional model $\underline{T}|\underline{X}^{(p)},\underline{\xi}$ where $\underline{\xi}$ are unknown parameters with distribution $\pi(\underline{\xi})$. The distribution $\pi(\underline{\xi})$ will typically be a posterior distribution based on a training sample on $(\underline{T},\underline{X}^{(p)})$ and any prior information about the relationship between $\underline{T}$ and $\underline{X}^{(p)}$.

The costs of misclassification by the screen are as in the standard case of Chapter 3, with a cost of c_a for wrongly retaining a defective item, and c_r for rejecting a satisfactory item. Notice that as $\underline{T}$ will never be measured there is no cost c_m. The cost of operating the screen will depend on how many (and possibly which) components or elements of $\underline{X}^{(p)}$ are used. We write this cost as $c_s(p)$, making explicit the dependence on the dimension of the screening variable.

The expected cost of using a one–stage screen with acceptance region $C_{\underline{X}^{(p)}}$ is

$$\begin{aligned}\mathcal{K}\{C_{\underline{X}^{(p)}}\} &= c_r P(\text{reject a good item}) + c_a P(\text{accept a bad item}) + c_s(p) \\ &= c_r P(\underline{T} \in C_{\underline{T}}, \underline{X}^{(p)} \notin C_{\underline{X}^{(p)}}) + c_a P(\underline{T} \notin C_{\underline{T}}, \underline{X}^{(p)} \in C_{\underline{X}^{(p)}}) + c_s(p).\end{aligned}$$

Expressing this explicitly in terms of a model for $\underline{T}|\underline{X}^{(p)},\underline{\xi}$ the Bayes cost of using the screen is

$$\begin{aligned}\mathcal{K}\{C_{\underline{X}^{(p)}}\} = E_{\underline{X}^{(p)}} \Big[& c_r E_{\underline{\xi}} \left\{ P(\underline{T} \in C_{\underline{T}}|\underline{X}^{(p)},\underline{\xi}) \right\} I(\underline{X}^{(p)} \notin C_{\underline{X}^{(p)}}) \\ & + c_a E_{\underline{\xi}} \left\{ P(\underline{T} \notin C_{\underline{T}}|\underline{X}^{(p)},\underline{\xi}) \right\} I(\underline{X}^{(p)} \in C_{\underline{X}^{(p)}}) \Big] + c_s(p). \quad (5.1)\end{aligned}$$

where

$$I(\underline{X}^{(p)} \in A) = \begin{cases} 1, & \underline{X}^{(p)} \in A, \\ 0, & \text{otherwise.} \end{cases}$$

Definition 5.1 A *Bayes one–stage screen* $C^*_{\underline{X}^{(p)}}$ is any subset of $\Omega_{\underline{X}^{(p)}}$ satisfying

$$\mathcal{K}\{C^*_{\underline{X}^{(p)}}\} = \inf_{C_{\underline{X}^{(p)}}} \mathcal{K}\{C_{\underline{X}^{(p)}}\}, \quad (5.2)$$

the infimum in (5.2) being taken over all subsets of $\Omega_{\underline{X}^{(p)}}$.

The following corollary is a direct consequence of Turkman & Amaral Turkman (1989) and is also recoverable as a special case of Theorem 3.1 given in section 3.2.

For illustrative purposes we give two short proofs, the first follows Turkman & Amaral Turkman and the second follows the proof of Theorem 3.1. However, we stress that the following conjecture is merely a corollary, easily inferred from results proved elsewhere. We give proofs to illustrate how we arrive at the solution to our formulation of the screening problem. We use the convention that, when there is an arbitrary choice between accepting or rejecting an item in terms of Bayes cost then the item will be accepted.

Corollary 5.1 A *Bayes one-stage screen* is given by

$$C^*_{\underline{X}^{(p)}} = \left[\underline{x}^{(p)} \in \Omega_{\underline{X}^{(p)}} : E_{\underline{\xi}}\left\{P(\underline{T} \in C_{\underline{T}} | \underline{x}^{(p)}, \underline{\xi})\right\} \geq \frac{c_a}{c_a + c_r}\right]. \qquad (5.3)$$

Proof 1 (Turkman & Amaral Turkman)

The Bayes cost in (5.1) can be rewritten as

$$\begin{aligned} \mathcal{K}\{C_{\underline{X}^{(p)}}\} &= E_{\underline{X}^{(p)}}\left\{\tilde{c}_r(\underline{X}^{(p)}) I(\underline{X}^{(p)} \notin C_{\underline{X}^{(p)}})\right. \\ &\quad \left. + \tilde{c}_a(\underline{X}^{(p)}) I(\underline{X}^{(p)} \in C_{\underline{X}^{(p)}})\right\} + c_s(p), \qquad (5.4) \end{aligned}$$

where

$$\begin{aligned} \tilde{c}_r(\underline{x}^{(p)}) &= c_r E_{\underline{\xi}}\left\{P(\underline{T} \in C_{\underline{T}} | \underline{X}^{(p)}, \underline{\xi})\right\} \\ \tilde{c}_a(\underline{x}^{(p)}) &= c_a E_{\underline{\xi}}\left\{P(\underline{T} \notin C_{\underline{T}} | \underline{X}^{(p)}, \underline{\xi})\right\}. \end{aligned}$$

As $I(\underline{X}^{(p)} \notin C_{\underline{X}^{(p)}}) = 1 - I(\underline{X}^{(p)} \in C_{\underline{X}^{(p)}})$,

$$\mathcal{K}\{C_{\underline{X}^{(p)}}\} = E_{\underline{X}^{(p)}}\left[\tilde{c}_r(\underline{X}^{(p)}) + \left\{\tilde{c}_a(\underline{X}^{(p)}) - \tilde{c}_r(\underline{X}^{(p)})\right\} I(\underline{X}^{(p)} \in C_{\underline{X}^{(p)}})\right] + c_s(p).$$

Here it is easy to see that $\mathcal{K}\{C_{\underline{X}^{(p)}}\}$ is minimised when the second term above is negative for all $\underline{X}^{(p)} \in C_{\underline{X}^{(p)}}$, and hence

$$C^*_{\underline{X}^{(p)}} = \left\{\underline{x}^{(p)} \in \Omega_{\underline{X}^{(p)}} : \tilde{c}_a(\underline{x}^{(p)}) - \tilde{c}_r(\underline{x}^{(p)}) \leq 0\right\}.$$

The result follows from the definitions of $\tilde{c}_a(\underline{x}^{(p)})$ and $\tilde{c}_r(\underline{x}^{(p)})$ with a little algebra. □

Proof 2 (Theorem 3.1)

From equation (5.4) it follows that

$$\mathcal{K}\{C_{\underline{X}^{(p)}}\} \geq E_{\underline{X}^{(p)}}\left[\min\left\{\tilde{c}_r(\underline{X}^{(p)}), \tilde{c}_a(\underline{X}^{(p)})\right\}\right] + c_s(p).$$

The bound is clearly attained by the one–stage screen,

$$C^*_{\underline{X}^{(p)}} = \left[\underline{x}^{(p)} \in \Omega_{\underline{X}^{(p)}} : \tilde{c}_a(\underline{x}^{(p)}) = \min\left\{\tilde{c}_r(\underline{x}^{(p)}), \tilde{c}_a(\underline{x}^{(p)})\right\}\right]$$

and the result follows with a little manipulation. □

Note that the Bayes one–stage screen given by (5.3) has no dependence on the distribution of the screening variable $\underline{X}^{(p)}$ or the screening cost $c_s(p)$. As in Chapter 3, a more general result follows when the misclassification costs c_a and c_r are allowed to be functions of $\underline{T}$ and $\underline{X}^{(p)}$.

Calculation of the Bayes acceptance region is far from trivial and involves computationally intensive methods with many choices of model for $\underline{T}|\underline{X}^{(p)}, \underline{\xi}$ and $\underline{\xi}$. To find a simple and robust solution to this problem we follow earlier chapters which consider a binary performance variable T (which takes the value 1 when an item is acceptable and 0 otherwise) and a linear probit regression for $T|\underline{X}^{(p)}, \underline{\xi}$:

$$P(T = 1|\underline{X}^{(p)} = \underline{x}^{(p)}, \underline{\xi}) = \Phi\left(\underline{\xi}^T \underline{x}_0^{(p)}\right)$$

where $\underline{x}_0^{(p)} = (1, \underline{x}^{(p)^T})^T$.

As in earlier chapters, the choice of a probit link function is motivated by the fact that $E_{\underline{\xi}}[\Phi(\underline{\xi}^T \underline{x}_0^{(p)})]$ has a closed form when $\underline{\xi}$ is assumed to have a $N_{p+1}(\underline{m}, \boldsymbol{S})$ distribution. This is the case when the posterior distribution of $\underline{\xi}$ can be approximated by its asymptotic normal form, where $\underline{m}$ is the maximum likelihood estimate of $\underline{\xi}$ and $\boldsymbol{S}$ is the inverse of Fisher's information matrix evaluated at $\underline{m}$. For this model, given $\underline{x}_0^{(p)}$, following Lemma 3.2

$$E_{\underline{\xi}}\left[\Phi(\underline{\xi}^T \underline{x}_0^{(p)})\right] = \Phi\left\{\frac{\underline{m}^T \underline{x}_0^{(p)}}{\left(1 + \underline{x}_0^{(p)^T} \boldsymbol{S} \underline{x}_0^{(p)}\right)^{1/2}}\right\}$$

Thus under this model, the Bayes acceptance region is

$$C^*_{\underline{X}^{(p)}} = \left\{ \underline{x}^{(p)} \in \mathbb{R}^p : \frac{\underline{m}^T \underline{x}_0^{(p)}}{(1 + \underline{x}_0^{(p)T} \boldsymbol{S} \underline{x}_0^{(p)})^{1/2}} \geq \Phi^{-1}\left(\frac{c_a}{c_a + c_r}\right) \right\}. \tag{5.5}$$

Advantages of this solution are that it is simple to calculate and interpret. Also Boys and Glazebrook (1992) show that for the $p = 1$ case this solution is robust to modest departures from modelling assumptions concerning the link function and the distribution of $\underline{\xi}$. Clearly cases with $p > 1$ will inherit such robustness properties.

5.3 Reduction in dimensionality

In many situations there is a significant price to pay for measuring each component of the screening variable. It may be cheaper to operate a d–dimensional screen ($d < p$) and save some of the screening cost. In this section we investigate the calculation of marginal Bayes costs and explore various strategies for determining which covariates to use in the screen.

