Model Robust Designs for Binary Response Experiments

by

Shi-Hau Hwang

Advisor

Mong-Na Lo Huang

Department of Applied Mathematics
National Sun Yat-sen University
Kaohsiung, Taiwan, 804, R.O.C.

July, 2004

Abstract

The binary response experiments are often used in many areas. In many investigations, different kinds of optimal designs are discussed under an assumed model. There are also some discussions on optimal designs for discriminating models. The main goal in this work is to find an optimal design with two support points which minimizes the maximal probability differences between possible models from two types of symmetric location and scale families. It is called the minimum bias two-points design, or the mB_2 design in short here. D- and A-efficiencies of the mB_2 design obtained here are evaluated under an assumed model. Furthermore, when the assumed model is incorrect, the biases and the mean square errors in evaluating the true probabilities are computed and compared with that by using the D- and A-optimal designs for the incorrectly assumed model.

Keywords : Binary response, symmetric location and scale family, mB_2 design, D-efficiency, A-efficiency, bias, mean square error.

Contents

1 Introduction **1**
1.1 Preliminaries 1
1.2 Optimization criterion 3

2 The min-max results for two models **9**
2.1 The probit and logit case 9
2.2 General cases 18

3 Efficiencies and biases comparisons **20**
3.1 The probit and logit case 21
3.2 The probit and double reciprocal case 23

4 Discussions and conclusions **24**

Appendix **28**

A The convergence of MLEs for two-points designs with a misspecified link model **28**

B Properties of the scale function $\beta_{F,G}(p)$ and the distance function $D_{F,G}(b, x)$ **30**

C Figures of difference between two models **37**

D Tables for probit being the true model with logit link function **39**

E Tables for double-reciprocal being the true model with probit link **40**

F Some further works about mB_3 design for the probit and double reciprocal case **41**

1 Introduction

1.1 Preliminaries

A binary response experiment is one that the response variable y is 1 or 0. This response y depends on a random variable Z which may not be observed completely, and it is less than the predictor variable x if and only if the response y is 1. What we know only is whether the event $\{Z < x\}$ happens or not. It can be expressed by the following,

$$P\{y(x) = 1\} = P\{Z < x\} \quad \text{and} \quad P\{y(x) = 0\} = P\{Z \geq x\}.$$

We may not know exactly what the distribution of Z is, but may have the information that the distribution is continuous. The following four possible families have appeared in many literatures,

1. probit : $F_Z(x) = F_1(x; \mu, \sigma) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{(x-\mu)/\sigma} e^{\frac{-t^2}{2}} dt$
2. logit : $F_Z(x) = F_2(x; \mu, \sigma) = \dfrac{e^{\frac{x-\mu}{\sigma}}}{1 + e^{\frac{x-\mu}{\sigma}}}$
3. double exponential : $F_Z(x) = F_3(x; \mu, \sigma) = \displaystyle\int_{-\infty}^{(x-\mu)/\sigma} \frac{1}{2} e^{-|t|} dt$
4. double reciprocal : $F_Z(x) = F_4(x; \mu, \sigma) = \displaystyle\int_{-\infty}^{(x-\mu)/\sigma} \frac{1}{2(1+|t|)^2} dt$

where μ and σ are unknown parameters. We denote that $f_i = F_i'$ for $i = 1 \ldots 4$ in the following work. Among these models above, the probit model and the logit model are used most often. It is conventional that Z is from a symmetric location and scale family, that is,

$$F_Z(x; \mu, \sigma) = F(\frac{x-\mu}{\sigma}) \quad \text{and} \quad F_Z(x; \mu, \sigma) = 1 - F_Z(-x; \mu, \sigma)$$

for all $x \in \mathbf{R}$, where F is called the standard distribution of this family. Under this conventionality, there are a lot of investigations discussing the optimal designs for a binary response experiment with a given model. For example Wu (1985), Minkin (1987), Wu (1988), Khan and Yazdi (1988), Sitter and Wu (1993), Sitter and Fainaru (1997), Dette and Sahm (1997), and Mathew and Sinha (2001).

While doing a binary response experiment, it is of great interest in finding optimal designs for estimating the predicted variable x_p such that the response variable y is equal to 1 with probability p as stated in Wu (1988), i.e.

$$P(Z < x_p) = F_Z(x_p) = F(\frac{x_p - \mu}{\sigma}) = p,$$

and $x_p = \mu + \sigma F^{-1}(p)$ is an unknown parameter. Here, x_p is called the p^{th} quantile of this model. If we know what the model is, we can find estimations of μ and σ to estimate the p^{th} quantile x_p. The corresponding design problems with $p = \frac{1}{2}$ has already been discussed in earlier investigations because $x_{\frac{1}{2}}$ only depend on μ and does not depend on σ. When p is not equal to $\frac{1}{2}$, the optimal design in estimating x_p depends on the distribution of the assumed model. In some applications, it may request the estimation of some extreme quantile x_p with p close to 1. For example, the probability p of detonation in a pyrotechnics experiment is usually demanded to be 0.99 or even larger. The values of x_p will be quite different between distinct models as p close to 1. For instance, the value of $x_{0.999}$ of probit model with mean 0 and variance 1 is 3.09023 and $x_{0.999}$ of logit model with the same mean and variance is equal to 3.80789. These two points are quite far away from each other.