Consider a partition of the p–dimensional screening variable into two parts, one d–dimensional, corresponding to the components to remain as part of the screen, and the other $(p-d)$–dimensional, corresponding to the components to be omitted from the screen: $\underline{X}^{(p)} \equiv (\underline{X}^{(d)}, \underline{X}^{(p-d)})$. Our main task is to find a d–dimensional one–stage screen $C_{\underline{X}^{(d)}}$ that minimises Bayes cost

$$\begin{aligned} \mathcal{K}\{C_{\underline{X}^{(d)}}\} &= E_{\underline{X}^{(d)}}\left[c_r E_{\underline{X}^{(p-d)}|\underline{X}^{(d)}}\left\{E_{\underline{\xi}}\left[P(\underline{T} \in C_{\underline{T}}|\underline{X}^{(d)}, \underline{X}^{(p-d)}, \underline{\xi})\right]\right\} I(\underline{X}^{(d)} \notin C_{\underline{X}^{(d)}}) \right. \\ &+ \left. c_a E_{\underline{X}^{(p-d)}|\underline{X}^{(d)}}\left\{E_{\underline{\xi}}\left[P(\underline{T} \notin C_{\underline{T}}|\underline{X}^{(d)}, \underline{X}^{(p-d)}, \underline{\xi})\right]\right\} I(\underline{X}^{(d)} \in C_{\underline{X}^{(d)}})\right] \\ &+ c_s(d). \end{aligned} \tag{5.6}$$

Comparing the Bayes costs (5.6) and (5.1) it follows from Corollary 5.1 that the d-dimensional Bayes one-stage screen is

$$C^*_{\underline{X}^{(d)}} = \left\{ \underline{x}^{(d)} \in \Omega_{\underline{X}^{(d)}} : E_{\underline{X}^{(p-d)}|\underline{X}^{(d)}=\underline{x}^{(d)}}\left\{E_{\underline{\xi}}\left[P(\underline{T} \in C_{\underline{T}}|\underline{x}^{(d)}, \underline{X}^{(p-d)}, \underline{\xi})\right]\right\} \geq \frac{c_a}{c_a + c_r} \right\}. \tag{5.7}$$

Hence if we build a model that uses a p–dimensional screening variable but we only wish to use d of those dimensions when performing the screen we must account for the unused $p-d$ dimension via a suitable conditional expectation. It would be possible to construct a new model involving only the required d screening components but care would be needed to ensure that the marginal model was consistent with the full p–dimensional model. Assuming the marginal model was of the same form as the full model would generally only be suitable as an approximate method. So, in order to use the Bayes acceptance region (5.7) we need to average our diagnostic probabilities over the distribution of $\underline{X}^{(p-d)}|\underline{X}^{(d)} = \underline{x}^{(d)}$. This will generally involve numerical integration methods such as quadrature or Monte-Carlo techniques. In terms of the linear probit regression model considered above, the d-dimensional Bayes one-stage screen is

$$C^*_{\underline{X}^{(d)}} = \left\{ \underline{x}^{(d)} \in \Omega_{\underline{X}^{(d)}} : E_{\underline{X}^{(p-d)}|\underline{X}^{(d)}=\underline{x}^{(d)}} \left[\Phi \left\{ \frac{\underline{m}_1^T \underline{x}_0^{(d)} + \underline{m}_2^T \underline{X}^{(p-d)}}{f\left(\underline{x}_0^{(d)}, \underline{X}^{(p-d)}\right)^{1/2}} \right\} \right] \geq \frac{c_a}{c_a + c_r} \right\}, \tag{5.8}$$

where

$$f\left(\underline{x}_0^{(d)}, \underline{X}^{(p-d)}\right) = 1 + \underline{x}_0^{(d)T} \boldsymbol{S}_{11} \underline{x}_0^{(d)} + \underline{X}^{(p-d)T} \boldsymbol{S}_{22} \underline{X}^{(p-d)} + 2\underline{x}_0^{(d)T} \boldsymbol{S}_{12} \underline{X}^{(p-d)},$$

$\underline{x}_0^{(d)} = \left(1, \underline{x}^{(d)T}\right)^T$, and the $\boldsymbol{S}_{ij}$ and m_i are the $d+1$ and $p-d$ constituent parts of $\boldsymbol{S}$ and $\underline{m}$ respectively. Note that the region is more complex than in (5.5).

We now have an optimal one–stage screen for a given set of $d(\leq p)$ screening components. In order to compare the Bayes cost for optimal regions of different components, the full distribution of $\underline{X}^{(p)}$ must be specified. The cost calculation will in general involve numerical methods.

An important ancillary question in the reduction of the dimension of the screen is how to choose which screening components to use in a d–dimensional screen. In cases in which the cost of measuring the screening variable $c_s(d)$ is a constant for each dimension d, that is, $c_s(d)$ is the same under each of the different combinations of covariates that could make up a d-dimensional screening variable, the problem of finding the optimal screen design can be split into two stages. Firstly, that of choosing which components are the cheapest to use for each $d = 1, \ldots, p$, and secondly,

comparing the Bayes cost of the optimal components for different d via consideration of screening cost $c_s(d)$.

Let 2^p denote all the subsets of screening components $\{X_1, X_2, \ldots, X_p\}$ and let S_d denote all ${}^pC_d$ subsets of $\{X_1, X_2, \ldots, X_p\}$ of order d. Consider a choice of d screening components, $\sigma(\in 2^p)$, and its minimum Bayes cost excepting the screening cost

$$K(\sigma) = \mathcal{K}\{C_\sigma^*\} - c_s(d).$$

The best choice of components, σ, of a given dimension d, written σ_d, satisfies

$$K(\sigma_d) = \min_{\sigma \in S_d} K(\sigma).$$

Comparing the minimum Bayes cost for each member of S_d will yield σ_d. However, this can be computationally intensive and in section 5.4 we look at some heuristic approaches to finding σ_d.

Once we have σ_d for all choices of d $(= 1, \ldots, p)$ we can incorporate the screening cost and find the optimal screening design to employ.

Definition 5.2 The *Bayes design* is σ_{d^*} such that $0 \leq d^* \leq p$ and

$$c_s(d^*) + K(\sigma_{d^*}) = \min_{0 \leq d \leq p} \{c_s(d) + K(\sigma_d)\}, \tag{5.9}$$

where $d = 0$ corresponds to rejecting all items or accepting all items without using any screening procedure.

Note that it is natural to assume that $c_s(d)$ is increasing in d. $K(\sigma_d)$ is certainly decreasing in d, since any screen based on d components may be considered as a $d+1$ component screen. The best choice of components for dimension $d+1$ is σ_{d+1} and so $K(\sigma_{d+1}) \leq K(\sigma_d)$. Hence we are seeking an optimum tradeoff between complexity and precision.

5.4 Heuristic approaches for choosing screening components

When there are only small differences between the costs of measuring different components in the screening variable it is intuitively sensible that those components with the most joint discriminative power will minimise Bayes cost. In trying to find a very simple heuristic choice of components to include in the screen, some measure of individual discriminative power that also accounts for the uncertainty about that power would seem appropriate as a starting point. Boys & Glazebrook (1992) show that when $T|\underline{X}^{(p)}, \underline{\xi}$ follows a linear probit model, the magnitude of $m_i/s_i \equiv E(\xi_i)/s.d.(\xi_i)$ is a good measure of the discriminative power of screening component i alone. Therefore this measure can be used here to rank components. Section 5.5 presents some numerical work that shows this technique to be a good heuristic for the choice of screening components. However, some account may need to be taken of the correlation structure of the probit regression parameters ξ.

When all of the components are of roughly equal discriminative power with $m_i \simeq m_j$, $i, j = 1, 2, \ldots, p$ the ranking method above degenerates to a ranking based on the precision of ξ_i given by s_i^{-2}. Assuming that the distribution of $\underline{\xi}$ is based on a moderately large sample and weak prior information, $\underline{\xi}$ is approximately normally distributed with mean $\underline{m}$ given by the maximum likelihood estimate (posterior mode) of $\underline{\xi}$ and the inverse of the variance matrix $\boldsymbol{S}^{-1}$ given by Fisher's information matrix $\mathcal{I}(\underline{\xi})$ evaluated at $\underline{\xi} = \underline{m}$. Hence, given $\underline{m}$, the matrix $\boldsymbol{S}^{-1}$ is a function of the data on $\underline{X}$ and has expected value $E_{\underline{X}}\{\mathcal{I}(\underline{m})\}$. So, as an alternative to a ranking based on s_i^{-2}, here it seems natural to look at the properties of the distribution of $\underline{X}$ for an appropriate heuristic. With each covariate assumed standardised with zero mean and unit variance, a possible base for a ranking of components is the correlation structure of $\underline{X}^{(p)}$. When the correlations between screening components are equal, $(corr(X_i, X_j) = \rho$ for $i, j = 1, \ldots, p$, $(i \neq j)$ and $-(p-1)^{-1} < \rho < 1)$, all of the screening components are exchangeable and any d–dimensional choice will be

optimal. The design problem is reduced to finding the optimal number of screening components to use (discussed in section 5.3).

Another special case, which offers some insight, is that of serial correlation. Here $corr(X_i, X_j) = \rho^{|i-j|}$ for $i, j = 1, \ldots, p$, $i \neq j$ and $|\rho| < 1$. A useful heuristic would be a univariate summary of the value of any particular choice of components which would, in turn, enable a ranking of choices. As the focus is on correlation structure a natural choice is a ranking based on maximising the determinant of the correlation matrix of the screening components being considered. We now investigate this heuristic in detail.

Consider a choice of d screening components, $\sigma = \{X_{\ell_1}, X_{\ell_2}, \ldots, X_{\ell_d}\}$, taken from a p–dimensional screening variable with serial correlation, where $\ell_i \in \{1, 2, \ldots, p\}$ and $\ell_1 < \ell_2 < \ldots < \ell_d$. Denoting $M_d(\sigma) = Corr(\sigma)$, the correlation matrix of the (standardised) screening components σ, then

$$M_d(\ell_1, \ell_2, \ldots, \ell_d) = \begin{pmatrix} 1 & \rho^{\ell_2-\ell_1} & \ldots & \rho^{\ell_{d-1}-\ell_1} & \rho^{\ell_d-\ell_1} \\ \rho^{\ell_2-\ell_1} & 1 & \ldots & \rho^{\ell_{d-1}-\ell_2} & \rho^{\ell_d-\ell_2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho^{\ell_{d-1}-\ell_1} & \rho^{\ell_{d-1}-\ell_2} & \ldots & 1 & \rho^{\ell_d-\ell_{d-1}} \\ \rho^{\ell_d-\ell_1} & \rho^{\ell_d-\ell_2} & \ldots & \rho^{\ell_d-\ell_{d-1}} & 1 \end{pmatrix}$$

By subtracting $\rho^{\ell_2-\ell_1} \times$row 2 from row 1 of the above matrix it is easy to see that the determinant of the matrix satisfies the relation:

$$\det\{M_d(\ell_1, \ell_2, \ldots, \ell_d)\} = \left(1 - \rho^{2(\ell_2-\ell_1)}\right) \det\{M_{d-1}(\ell_2, \ldots, \ell_d)\}.$$

Denoting as $D_d(\sigma)$ the determinant of $M_d(\sigma)$,

$$D_d(\ell_1, \ell_2, \ldots, \ell_d) = \prod_{i=1}^{d-1} \left(1 - \rho^{2(\ell_{i+1}-\ell_i)}\right). \tag{5.10}$$

As the evaluation of D_d is based on the differences $\ell_{i+1} - \ell_i$, we can redefine the problem in terms of $k_i = \ell_{i+1} - \ell_i$, $i = 1, 2, \ldots, d-1$. Now our task is to choose $\underline{k} = (k_1, k_2, \ldots, k_{d-1})$ which maximises $D_d(\underline{k})$. Note that for all possible choices of $\underline{k}$,

$$R(\underline{k}) \equiv \sum_{i=1}^{d-1} k_i = \ell_d - \ell_1 \leq p - 1. \tag{5.11}$$

It follows from the following theorem that D_d is maximised when all the elements of $\underline{k}$ take one of two values, r or $r+1$, subject to satisfying (5.11).

Theorem 5.1 For any choice of $\underline{k}$ such that $|k_i - k_j| > 1$, for some $i, j \in \{1, 2, \ldots, d-1\}$, there exists an alternative choice of $\underline{k} = \underline{k}^*$, with $D_d(\underline{k}^*) > D_d(\underline{k})$.

Proof

Consider $\underline{k} = (k_1, k_2, \ldots, k_i, \ldots, k_j, \ldots, k_{d-1})$, where $|k_j - k_i| > 1$ and, without loss of generality, $k_j > k_i$. Now consider an alternative choice $\underline{k}^* = (k_1, k_2, \ldots, k_i + 1, \ldots, k_j - 1, \ldots, k_{d-1})$ with $R(\underline{k}) = R(\underline{k}^*)$. From (5.10) it can be seen that $D_d(\underline{k}^*) > D_d(\underline{k})$ if

$$(1 - \rho^{2(k_i+1)})(1 - \rho^{2(k_j-1)}) > (1 - \rho^{2k_i})(1 - \rho^{2k_j}).$$

After a little algebra this is equivalent to

$$(1 - \rho^2)(\rho^{2k_i} - \rho^{2(k_j-1)}) > 0,$$

which is true as $k_j - 1 > k_i$ and $|\rho| < 1$. □

With the result given in Theorem 5.1 we can solve the problem of choosing a suitable $\underline{k}$ and hence σ.

Theorem 5.2 The choice of $\underline{k} = (k_1, k_2, \ldots, k_{d-1})$ which maximises $D_d(\underline{k})$ is:

Any	$(d-1) - [(p-1) \bmod (d-1)]$	of the $k_i = (p-1) \operatorname{div} (d-1)$
and the other	$(p-1) \bmod (d-1)$	of the $k_i = (p-1) \operatorname{div} (d-1) + 1$,

where $i = 1, 2, \ldots, d-1$.

Proof

Theorem 5.1 shows that the optimal choice of $\underline{k}$ must of the form: any a of the $k_i = r$, say, and the other $b = (d-1) - a$ of the $k_i = r+1$. So we need to find r, a and b that maximise D_d, where

$$D_d = \left(1 - \rho^{2r}\right)^a \left(1 - \rho^{2(r+1)}\right)^b,$$

subject to

$$ar + b(r+1) \leq p - 1 \tag{5.12}$$

and $a + b = d - 1$. Given r, it is clear that b should be as large as possible. Under (5.12) it follows that $(d-1)r \leq (p-1)$ and the largest possible value r can take is

$$r^* = (p-1) \text{ div } (d-1).$$

Maximising D_d for this choice of r, the largest value of b, subject to (5.12), satisfies equality in (5.12) and is given by:

$$\begin{aligned} b^* &= (p-1) - (d-1) \times (p-1) \text{ div } (d-1) \\ &= (p-1) \bmod (d-1). \end{aligned}$$

Hence, these choices of r and b give

$$D_d = \left(1 - \rho^{2r^*}\right)^{d-1-b^*} \left(1 - \rho^{2(r^*+1)}\right)^{b^*}. \tag{5.13}$$

For choices of $r < r^*$, we can choose $b = d - 1$ and still satisfy the constraint (5.12). Hence, for such choices of r, the largest value of D_d is

$$D_d = \left(1 - \rho^{2(r^*-n+1)}\right)^{d-1},$$

where $0 < n < r^*$. Plainly this is smaller than D_d in (5.13) above for all n, $0 < n < r^*$. Hence r^* and b^* give the location of the constrained maximum of D_d and the result follows from these choices of r and b. □

When we have the special case of serial correlation, Theorem 5.2 gives us the choice of screening components which maximise the determinant of their correlation matrix. In essence it says that when choosing the components to use, $\{X_{\ell_1}, X_{\ell_2}, \ldots, X_{\ell_d}\}$, the ℓ's should be as spaced out as possible. For example, if we have a 10–dimensional screening variable and we wish to use 3 screening components, there are two choices of screening components that satisfy Theorem 5.2, namely $\{X_1, X_5, X_{10}\}$ and $\{X_1, X_6, X_{10}\}$. This heuristic choice of screening components is, of course, applicable to any correlation matrix. Its main advantage is its simplicity, though its performance needs to be

assessed. Section 5.5 investigates the performance of this heuristic choice of screening components within a serial correlation framework.

The heuristic above is based solely on the correlation structure of those elements to be employed in the screen and takes no account of the influence of those components not to be included. Other sensible heuristics based on the correlation matrix involve the maximisation or minimisation (as appropriate) of the determinant or trace of the matrices $E_{\underline{X}^{(p-d)}}\left[Corr(\underline{X}^{(d)}|\underline{X}^{(p-d)})\right]$ and/or $E_{\underline{X}^{(d)}}\left[Corr(\underline{X}^{(p-d)}|\underline{X}^{(d)})\right]$. When $\underline{X}^{(p)}$ is multivariate normal these expected conditional correlation matrices are straightforward to calculate as the conditional matrices do not depend on the components over which the expectation is taken. However, in general this is not the case and the simpler heuristic described earlier may be more attractive. The performance of heuristics that use these procedures in the context of a multivariate normal screening variable are investigated in section 5.5.

5.5 Some numerical examples

Under the probit model for $T|\underline{X}^{(p)}$, in this section we give some numerical examples of d-dimensional screen design and evaluate the heuristic proposals of section 5.4 numerically. Throughout this section we fix the cost of incorrectly accepting or rejecting an item as equal ($c_a = c_r = 1$) and assume that the cost of measuring the d-dimensional screening variable $\underline{X}^{(d)}$ is of the form $c_s(d) = dc_s$, where c_s is the constant cost of measuring any one of the covariates. For convenience we also assume that the (standardised) screening variable has a standard normal distribution with correlation matrix Σ.

As necessary preliminaries, in section 5.5.1 we describe the calculation of the screening region and section 5.5.2 describes an algorithm to compute a covariance matrix for the regression parameters that would be typical if the screen were based on large sample. Section 5.5.3 describes the estimation of Bayes cost of the screen and we finally turn to the performance of the heuristics in section 5.5.4.

5.5.1 The optimal one–stage screen

When all of the available covariates are used in the screen the form of the Bayes one–stage screen given in (5.5) is easy to calculate. However, the optimal d–dimensional screen ($d < p$) given in (5.8) is more complex and is determined by an expectation over the conditional distribution of the unused covariates. In our numerical work we calculate this expectation as follows.

Partitioning the correlation matrix of $\underline{X}^{(p)}$ into its d and $p-d$ constituent parts we write

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^T & \Sigma_{22} \end{pmatrix}.$$

Then, under the properties of the normal distribution it is easy to show that

$$\underline{X}^{(p-d)} | \underline{X}^{(d)} = \underline{x}^{(d)} \sim N_{p-d}\left(\mu_{\underline{x}^{(d)}}, \Sigma_{2|1}\right),$$

where $\mu = \Sigma_{12}^T \Sigma_{11}^{-1}$ and $\Sigma_{2|1} = \Sigma_{22} - \Sigma_{12}^T \Sigma_{11}^{-1} \Sigma_{12}$. Hence, the conditional expectation in (5.8) is over a normal distribution. Making a change of variables, we can rephrase the expectation as one over a multivariate standard normal random variable. We write

$$\underline{X}^{(p-d)} | \underline{X}^{(d)} = \underline{x}^{(d)} = \mu_{\underline{x}^{(d)}} + \Sigma_{2|1}^{1/2} \underline{Z}$$

where $\underline{Z} \sim N_{(p-d)}(\underline{0}, I)$ and I is the identity matrix. After some algebra the expectation in (5.8) becomes

$$P_T \equiv E_{\underline{Z}}\left[\Phi\left\{\frac{\alpha + \underline{\beta}^T \underline{Z}}{\left(\gamma + \underline{\lambda}^T \underline{Z} + \underline{Z}^T \Delta \underline{Z}\right)^{1/2}}\right\}\right], \tag{5.14}$$

where

$$\begin{aligned} \alpha &= \underline{m}_1^T \underline{x}_0^{(d)} + \underline{m}_2^T \mu_{\underline{x}^{(d)}} \\ \underline{\beta} &= \Sigma_{2|1}^{1/2^T} \underline{m}_2 \\ \gamma &= 1 + \underline{x}_0^{(d)^T} S_{11} \underline{x}^{(d)} + 2\underline{x}_0^{(d)^T} S_{12} \mu_{\underline{x}^{(d)}} + (\mu_{\underline{x}^{(d)}})^T S_{22} \mu_{\underline{x}^{(d)}} \\ \underline{\lambda} &= 2(S_{02}\Sigma_{2|1}^{1/2})^T \underline{x}^{(d)} \\ \Delta &= \Sigma_{2|1}^{1/2^T} S_{22} \Sigma_{2|1}^{1/2} \end{aligned}$$

To compute the expectation in (5.14) we use the Monte—Carlo integration method discussed in section 3.3.3, with the multivariate standard normal distribution as the sampled distribution and the $h(\cdot)$–function given by

$$h(\underline{z}) = \Phi\left\{\frac{\alpha + \underline{\beta}^T\underline{z}}{\left(\gamma + \underline{\lambda}^T\underline{z} + \underline{z}^T\Delta\underline{z}\right)^{1/2}}\right\} \tag{5.15}$$

Recall that the precision of the estimate of the integral (expectation) increases with more sampling. Hence we can sentence an item as soon as we have done enough sampling to be reasonably sure whether the expectation is greater than or less than $c_a/(c_r + c_a)$. The following algorithm was used to determine the suitability of items based on a measurement $\underline{x}^{(d)}$, of a d–dimensional screening variable.

Algorithm

1. Compute μ, $\Sigma_{2|1}^{1/2}$ and other quantities in (5.15) that can be computed without knowing $\underline{x}^{(d)}$ and $\underline{z}$.

2. Now compute quantities in (5.15) that include $\underline{x}^{(d)}$ and exclude $\underline{z}$.

3. Set at zero storage bins for the sum of the simulated values of the h–function and the sum of squared values of the h–function, denoted Φ_Σ and Φ_{Σ^2} say. Also set $i = 0$.

4. Let $i = i + 1$ and generate $p - d$ standard normal pseudo-random observations $z_1^{(i)}, z_2^{(i)}, \ldots, z_{p-d}^{(i)}$, say, using N.A.G. routine G05DDF, N.A.G. (1990). As $\underline{Z}$ is standardised with each component of $\underline{Z}$ independent, set $\underline{z}^{(i)} = (z_1^{(i)}, z_2^{(i)}, \ldots, z_{p-d}^{(i)})$.

5. Calculate $h(\underline{z}^{(i)})$, where $h(\cdot)$ is as in (5.15) and let $\Phi_\Sigma = \Phi_\Sigma + h(\underline{z}^{(i)})$ and $\Phi_{\Sigma^2} = \Phi_{\Sigma^2} + h(\underline{z}^{(i)})^2$.

6. If $i < i_{\min}$ then return to step 4, otherwise calculate an estimate of the expectation (5.14) and an estimate of the variance of the estimator as follows,

$$\begin{aligned} \hat{P}_T &= \Phi_\Sigma/i, \\ \widehat{Var(\hat{P}_T)} &= (\Phi_{\Sigma^2} - i\hat{P}_T^2)/\{i(i-1)\}. \end{aligned}$$

7. If $\hat{P}_T - 1.96\sqrt{\widehat{Var}(\hat{P}_T)} \geq c_a/(c_a + c_r)$ then accept the item and stop,
   if $\hat{P}_T + 1.96\sqrt{\widehat{Var}(\hat{P}_T)} < c_a/(c_a + c_r)$ then reject the item and stop.

8. If $i \leq i_{\max}$ then return to step 4, otherwise, if $\hat{P}_T \geq c_a/(c_a + c_r)$ accept the item, else reject the item, in both cases noting the possible inaccuracy of the allocation.

Step 7 uses a normal approximation for P_T to sentence the item if the estimate of the expectation is more than 1.96 standard deviations above or below $c_a/(c_a + c_r)$. The bounds $i_{\min}$ and $i_{\max}$ are the minimum and maximum number of simulations that should be generated to sentence an item. In the last step the estimate of the expectation is not accurate enough to be sure of a correct sentence. Here, if $i = i_{\max}$, we sentence based on the final estimate and flag that the allocation may be incorrect. In the studies of the performance of the heuristic in section 5.5.4 we set $i_{\min} = 1000$ and $i_{\max} = 100000$ when using the algorithm in section 5.5.3 to estimate Bayes cost. These choices resulted in less than 0.5% of allocations being made by step 8.