Suppose the design is with only two support points, say a_1 and a_2. If we have two possible models with distributions F and G having the same values at a_1 and a_2, i.e.

$$F(a_1) = G(a_1) \quad \text{and} \quad F(a_2) = G(a_2),$$

we will not be able to discriminate which model is true. In Chao and Fuh (1999), it is observed that if the model fitted has been misspecified, the estimation of an extreme quantile will not be consistent because of assuming an incorrect model. Uncertainty about the model has been an important issue in designing experiments. Many investigations have discussed the design problems on how to discriminate between models. See Atkinson and Fedorov (1975), Yanagisawa (1988), Yanagisawa (1990), and Muller and Ponce de Leon (1996) etc. If we use these methods, we usually need to do experiments at a lot of different design points for many times. It is not very efficient and economic to do so if the cost is high for each experiment.

Note that if the true model is either probit or logit, we may compare the probabilities of these two 0.999^{th} quantiles for probit and logit models with mean 0 and variance 1 are

$$F_1(3.09023;0,1) = 0.9990, \quad F_1(3.80789;0,1) = 0.9999, \quad \text{and}$$
$$F_2(3.09023;0,\frac{\pi}{\sqrt{3}}) = 0.9963, \quad F_2(3.80789;0,\frac{\pi}{\sqrt{3}}) = 0.9990.$$

It can be seen that although the 0.999^{th} quantiles for these two distributions may be far away, the difference of the probabilities at these two points is not very large in either one of the two cases. It is observed that even the quantile estimation is far from the true quantile, if there is no much difference between the probabilities of them, we accept that the quantile estimation is not bad.

1.2 Optimization criterion

Suppose the possible models are from two symmetric location and scale families with standard distributions F and G, and we do not know which one is the true model. If we choose the wrong one, it is impossible that the estimated p^{th} quantiles are consistent for all $p \in (0,1)$. Here we cite the definition of the distance between two distribution functions F and G is

$$d(F,G) = \sup_{x\in\mathbf{R}} |F(x) - G(x)|.$$

In this work, if the true model is with distribution $F(\,\cdot\,;\mu_0,\sigma_0)$ but we choose the other model with standard distribution G, we say that $G(\,\cdot\,;\mu_1,\sigma_1)$ is closer to $F(\,\cdot\,;\mu_0,\sigma_0)$ than $G(\,\cdot\,;\mu_2,\sigma_2)$ is if the distance between $F(\,\cdot\,;\mu_0,\sigma_0)$ and $G(\,\cdot\,;\mu_1,\sigma_1)$ is less or equal to the distance between $F(\,\cdot\,;\mu_0,\sigma_0)$ and $G(\,\cdot\,;\mu_2,\sigma_2)$, i.e.

$$\sup_{x\in\mathbf{R}} |F(x;\mu_0,\sigma_0) - G(x;\mu_1,\sigma_1)| \le \sup_{x\in\mathbf{R}} |F(x;\mu_0,\sigma_0) - G(x;\mu_2,\sigma_2)|.$$

So the distribution which is the closest to $F(\,\cdot\,;\mu_0,\sigma_0)$ from families with G is $G(\,\cdot\,;\mu^*,\sigma^*)$ where

$$(\mu^*,\sigma^*) = \arg \inf_{\mu\in\mathbf{R},\sigma>0} \sup_{x\in\mathbf{R}} |F(x;\mu_0,\sigma_0) - G(x;\mu,\sigma)|.$$

Since $F(\,\cdot\,;\mu_0,\sigma_0)$ is symmetric at μ_0 and $G(\,\cdot\,;\mu,\sigma)$ is symmetric at μ, therefore

$$\begin{aligned}
&\max\{|F(\mu_0+c;\mu_0,\sigma_0) - G(\mu_0+c;\mu,\sigma)|, |F(\mu_0-c;\mu_0,\sigma_0) - G(\mu_0-c;\mu,\sigma)|\}\\
&= \max\{|F(\frac{c}{\sigma_0}) - G(\frac{c+(\mu_0-\mu)}{\sigma})|, |F(\frac{-c}{\sigma_0}) - G(\frac{-c+(\mu_0-\mu)}{\sigma})|\}\\
&= \max\{F(\frac{c}{\sigma_0}) - G(\frac{c+(\mu_0-\mu)}{\sigma})|, |1-F(\frac{c}{\sigma_0}) - 1 + G(\frac{c+(\mu-\mu_0)}{\sigma})|\}\\
&= \max\{F(\frac{c}{\sigma_0}) - G(\frac{c+(\mu_0-\mu)}{\sigma})|, |F(\frac{c}{\sigma_0}) - G(\frac{c+(\mu-\mu_0)}{\sigma})|\}\\
&\ge |F(\frac{c}{\sigma_0}) - G(\frac{c}{\sigma})|, \quad \text{for all } \mu\in\mathbf{R} \text{ and } \sigma>0.
\end{aligned}$$