5.5.2 Covariance structure of the regression parameters

Recall that when the probit regression model on $T|\underline{X}$ is based on a large sample of data, the parameters of the regression model $\underline{\xi}$ are approximately normally distributed with mean given by the maximum likelihood estimate of $\underline{\xi}$ and the covariance as the inverse of Fisher's information matrix evaluated at the maximum likelihood estimator. Hence given the mean $E(\underline{\xi}) = \underline{m}$ and the distribution of $\underline{X}^{(p)}$,

$$E_{\underline{X}^{(p)}}\left\{Var(\underline{\xi})^{-1}\right\} = E_{\underline{X}^{(p)}}(\boldsymbol{S}^{-1}) = E_{\underline{X}^{(p)}}\left\{\mathcal{I}(\underline{m})\right\}, \tag{5.16}$$

where $\mathcal{I}(\underline{m})$ is Fisher's information matrix of $\underline{\xi}$ based on a sample of size n evaluated at $\underline{m}$. Throughout the numerical work presented in this section, given $\underline{m}$ and $\boldsymbol{\Sigma}$, the correlation matrix of $\underline{X}^{(p)}$, we compute the covariance matrix of $\underline{\xi}$ as the inverse of the expectation in (5.16). Hence $\boldsymbol{S}$ will be typical of the covariance matrices found when $p(\underline{\xi})$ is based on a moderate sample of data.

Following (2.13) and (4.25), under a probit regression for $T|\underline{X}^{(p)}$ the (j,k)th element of Fisher's information matrix of the parameters $\underline{\xi}$ from one observation $(t, \underline{x}^{(p)})$ on $(T, \underline{X}^{(p)})$ is, $j, k = 0, 1, \ldots, p$,

$$\imath_{jk}(\underline{\xi}, \underline{X}^{(p)}) = \frac{x_j x_k \phi(\underline{\xi}^T \underline{x}_0^{(p)})^2}{\Phi(\underline{\xi}^T \underline{x}_0^{(p)})\{1 - \Phi(\underline{\xi}^T \underline{x}_0^{(p)})\}}, \tag{5.17}$$

with $x_0 = 1$. Evaluating (5.17) at $\underline{\xi} = \underline{m}$ and taking an expectation over the distribution of $\underline{X}^{(p)}$ gives the elements of the expected value of Fisher's information matrix as

$$E_{\underline{X}^{(p)}} \left\{ \imath_{jk}(\underline{m}, \underline{X}^{(p)}) \right\} = \int_{\mathbf{R}^p} \imath_{jk}(\underline{m}, \underline{x}^{(p)}) \phi_p(\underline{x}^{(p)}|\Sigma)\, d\underline{x}^{(p)}, \tag{5.18}$$

where $\phi_p(\underline{x}^{(p)}|\Sigma)$ is the standardised multivariate normal distribution with correlation matrix Σ. For a sample of n observations, the elements of the expected value of Fisher's information matrix $E_{\underline{X}^{(p)}} \{\mathcal{I}(\underline{m})\}$ in (5.16) has (j,k)th element

$$E_{\underline{X}^{(p)}} \{I_{jk}(\underline{m})\} = \sum_{i=1}^{n} E_{\underline{X}^{(p)}} \left\{ \imath_{jk}(\underline{m}, \underline{X}^{(p)}) \right\} = n E_{\underline{X}^{(p)}} \left\{ \imath_{jk}(\underline{m}, \underline{X}^{(p)}) \right\}.$$

As in Chapter 4, throughout our examples we suppose that the sample size is $n = 30$.

To compute the integrals in (5.18) we use Monte–Carlo integration in preference to Gauss–Hermite quadrature as used in section 4.4.2. The reasoning behind this choice is that here the computation will involve integration over a larger number of dimensions than in section 4.4.2 and quadrature may not perform as well. Also, the accuracy of the estimate of the integral is not essential as we only want the information matrix and hence covariance matrix to be roughly typical of those from a large sample. In fact, the implementation of the Monte–Carlo method that is described below essentially calculates the information matrix at $\underline{m}$ for a large number of pseudo observations on $\underline{X}^{(p)}$ and then rescales to the smaller sample size of $n = 30$.

The Monte–Carlo technique has already been described in detail elsewhere and so we simply give the algorithm for computing an estimate of the expectation in (5.16):

Algorithm

1. Initialise at zero $\hat{\mathcal{I}}$, a $(p+1) \times (p+1)$ matrix of storage bins to hold the information matrix. Set the number of simulations generated so far $i = 0$.

2. Set $i = i + 1$ and generate a pseudo-random observation $\underline{x}^{(i)}$ on $\underline{X}^{(p)}$, using N.A.G. routine G05EZF say, N.A.G. (1990).

3. For $j, k = 0, 1, \ldots, p$, calculate the information from this observation $\imath_{jk}(\underline{m}, \underline{x}^{(i)})$ and add it to the elements of current information matrix $\hat{\mathcal{I}}_{jk} = \hat{\mathcal{I}}_{jk} + \imath_{jk}(\underline{m}, \underline{x}^{(i)})$.

4. If $i < i_{\max}$ then return to step 2, otherwise compute the estimate of the expected observed information matrix (5.16) for 30 observations as $\hat{\mathcal{I}} = 30\hat{\mathcal{I}}/i_{\max}$.

5. Invert $\hat{\mathcal{I}}$ to obtain the covariance matrix.

Notice that we do not stop the simulation process when a desired accuracy is achieved, as in earlier implementations of the Monte–Carlo method. This is because here it seems more natural to compute the covariance matrix for all problems based on the same number of pseudo–observations. Some experimentation can be carried out to find an $i_{\max}$ that gives a reasonable accuracy by computing the standard error of the estimate $\hat{\mathcal{I}}$ in the usual way. In the numerical studies in section 5.5.4 we found that $i_{\max} = 1,000,000$ gave a suitably accurate answer (the information matrix correct to at least 2 decimal places).

5.5.3 Computing Bayes cost

For our model, the Bayes cost of a d-dimensional Bayes one–stage screen is,

$$\mathcal{K}\left\{C^*_{\underline{X}^{(d)}}\right\} = \int_{\mathbf{R}^p} \left[c_r \Phi \left\{ \frac{\underline{m}^T \underline{x}_0^{(p)}}{(1 + \underline{x}_0^{(p)T} \boldsymbol{S} \underline{x}_0^{(p)})^{1/2}} \right\} I\left(\underline{x}^{(d)} \notin C^*_{\underline{X}^{(d)}}\right) + c_a \Phi \left\{ \frac{-\underline{m}^T \underline{x}_0^{(p)}}{(1 + \underline{x}_0^{(p)T} \boldsymbol{S} \underline{x}_0^{(p)})^{1/2}} \right\} I\left(\underline{x}^{(d)} \in C^*_{\underline{X}^{(d)}}\right) \right] \phi_p(\underline{x}^{(p)} | \boldsymbol{\Sigma})\, d\underline{x}^{(p)} + c_s(d).$$

The Monte–Carlo algorithm used to estimate this Bayes cost follows easily from that given to calculate the Bayes cost of a two–stage screen in section 3.3.3. The main difference is that in step 3 we partition $\underline{x}^{(i)}$, the simulated observation on $\underline{X}^{(p)}$ into its

d and $p-d$ constituent parts, $\underline{x}^{((i)^{(d)})}$ and $\underline{x}^{((i)^{(p-d)})}$ and submit $\underline{x}^{((i)^{(d)})}$ to the algorithm described in section 5.5.1 to determine whether it is at an acceptable level or not. If $\underline{x}^{((i)^{(d)})}$ is acceptable we set $h(\underline{x}^{(i)}) = c_a\Phi\{-\underline{m}^T\underline{x}_0^{(p)}/(1+\underline{x}_0^{(p)^T}\boldsymbol{S}\underline{x}_0^{(p)})^{1/2}\}\}$, otherwise we set $h(\underline{x}^{(i)}) = c_r\Phi\{\underline{m}^T\underline{x}_0^{(p)}/(1+\underline{x}_0^{(p)^T}\boldsymbol{S}\underline{x}_0^{(p)})^{1/2}\}\}$. When $d = p$ the acceptability of an item is determined by the simple rule given in 5.5. Unfortunately, the extra step of having to take an expectation to determine whether $\underline{x}^{(d)}$ is acceptable increases the computational time needed to compute Bayes cost. Hence, in the study of the performance of the heuristics in the next section we were only able to obtain estimates of Bayes cost with a standard error typically just under 0.5% of the Bayes cost.

5.5.4 Performance of the heuristics

We now turn to the performance of the suggested heuristics, whose aim is to determine the optimal d–dimensional choice of components without resorting to the numerically intensive calculation of Bayes cost.

First we examine the heuristic that ranks components according to the magnitude of $m_i^* \equiv m_i/s_i = E(\xi_i)/s.d.(\xi_i)$. We suppose that we have available three covariates on which to base the screen. We compare the ranking of components under the heuristic with the Bayes–optimal ranking by estimating Bayes costs, as in section 5.5.3. The heuristic is intended for ranking components and so we are interested in finding σ_d, the optimal components in the screen for fixed d. Hence, we compare Bayes costs with the cost of performing the screen, $c_s(d)$, subtracted. To make the study manageable we assume that each covariate is independent of the others so that $Corr(X_i, X_j) = 0$, $i, j = 1, 2, 3$ $(i \neq j)$ and compute the covariance matrix as the inverse of the expected value of Fisher's information matrix using the algorithm given in section 5.5.2. Hence the only inputs to the screen that we vary in our study are the mean regression parameters $\underline{m}$.

We set $m_0 = 0$ and $m_1 = 1$ and considered all possible designs in which $m_2 = 0.9$, 0.95, 0.975, 0.99, 0.995, 0.999, 0.9995, 1.0 and $m_3 = 1.0$, 1.0005, 1.001, 1.005, 1.01, 1.025, 1.05, 1.1. Hence there were 64 different values for $\underline{m}$ and for each value of $\underline{m}$

there were 7 different designs (three in which only one screening component is used, three in which two are used and one in which all three screening components are used – the last of these is of little interest here). The structure of the study was chosen so that the spacing in the grid of m-values gradually increases as the difference in m–values increases.

Consider choosing the optimal component for use in a one dimensional screen. In the above setup, a ranking of components by m–value will always give an ordering of 3, 1, 2 (other than when $m_2 = 1.0$ and/or $m_3 = 1.0$). However, recall that we only compute the covariance structure to be roughly as expected. Hence an ordering by m–value may not correspond to ordering by m^*–value. All cases with $m_i > m_j$, $i, j = 1, 2, 3$ in our study, resulted in $m_i^* > m_j^*$ but those cases with $m_i = m_j$ did not give $m_i^* = m_j^*$. To compare the ranking of components by m^*–value with a ranking by the Bayes cost of a one component screen, we compare the differences in m^*–value with the differences in Bayes cost. That is, denoting as a_i the Bayes cost of operating a screen based only on component i, $i = 1, 2, 3$, we compare the differences $m_1^* - m_2^*$, $m_3^* - m_1^*$ and $m_3^* - m_2^*$ with $a_2 - a_1$, $a_1 - a_3$ and $a_2 - a_3$ respectively. Hence we define

$$\eta_{2,1} = \frac{a_2 - a_1}{\min(a_1, a_2)} \times 100,$$

where the magnitude of positive values of $\eta_{2,1}$ gives the percentage loss in Bayes cost by using a screen that involves component 2 if a screen using component 1 is cheaper. The magnitude of negative values correspond to the reverse — the percentage loss in Bayes cost by using a screen that includes component 1 if a screen using component 2 is cheaper. Figure 5.1(a) plots estimates of $\eta_{2,1}$ against the difference $m_1^* - m_2^*$, each point corresponding to a different choice of $\underline{m}$. However, Figure 5.1(a) is constructed with Monte–Carlo estimates of Bayes cost and, as already noted, we are only able to estimate Bayes cost to a given accuracy. Recall that under the Monte–Carlo scheme for estimating Bayes cost, the Bayes cost of the screen a_i is approximately normally distributed with mean given by the Monte–Carlo estimate, $\hat{a}_i$ say, and variance given by the variance of the estimator, estimated by $\widehat{Var(\hat{a}_i)}$ say. In cases in which each Monte-Carlo estimate of Bayes cost is computed using different pseudo–samples from

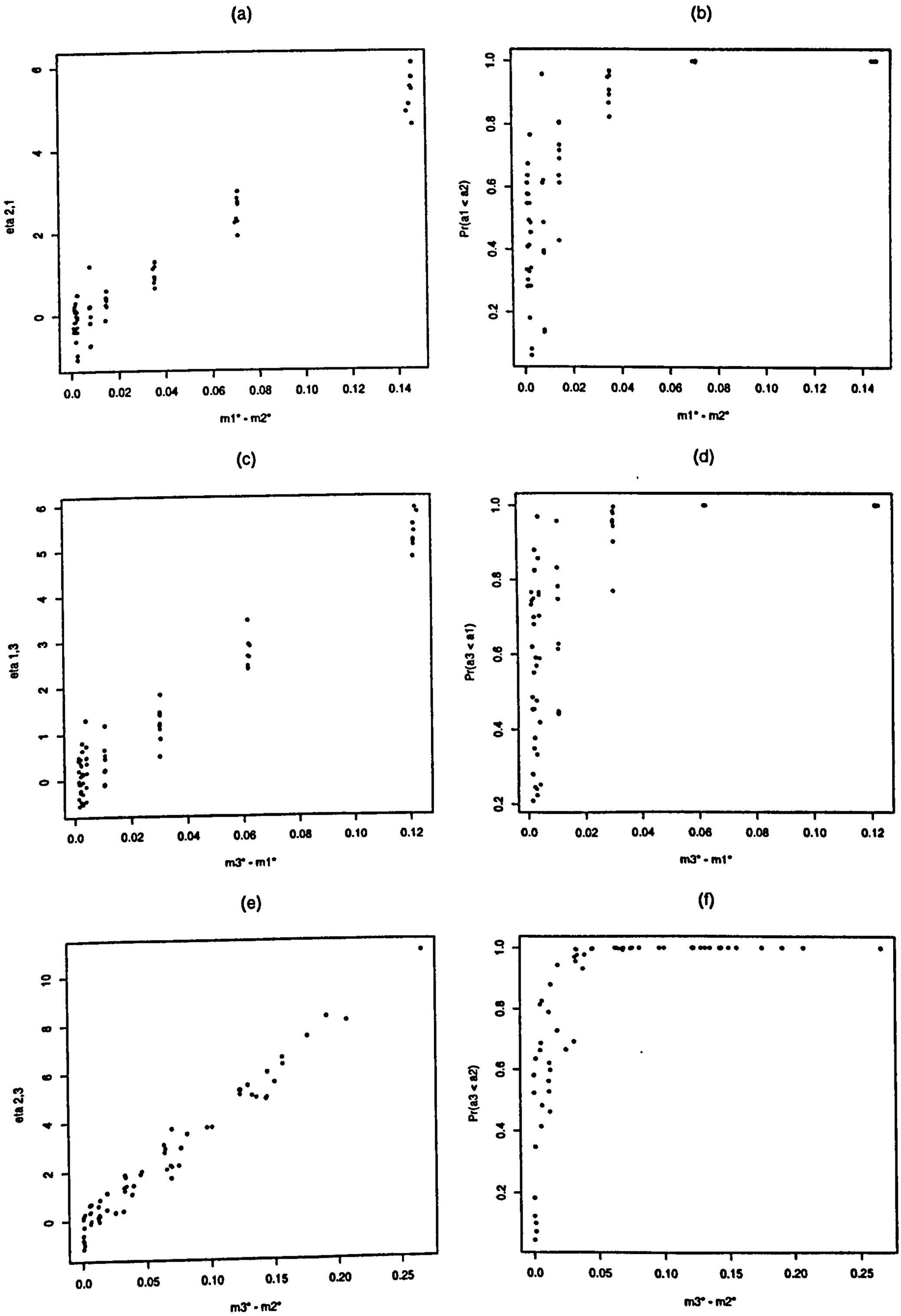


Figure 5.1: Performance of m^*–ordering heuristic for 1–d screening

the multivariate normal distribution, the Bayes costs are independent. In our study this was the case and so the difference in Bayes costs $a_i - a_j$ is approximately normal with mean $\hat{a}_i - \hat{a}_j$ and variance $\widehat{Var(\hat{a}_i)} + \widehat{Var(\hat{a}_j)}$. Based on this normal approximation we use the estimates of Bayes cost and their standard error to construct Figure 5.1(b) which plots $P(a_1 < a_2)$ versus $m_1^* - m_2^*$. When $P(a_1 < a_2)$ is close to 0 or 1 it is clear which Bayes cost is greater, otherwise the estimates of Bayes cost are not accurate enough to make any inferences. We also define

$$\eta_{1,3} = \frac{a_1 - a_3}{\min(a_1, a_3)} \times 100 \quad \text{and} \quad \eta_{2,3} = \frac{a_2 - a_3}{\min(a_2, a_3)} \times 100.$$

The interpretations of $\eta_{1,3}$ and $\eta_{2,3}$ follow naturally from the interpretation of $\eta_{2,1}$. Figure 5.1(c) plots estimates of $\eta_{1,3}$ against $m_3^* - m_1^*$ and Figure 5.1(e) plots estimates of $\eta_{2,3}$ against $m_3^* - m_2^*$. Also, Figure 5.1(d) plots $P(a_3 < a_1)$ versus $m_3^* - m_1^*$, and Figure 5.1(e) shows $P(a_3 < a_2)$ against $m_3^* - m_2^*$.

These plots show that when the difference in m^*–values is large enough, say greater than 0.05, it is clear that ranking by m^*–value works well. For example, in cases when m_1^* is much larger than m_2^*, $\eta_{2,1}$ is large and positive and $P(a_1 < a_2)$ is close to 1. However, when m^*–values are close, the inaccuracy in the computation of Bayes costs does not allow us to assess whether a ranking by m^*–value performs well.

When selecting components to use in a 2–dimensional screen, a ranking by m–value orders pairs of components as $\{1,3\}$, $\{2,3\}$ then $\{1,2\}$. Denoting as a_{ij}, $i,j = 1,2,3$, $i < j$, the Bayes cost of operating a screen incorporating measurements on components i and j, we define

$$\eta_{12,13} = \frac{a_{12} - a_{13}}{\min(a_{12}, a_{13})} \times 100,$$

with $\eta_{12,23}$ and $\eta_{23,13}$ defined similarly. Positive values of $\eta_{12,13}$ give the percentage of Bayes cost lost by using a screen that incorporates components 1 and 2 when a screen that uses components 1 and 3 is cheaper. The magnitude of negative values of $\eta_{12,13}$ gives the percentage loss in Bayes cost by using a screen with components 1 and 3 when a screen with components 1 and 2 is cheaper. The interpretations of $\eta_{12,23}$ and $\eta_{23,13}$ follow naturally. To compare the ranking of pairs of components by

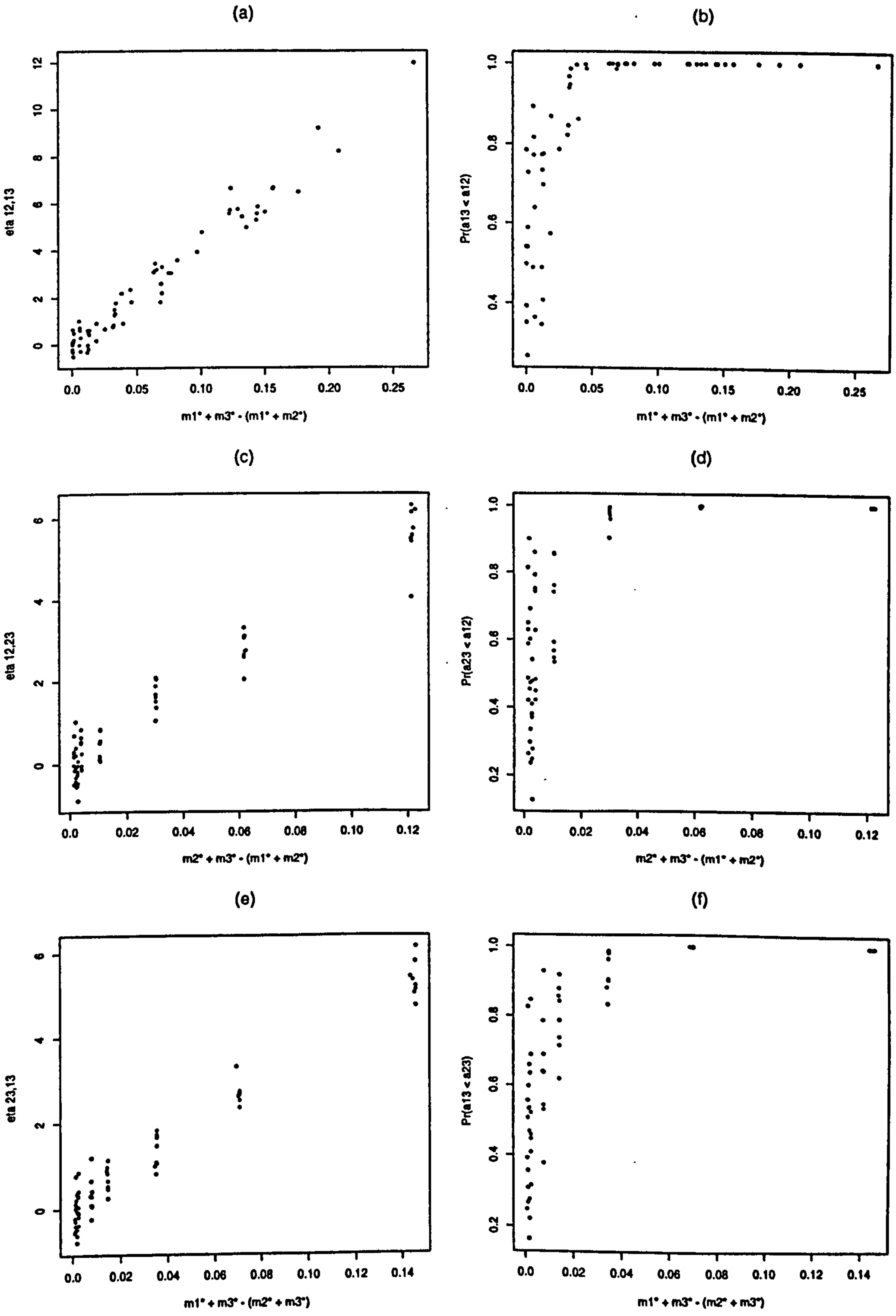


Figure 5.2: Performance of m^*–ordering heuristic for 2–d screening

m^*-value with ranking by Bayes cost we compare the differences $m_i^* + m_j^* - (m_k^* + m_l^*)$ with $a_{kl} - a_{ij}$ for $i, j, k, l = 1, 2, 3$, $i < j$, $k < l$. Hence, Figure 5.2(a) plots estimates of $\eta_{12,13}$ against $m_1^* + m_3^* - (m_1^* + m_2^*)$, Figure 5.2(c) plots estimates of $\eta_{12,23}$ against $m_2^* + m_3^* - (m_1^* + m_2^*)$ and Figure 5.2(e) plots estimates of $\eta_{23,13}$ against $m_1^* + m_3^* - (m_2^* + m_3^*)$. As in the 1–dimensional case we should assess the performance of the heuristic in light of the accuracy of the estimates of Bayes cost and we use the normal approximation for Bayes cost to plot probabilities. Figure 5.2(b) plots $P(a_{13} < a_{12})$ versus $m_1^* + m_3^* - (m_1^* + m_2^*)$, Figure 5.2(d) plots $P(a_{23} < a_{12})$ versus $m_2^* + m_3^* - (m_1^* + m_2^*)$ and Figure 5.2(f) plots $P(a_{13} < a_{23})$ versus $m_1^* + m_3^* - (m_2^* + m_3^*)$. When the difference in the sum of m^*-values is large enough the heuristic clearly performs well, with uncertainty about the difference in Bayes costs, and hence uncertainty about the performance of the heuristic, present when the sums of m^*-values are close.

We now consider situations in which each covariate available for screening has roughly equal discriminative power and the uncertainty about such power is as expected from a typical (large) sample. Here we assess the performance of the heuristics based on the correlation matrix Σ, focusing on the special cases in which the covariates are equi–correlated or serially correlated.

Consider such a case in which we have a possible five (standardised) screening components on which to base our screen and the mean regression parameters in the linear probit model are equal ($m_0 = 0, m_1 = m_2 = \ldots = m_5 = 1$). If the screening components are equally correlated then the problem reduces to choosing the optimal number of components to use in the screen. Figure 5.3(a) shows the estimated Bayes cost when using different numbers of components (d) for the case $c_s = 0.03$ and correlation parameter $\rho = 0.4$. It demonstrates that the optimal screen should be based on (any) three of the five components. This scree plot is typical of those used in multivariate analysis for dimensionality reduction. The sensitivity of the dimension calculation to the screening cost per component (c_s) is easily assessed.

The calculation for the serial correlation case is more complex. Not only must the dimension of the screen be determined but also which components to use. Fig-

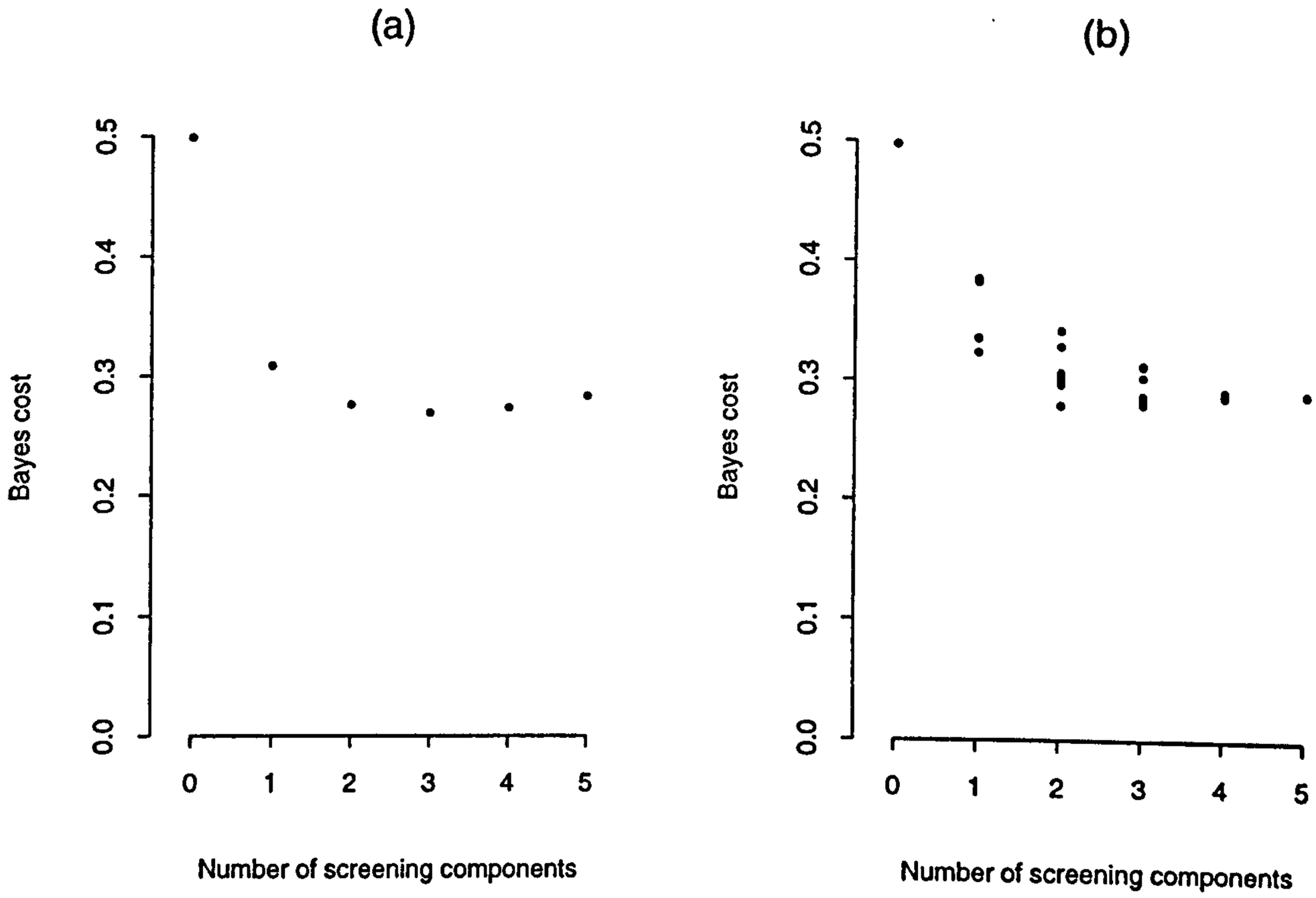


Figure 5.3: Bayes costs for d–dimensional screens

ure 5.3(b) shows the estimated Bayes cost, again for the case $c_s = 0.03$ and correlation parameter $\rho = 0.4$. The plot shows the cost of each of the ${}^5C_d$ choices of components. It is clear that the optimal choice of components decreases in importance as the number of components included increases.

In the case of serial correlation, we now assess the performance of the heuristics that are based on the correlation matrix Σ. Let $\sigma_d^{(1)}$ denote the d–dimensional choice of the "equal spacing" heuristic based on Theorem 5.2. Also, let $\sigma_d^{(2)}$ and $\sigma_d^{(3)}$ denote the choices of heuristics based on the minimisation of $tr\left\{E_{\underline{X}^{(p-d)}}\left[Corr(\underline{X}^{(d)}|\underline{X}^{(p-d)})\right]\right\}$ and $tr\left\{E_{\underline{X}^{(d)}}\left[Corr(\underline{X}^{(p-d)}|\underline{X}^{(d)})\right]\right\}$ respectively. Under our assumption of normality for the standardised screening variable $\underline{X}^{(p)}$,

$$\begin{aligned} Corr(\underline{X}^{(d)}|\underline{X}^{(p-d)}) &= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}, \\ Corr(\underline{X}^{(p-d)}|\underline{X}^{(d)}) &= \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}, \end{aligned}$$

where the Σ_{ij} are the d and $p-d$ constituent parts of of the correlation matrix. Hence

Table 5.1: Performance of heuristics

			ρ			
	d	${}^5C_d$	0.2	0.4	0.6	0.8
$\sigma_d^{(1)}$	1	5	–	–	–	–
	2	10	N (14%)	N (23%)	N (24%)	N (15%)
	3	10	Y	Y	Y	Y
	4	5	Y	Y	Y	Y
$\sigma_d^{(2)}$	1	5	Y	Y	Y	Y
	2	10	Y	Y	Y	Y
	3	10	Y	Y	Y	N (3%)
	4	5	N (0.33%)	N (0.33%)	N (0.11%)	N (0.41%)
$\sigma_d^{(3)}$	1	5	Y	Y	Y	Y
	2	10	Y	Y	Y	N (4%)
	3	10	Y	Y	Y	Y
	4	5	Y	Y	Y	Y

the correlation matrices of the conditional variables are not functions of the random variables over which the expectation is taken and the calculation of the proposed heuristics is reasonably straightforward.

The heuristics performed well in a variety of serial correlation models, with up to five possible components, in the majority of situations giving the choice with the lowest estimated Bayes cost. Table 5.1 shows whether the choice with lowest estimated cost was obtained for various dimensions and values of ρ (indicated Y or N) and where the choice was different (N), the percentage difference in Bayes cost. Again these results need to be assessed in the light of the accuracy of the estimates of Bayes cost. Under the normality approximation for Bayes cost, Table 5.2 gives some indication of the accuracy of the results in Table 5.1. For those cases in which the design with lowest estimated Bayes cost was selected by the heuristic (Y), Table 5.2 gives the probability (correct to 4 significant figures) that the choice with

Table 5.2: Accuracy of results

			ρ			
	d	${}^5C_d$	0.2	0.4	0.6	0.8
$\sigma_d^{(1)}$	1	5	–	–	–	–
	2	10	(1.000)	(1.000)	(1.000)	(1.000)
	3	10	0.1810	0.9297	0.9970	0.9364
	4	5	0.5641	0.7578	0.7364	0.7675
$\sigma_d^{(2)}$	1	5	0.9253	1.000	1.000	1.000
	2	10	1.000	1.000	1.000	0.9990
	3	10	0.1810	0.9297	0.9970	(0.9670)
	4	5	(0.6354)	(0.6233)	(0.5394)	(0.6334)
$\sigma_d^{(3)}$	1	5	0.8296	1.000	1.000	1.000
	2	10	1.000	1.000	1.000	(0.9990)
	3	10	0.1810	0.9297	0.9970	0.9364
	4	5	0.5641	0.7578	0.7364	0.7675

lowest estimated cost has lowest Bayes cost — remember that we already know that it is the choice that is *most likely* to have lowest Bayes cost. When the heuristic chooses a design different to that with lowest estimated Bayes cost, Table 5.2 gives the probability that the choice with lowest estimated Bayes cost has Bayes cost lower than the heuristic choice. Notice that in some cases the uncertainty over Bayes cost results in uncertainty over the performance of the heuristic, even so it is clear that the heuristics perform well.

Chapter 6

Further work

In this chapter we discuss alternatives, improvements and extensions to the techniques described in the thesis and end with some concluding remarks.

6.1 Designing two–stage screens under the sampling paradigm

Recall the case in which T is a binary performance variable, $\underline{X}$ is the screening variable and $\underline{\theta}$ denotes the parameters of the joint model for $(T, \underline{X})$. In Chapter 2 and the probit example in Chapter 3 optimal two–stage designs are achieved when the joint density $p(t, \underline{x}|\underline{\theta})$ is factorised as $p(t|\underline{x}, \underline{\xi})p(\underline{x}|\underline{\phi})$, with $\underline{\theta}^T = (\underline{\xi}^T, \underline{\phi}^T)$. We discuss here an alternative which follows the sampling approach in which $p(t, \underline{x}|\underline{\theta}) = p(\underline{x}|t, \underline{\eta})p(t|\underline{\psi})$.

With the standard case of constant misclassification costs given by c_a and c_r and the cost of measuring the performance variable written as c_m, Lemma 3.1 gives the Bayes two–stage screen as an optimal partition $(\Omega_A^*, \Omega_R^*, \Omega_M^*)$ of $\Omega_{\underline{X}}$, where $\Omega_{\underline{X}}$ is the sample space of $\underline{X}$. It is clear from Lemma 3.1 that an expression for $P(T = 1|\underline{x}) = E_{\underline{\xi}}[P(T = 1|\underline{x}, \xi)]$ is required to investigate the form of the screening regions

$(\Omega_A^*, \Omega_R^*, \Omega_M^*)$. Under the sampling approach, Bayes Theorem gives

$$P(T=1|\underline{x}) = \frac{p(\underline{x}|t=1)P(T=1)}{p(\underline{x}|t=1)P(T=1)+p(\underline{x}|t=0)P(T=0)}. \tag{6.1}$$

Inserting (6.1) into Lemma 3.1, with $c_m/c_r < 1-c_m/c_a$ the optimal screening partition is given by

$$\begin{aligned}
\Omega_A^* &= \left\{\underline{x} \in \Omega_{\underline{X}} : \frac{p(\underline{x}|t=1)}{p(\underline{x}|t=0)} \geq \frac{P(T=0)}{P(T=1)}\left(\frac{k_2}{1-k_2}\right)\right\}, \\
\Omega_R^* &= \left\{\underline{x} \in \Omega_{\underline{X}} : \frac{p(\underline{x}|t=1)}{p(\underline{x}|t=0)} \leq \frac{P(T=0)}{P(T=1)}\left(\frac{k_1}{1-k_1}\right)\right\}, \\
\Omega_M^* &= \left\{\underline{x} \in \Omega_{\underline{X}} : \frac{P(T=0)}{P(T=1)}\left(\frac{k_1}{1-k_1}\right) \leq \frac{p(\underline{x}|t=1)}{p(\underline{x}|t=0)} \leq \frac{P(T=0)}{P(T=1)}\left(\frac{k_2}{1-k_2}\right)\right\},
\end{aligned}$$

where $k_1 = c_m/c_r$ and $k_2 = 1 - c_m/c_a$. Note that the rule for accepting an item is similar to the acceptance rule obtained by taking a sampling approach to the design of a one–stage screen that satisfies the local criterion given in (1.4), see Dunsmore & Boys (1988).

The designs obtained in Chapter 2 and the example of Chapter 3 have the advantages of simplicity and robustness. The advantages of taking the approach described here need to be assessed. Boys & Dunsmore (1987) and Dunsmore & Boys (1988) are useful resources for this purpose. They propose models for $p(\underline{x}|t,\underline{\xi})$ and $p(t|\underline{\phi})$ and describe methods for obtaining the predictive probabilities $P(T=i)$, $i=1,0$ and densities $p(\underline{x}|t=i)$, $i=1,0$.

Other two–stage designs

The very general result given by Theorem 3.1 will allow us to develop two–stage procedures for a variety of models for $(\underline{T},\underline{X})$ and loss functions for misclassification. In particular, it would be interesting to use Theorem 3.1, or, if necessary, extend it, to obtain screen designs for the loss functions proposed by Tang (1988b), taking a predictive approach to the model for $(\underline{T},\underline{X})$.

6.2 Multiple alternatives

The idea of using a screening variable $\underline{X}$ to predict the binary result of whether $\underline{T} \in C_{\underline{T}}$ or $\underline{T} \notin C_{\underline{T}}$ can be extended to cases in which an item should be allocated to one of a number (> 2) of classes depending on $\underline{T}$. For example, Bai & Hong (1992) suppose that the quality of an item, defined by $\underline{T}$, will determine which of an ordered choice of markets an item should be shipped to. That is, if $\underline{T} \in M_i$ say, then the item should be allocated to market i. A procedure based on a correlated screening variable $\underline{X}$ is again proposed as a cheap method of classifying items.

Suppose that there are N markets and the screen allocates an item to market i if $\underline{X} \in \Omega_i$, where $(\Omega_1, \Omega_2, \ldots, \Omega_N)$ is a partition of the sample space $\Omega_{\underline{X}}$. Denote the cost associated with the action of allocating an item to market i as $c_i(\underline{x})$. Generalising Theorem 3.1 as follows provides the optimal partition $(\Omega_1^*, \Omega_2^*, \ldots, \Omega_N^*)$. The Bayes cost of operating the screen is

$$\sum_{i=1}^{N} \int_{\Omega_i} c_i(\underline{x}) p(\underline{x})\, d\underline{x} \;\geq\; \int_{\Omega_{\underline{X}}} \min_j \{c_j(\underline{x})\}\, p(\underline{x})\, d\underline{x}.$$

Therefore, the bound is attained when

$$\Omega_i^* = \left\{ \underline{x} \in \Omega_{\underline{X}} : c_i(\underline{x}) = \min_j \{c_j(\underline{x})\} \right\}, \tag{6.2}$$

for $i = 1, 2, \ldots, N$. This result is similar to the *Bayes minimum risk decision rule*, see Hand (1981) p.6. Theorem 3.1 concerns the case in which

$$\begin{aligned} N = 3, \quad \Omega_1 = \Omega_A, \quad & c_1(\underline{x}) = \tilde{c}_a(\underline{x}), \\ \Omega_2 = \Omega_R, \quad & c_2(\underline{x}) = \tilde{c}_r(\underline{x}), \\ \Omega_3 = \Omega_M, \quad & c_3(\underline{x}) = \tilde{c}_m(\underline{x}). \end{aligned}$$

In the simple case in which T denotes the number of the correct market for an item, the probit modelling approach taken by Boys & Glazebrook (1992) can be generalised. Suppose that losses are only incurred by the misclassification of items and c_{ij} denotes the cost of wrongly allocating an item to market i that should have been dispatched to market j, $i, j = 1, 2, \ldots, N$, $i \neq j$. The Bayes cost is then given

by

$$\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N} c_{ij}P(T=j, X\in\Omega_i) = \sum_{i=1}^{N}\int_{\Omega_i}\left\{\sum_{\substack{j=1\\ j\neq i}}^{N} c_{ij}P(T=j|x)\right\}p(x)\,dx.$$

Following the result given by (6.2) above

$$\Omega_i^* = \left\{\underline{x}\in\Omega_{\underline{X}} : \sum_{\substack{j=1\\ j\neq i}}^{N} c_{ij}P(T=j|\underline{x}) = \min_k\left[\sum_{\substack{j=1\\ j\neq k}}^{N} c_{kj}P(T=j|\underline{x})\right]\right\}. \qquad (6.3)$$

To obtain the predictive probability $P(T=i|x)$ we propose the N-category ordered probit model for $P(T=i|\underline{x},\underline{\xi})$. Such a model may be appropriate when the number of the market denotes its position in a ranking of markets. The model is written as

$$P(T\leq i|\underline{x},\lambda_i,\underline{\beta}) = \Phi\left(\lambda_i+\underline{\beta}^T\underline{x}\right), \qquad i=1,2,\ldots,N-1,$$

where $\underline{\beta}$ and $\underline{\lambda}$ are the parameters of the model and $\lambda_1\leq\lambda_2\leq\ldots\leq\lambda_{N-1}$. When the parameters of the model are assessed via a large sample on $(T,\underline{X})$ and relatively weak prior information,

$$\begin{pmatrix}\underline{\lambda}\\ \underline{\beta}\end{pmatrix} \sim \text{Normal}_{N+p-1}\left\{\begin{pmatrix}\underline{\ell}\\ \underline{b}\end{pmatrix}, V\right\}$$

where

$\underline{\ell}$ is the maximum likelihood estimate of $\underline{\lambda}$,

$\underline{b}$ is the maximum likelihood estimate of $\underline{\beta}$,

V is the inverse of Fisher's information matrix evaluated at $\begin{pmatrix}\underline{\ell}\\ \underline{b}\end{pmatrix}$.

A little algebra gives a similar result to Lemma 3.2,

$$P(T\leq i|\underline{x}) = \Phi\left(\frac{\ell_i+\underline{b}^T\underline{x}}{(1+s_i^2+2\underline{r}_i^T\underline{x}+\underline{x}^TS_\beta\underline{x})^{1/2}}\right),$$

where $s_i^2 = Var(\lambda_i)$, $S_\beta = Var(\underline{\beta})$ and $\underline{r}_i = Cov(\lambda_i,\underline{\beta})$. Under such a construction the predictive probability $P(T=i|\underline{x})$ follows easily. For $i=2,3,\ldots,N-1$,

$$P(T=i|\underline{x}) = \Phi\left(\frac{\ell_i+\underline{b}^T\underline{x}}{(1+s_i^2+2\underline{r}_i^T\underline{x}+\underline{x}^TS_\beta\underline{x})^{1/2}}\right) - \Phi\left(\frac{\ell_{i-1}+\underline{b}^T\underline{x}}{(1+s_{i-1}^2+2\underline{r}_{i-1}^T\underline{x}+\underline{x}^TS_\beta\underline{x})^{1/2}}\right).$$

The case in which $i = 1$ has

$$P(T = 1|\underline{x}) = \Phi\left(\frac{\ell_1 + \underline{b}^T\underline{x}}{(1 + s_1^2 + 2\underline{r}_1^T\underline{x} + \underline{x}^T S_\beta \underline{x})^{1/2}}\right)$$

and for $i = N$

$$P(T = N|\underline{x}) = 1 - \Phi\left(\frac{\ell_{N-1} + \underline{b}^T\underline{x}}{(1 + s_{N-1}^2 + 2\underline{r}_{N-1}^T\underline{x} + \underline{x}^T S_\beta \underline{x})^{1/2}}\right).$$

The optimal allocation regions given by (6.3) can now be analysed.

6.3 Other ideas

There are other, less well–developed ideas we have for research into topics important in screening. Here we discuss some of these ideas.

Sequential versus batch screening

In the context of the Conn's syndrome example, in section 4.5 we saw that the Bayes cost of the optimal sequential screen was more than that of the optimal design of a 'batch' two–stage screen that assumes no fixed form for the X–stage. However, there we assumed that the costs of performing the screen were negligible in comparison with misclassification costs. When the cost of screening is higher it is intuitive that a sequential screen will become more competitive as all the covariates will not need to be measured on all items. Work could be carried out to assess the cost levels at which a sequential screen becomes economic and the effects of model parameters on the choice of sequential or batch screen. Moskowitz & Tsai (1988) compare their sequential procedure with a batch screen for the case in which there are two covariates. Also, it may be optimal to mix sequential and batch screening. That is, one or more of the stages in the sequential screen may be based on two or more covariates.

Stopping and updating

In a sense, obtaining a sample on $(\underline{T}, \underline{X})$ is equivalent to operating a two-stage screen in which all items are passed to the second stage. One idea is to continue operating such a screen until it becomes economic to start implementing the $\underline{X}$-stage of the screen. That is, until enough is known about $(\underline{T}, \underline{X})$ for the $\underline{X}$-stage to be effective. Hence we could devise a rule that gives the optimal stage at which the $\underline{X}$-filter should come into operation. Once the $\underline{X}$-stage of the screen is implemented, the question becomes how to update the screen design with new data. A measurement on $\underline{X}$ is available for those items sentenced at the first stage of the screen and, for the remainder, a full measurement on $(\underline{T}, \underline{X})$ is taken. For the case in which T is binary, the diagnostic modelling approach factorises the likelihood in such a way that the updating scheme should be straightforward. All the observations on $\underline{X}$ can be used to update the parameters of the model for $\underline{X}$ and observations on T at those $\underline{X}$ values that pass the item to the next stage can be used to update the model for $T|\underline{X}$. The sampling scheme can be thought of as separate sampling schemes for $T|\underline{X}$ and $\underline{X}$.

Unusual observations

Suppose that at the first stage of the screen the variable $\underline{X}$ is measured for an item and the value obtained, $\underline{x}$, is unlike any of the observations on $\underline{X}$ contained in the training sample used for screen construction. What inference should be made about $\underline{T}$ and, in particular, what decision should be made about the item in a two-stage screen, accept it, reject it or measure $\underline{T}$? The intuitive answer to the latter question is that $\underline{T}$ should be measured, as we should be very uncertain about the item having not encountered an item like it in the past. However, an implicit assumption in the linear structure assumed in the probit regression model of Chapter 2 is that items with relatively high magnitudes for the components of $\underline{X}$ are of high or low quality, as appropriate. Some thought should be given to whether this is a realistic assumption in the context of future observations that might have very high or very low values of $\underline{x}$. If such observations are considered to be a potential problem, then

the sampling modelling approach may become more appealing as it is based on the densities $p(\underline{x}|T = i)$, $i = 1, 0$ and is likely to give the intuitive allocation of rogue observations to the second stage of the screen. These arguments are based on simple intuition and need to be researched further.

Eliciting costs

Throughout the thesis little thought has been given to the problem of assessing costs. It will usually be difficult to accurately assess relative costs especially when they are of a different type. For example, in medical applications the financial cost of performing an operation may need to be assessed relative to the moral, psychological and physical cost of misdiagnosing the illness of a patient. Here the procedure described in section 3.4 might be useful. A clinician may find it easier to think about a constraint on the proportion of patients that are passed to the second stage of the screen rather than a cost for operating the second stage. A good text that describes procedures for assessing utility (loss) functions is Smith (1988). In fact, Smith points out that unbounded loss functions, such as the latter two in (1.5) used by Tang (1987,1988b), can be problematical. The Bayes cost can be sensitive to small changes in the distribution of the random variables concerned and so information about the variables must be specified very accurately, something which is practically difficult. In the quality control literature methods of cost assessment for the loss functions in (1.5) are given by Hald (1960), Campanella & Corcoran (1982), Bhuyan (1982) and Taguchi (1984).

Dynamic models

In many medical screening programmes, patients may be subjected to a screen at a number of different times throughout their lifetime. One idea would be to include such a case history of measurements in the model used to construct the next screen. As well as the problem of designing optimal screens, other questions arise. For example, what is the optimal time for the next measurement of the screening variable? Uncertainty is

likely to increase as the time since the last measurement increases and so this question essentially asks, when will there be enough uncertainty so that another measurement of $\underline{X}$ is necessary?

Normal approximation

Throughout the thesis we have cited a Normal approximation to the posterior distribution of parameters in which the mean is given by the maximum likelihood estimate (m.l.e.) of the parameters and the variance is the inverse of Fisher's information matrix evaluated at the m.l.e.. For a finite sample size, this approach can be inefficient or wasteful of information. A method of inference based on the data alone can be seriously flawed if real prior information is available that is strong enough to contribute substantialy to that contained in the likelihood function. As an alternative one might use a Normal approximation in which the mean is given by the posterior mode of the parameters and the variance is given by the inverse of the *observed* information (that is, minus the second derivative of the log-posterior). This approximation results from a Taylor series expansion of the log-posterior about the posterior mode. For a further discussion of these issues see Chapter 4 of Gelman et al. (1995).

6.4 Concluding remarks

Here we add some final conclusions to the comments made throughout the thesis.

The idea of the two–stage screen described by Tang (1988b) is a simple one but Tang's assumption of known parameters is unrealistic. In Chapter 2 we have taken a Bayesian approach to modelling and have provided simple designs that are robust to modelling assumptions.

Chapter 3 presented a general result for the optimal design of a two–stage screen under misclassification costs. The theory can be used to design screens under any appropriate model for $(\underline{T}, \underline{X})$ and any loss functions for the costs. Also in Chapter 3,

we described optimal two-stage screen designs under a constraint on the proportion of items passed to the second stage of the screen. The result can easily be adapted to the case of constraints on the proportion of items accepted or rejected at the first stage of the screen. We also suggested that the limited resources screen might be useful for eliciting costs and for solving the problem of allocating items in a situation when a maximum of k out of a batch of m items should be passed to the second stage of the screen.

The sequential screen described by Chapter 4 provided a low cost alternative when the cost of measuring covariates is high. The heuristic design proposed makes use of the simple and robust designs obtained in Chapter 2 and is seen to perform well when compared with optimal designs.

In Chapter 5 we made progress towards assessing which covariates are best to include in a screen. We proposed some heuristics for choosing covariates and these heuristics were seen to perform well in a variety of situations.

Finally, in this chapter we have provided a number of suggestions for improvements and extensions to our work. There is much scope for further research.

Bibliography

Aitchison, J. & Begg, C.B. (1976). Statistical diagnosis when basic cases are not classified with certainty. *Biometrika*, **63**, 1—12.

Aitchison, J. & Dunsmore, I.R. (1975). *Statistical Prediction Analysis.* Cambridge: Cambridge University Press.

American Society for Quality Control (1987). *Qual. Prog.*, **20**, 66-68.

Anderson, J.A. & Blair, V. (1982). Penalized maximum likelihood estimation in logistic regression and discrimination. *Biometrika*, **69**, 123—136.

Bai, D.S. & Hong, S.H. (1992). Economic screening procedures using a correlated variable with multi-decision alternatives. *Naval Res. Logist.*, **39**, 471—485.

Bergman, S. & Gittins, J.C. (1985). *Statistical Methods for Pharmaceutical Research Planning.* New York: Marcel Dekker.

Bhuyan, S.K. (1982). Cost of quality as a customer perception. *Trans. ASQC Quality Congr., Detroit*, pp. 459—464.

Boys R.J. (1985), *Predictive Screening Methods.* PhD Thesis, University of Sheffield.

Boys, R.J. (1992). On a kernel approach to a screening problem. *J. R. Statist. Soc. B*, **54**, 157—159.

Boys, R.J. & Dunsmore, I.R. (1986). Screening in a normal model. *J. R. Statist. Soc. B*, **48**, 60—69.

Boys, R.J. & Dunsmore, I.R. (1987). Diagnostic and sampling models in screening. *Biometrika*, 74, 365—374.

Boys, R.J. & Glazebrook, K.D. (1992). A robust design of a screen for a binary response. *Biometrika*, 79, 643—650.

Campanella, J. & Corcoran, F.J. (1982). Principles of quality costs. *Trans. ASQC Quality Congr., Detroit*, pp. 124—137.

Chambers, J.M. (1973). Fitting nonlinear models: numerical techniques. *Biometrika*, 60, 1—13.

Collett, D. (1991). *Modelling Binary Data.* London: Chapman & Hall.

Copas, J.B. (1983). Plotting p against x. *Appl. Statist.*, 32, 25—31.

Dawid, A.P. (1976). Properties of diagnostic data distributions. *Biometrics*, 32, 647—658.

Dobson, A. J. (1983). *Introduction to Statistical Modelling.* London: Chapman & Hall.

Dunsmore, I.R. & Boys, R.J. (1987). Predictive screening methods in binary repsonse models. In *Probability and Bayesian Statistics* (ed. R. Viertl), pp. 151—158. New York: Plenum.

Dunsmore, I.R. & Boys, R.J. (1988). Global versus local screening. In *Bayesian Statistics 3* (Eds. J.M. Bernardo, M. DeGroot, D.V. Lindley & A.F.M. Smith), pp. 593—599. Oxford: Oxford University Press.

Gastwirth, J.L. (1987). The statistical precision of medical screening procedures: Application to polygraph and AIDS antibodies test data. *Statist. Sci.*, 2, 213—238.

Geisser, S. & Johnson, W. (1992). Optimal administration of dual screening tests

for detecting a characteristic with special reference to low prevalence diseases. *Biometrics*, 48, 839—852.

Gelman, A., Carlin, J.B., Stern, H.S., & Rubin, D.B. (1995). *Bayesian Data Analysis.* London: Chapman & Hall.

Gnanadesikan, R. (1977). *Methods for Statistical Data Analysis of Multivariate Observations.* New York: Wiley.

Greenes, R.A. & Begg, C.B. (1985). Assessment of diagnostic technologies: Methodology for unbiased estimation from samples of selective verified patients. *Invest. Radiol.*, 20, 751—756.

Hald, A. (1960). The compound hypergeometric distribution and a system of single sampling plans based on a prior distribution and costs. *Technometrics*, 2, 275—340.

Hammersley, J.M. & Handscomb, D.C. (1964). *Monte Carlo Methods.* London: Chapman & Hall.

Hand D.J. (1981). *Discrimination and Classification.* Chichester: Wiley.

Kim, S.B., Bai, D.S. (1990). Economic screening procedures in logistic and normal models. *Naval Res. Logist.*, 37, 919—928.

Lindley, D.V. (1961). The use of prior probability distributions in statistical inference and decisions. *Proc. 4th Berkeley Symp.*, 1, 453—468.

Liu, W, (1992). Predictive Screening. *Comm. Statist. - Theory & Methods*, 21, 2349—2366.

Moskowitz, H. & Tsai, H.T. (1988). A one–sided double screening procedure using individual unit misclassification error. *Manag. Sci.*, 34, 1139—1153.

Moskowitz, H., Plante, R. & Tsai, H.T. (1993). A multi–stage screening model for

evaluation and control of misclassification error in the detection of hypertension. *Manag. Sci.*, **39**, 307—321.

Naylor J.C. & Smith A.F.M. (1982). Application of a method for the efficient computation of posterior distributions. *Appl. Statist.*, **31**, 214—225.

N.A.G. (1990). *The Numerical Algorithms Group Fortran Library Manual, Mark 14.* Oxford: N.A.G. Ltd..

Owen, D.B. & Boddie, J.W. (1976) A screening method for increasing acceptable product with some parameters unknown. *Technometrics*, **18**, 195–199.

Owen, D.B., Li, L. & Chou, Y.M. (1981). Prediction intervals for screening using a measured correlated variable. *Technometrics*, **23**, 165—170.

Owen, D.B., McIntyre, D. & Seymour, E. (1975). Tables using one or two screening variables to increase acceptance product under one–sided specifications. *J. Qual. Tech.*, **7**, 127—138.

Owen, D.B. & Su, Y.H. (1977). Screening based on normal variables. *Technometrics*, **19**, 65–68.

Smith, J.Q. (1988). *Decision Analysis: A Bayesian approach.* London: Chapman & Hall.

Taguchi, G. (1984). *Quality Evaluation for Quality Assurance.* Romulus: American Supplier Institute.

Tang, J. & Tang, K. (1989). A two–sided screening procedure using several correlated variables. *IIE Trans.*, **21**, 333-336.

Tang, K. (1987). Economic design of a one–sided screening procedure using a correlated variable. *Technometrics*, **29**, 477—485.

Tang, K. (1988a). Economic design of a two–sided screening procedure using a correlated variable. *Appl. Statist.*, **32**, 231—241.

Tang, K. (1988b). Design of a two—stage screening procedure using correlated variables: a loss function approach. *Naval Res. Logist.*, **35**, 513—533.

Tang, K. & Schneider, H. (1990). Cost effectiveness of using a correlated variable in a complete inspection plan when inspection error is present. *Naval Res. Logist.*, **37**, 893–904.

Tang, K. & Tang, J. (1994). Design of screening procedures: a review. *J. Qual. Tech.*, **26**, 209—226.

Thomas, J.G., Owen, D.B. & Gunst, R.F. (1977). Improving the use of educational tests as a selection tool. *J. Educ. Stud.*, **2**, 55—79.

Tierney, L. & Kadane, J.B. (1986). Accurate approximations for posterior moments and marginal densities. *J. Am. Statist. Assoc.*, **81**, 82—86.

Tsai, H.T. & Moskowitz, H. (1986). A single screening procedure using individual misclassification error under one–sided specification. Institute Paper No. 896, Krannert School of Management, Purdue University, West Lafayette, Indiana.

Turkman, K.F., & Amaral Turkman, M.A. (1989). Optimal Screening Methods. *J. R. Statist. Soc. B*, **51**, 287—295.

Whittle, P. (1971). *Optimisation under constraints.* Chichester: Wiley.

Wong, A., Meeker, J. & Selwyn, M.R. (1985). Screening on correlated variables: a Bayesian approach. *Technometrics*, **27**, 423—432