So we have the following inequality:

$$\sup_{x\in\mathbf{R}} |F(x;\mu_0,\sigma_0) - G(x;\mu_0,\sigma)| \le \sup_{x\in\mathbf{R}} |F(x;\mu_0,\sigma_0) - G(x;\mu,\sigma)|$$

for all $\mu \in \mathbf{R}$ and $\sigma > 0$. Then we have the result $\mu^* = \mu_0$ and we only need to find σ^* such that

$$\sigma^* = \arg\inf_{\sigma>0}\sup_{x\in\mathbf{R}} |F(x;\mu_0,\sigma_0) - G(x;\mu_0,\sigma)|.$$

Since F and G are from location and scale families, we can reduce the problem to finding $b > 0$ such that

$$\begin{aligned} b^* &= \arg\inf_{b>0}\sup_{x\in\mathbf{R}} |F(x;\mu_0,\sigma_0) - G(x;\mu_0,\frac{\sigma_0}{b})| \\ &= \arg\inf_{b>0}\sup_{x\in\mathbf{R}} |F(\frac{x-\mu_0}{\sigma_0}) - G(\frac{x-\mu_0}{\sigma_0/b})| \\ &= \arg\inf_{b>0}\sup_{x\in\mathbf{R}} |D_{F,G}(b,x)|, \end{aligned}$$

where $D_{F,G}(b,x) = F(x) - G(bx)$ is called the distance function for all $b > 0$ and $x \in \mathbf{R}$.

Two-points designs are the easiest kind of designs to decide two unknown parameters. When the possible models are from symmetric location and scale families, the best choices of two-points designs are symmetric, i.e. the support points are symmetric at μ and of the same weight. Suppose the support points are x_p and x_{1-p} for a percentage $p \neq \frac{1}{2}$ and there are two possible models from location and scale families of standard distribution F and G, which are agree some regularity conditions. First, if the true model is with distribution $F(\,\cdot\,;\mu_0,\sigma_0)$ but the assumed model is with distribution $G(\,\cdot\,,\mu_1,\sigma_1)$ where μ_1 and σ_1 are unknown, then $\{\hat{\mu}_{1,N},\hat{\sigma}_{1,N}\}$ which are the MLEs of $\{\mu_1,\sigma_1\}$ converges to $\{\mu_1,\sigma_1\}$ as the number of observations $N \to \infty$ (see Appendix A), i.e.

$$\lim_{N\to\infty}\hat{\mu}_{1,N} = \mu_1 = \frac{x_p + x_{1-p}}{2} = \mu_0, \quad \lim_{N\to\infty}\hat{\sigma}_{1,N} = \sigma_1 = \frac{x_p - x_{1-p}}{2G^{-1}(p)} = \frac{F^{-1}(p)}{G^{-1}(p)}\sigma_0.$$

We define the scale function $\beta_{F,G}$ as follows

$$\beta_{F,G}(p) = \begin{cases} \frac{G^{-1}(p)}{F^{-1}(p)} & \text{if } p \in (0,\frac{1}{2}) \cup (\frac{1}{2},1) \\ \lim_{p\to 1/2}\frac{G^{-1}(p)}{F^{-1}(p)} & \text{if } p = \frac{1}{2}. \end{cases}$$

For example, if the true model is probit and the assumed model is logit,

$$\beta_{F_1,F_2}(p) = \begin{cases} \frac{F_2^{-1}(p)}{F_1^{-1}(p)} & \text{if } p \in (0, \frac{1}{2}) \cup (\frac{1}{2}, 1) \\ \sqrt{\frac{8}{\pi}} & \text{if } p = \frac{1}{2}. \end{cases}$$

It is easy to show that

$$\beta_{F,G}(p) = \frac{1}{\beta_{G,F}(p)} \quad \text{for all } p \in (0, 1).$$

Therefore for all $q \in (0, 1)$, the true q^{th} quantile x_q is

$$F^{-1}(q; \mu_0, \sigma_0) = \mu_0 + \sigma_0 F^{-1}(q),$$

but the estimated q^{th} quantile through the assumed model as $N \to \infty$ is

$$\tilde{x}_q = \lim_{N\to\infty} G^{-1}(q; \hat{\mu}_{1,N}, \hat{\sigma}_{1,N}) = G^{-1}(q; \mu_1, \sigma_1) = \mu_0 + \frac{\sigma_0}{\beta_{F,G}(p)} G^{-1}(q).$$

The maximum difference between probabilities at x_q and $\tilde{x_q}$ for all $q \in (0, 1)$ is

$$\begin{aligned}
&\sup_{q\in(0,1)} |F(x_q; \mu_0, \sigma_0) - F(\tilde{x}_q; \mu_0, \sigma_0)| = \sup_{q\in(0,1)} |q - F(\frac{\tilde{x}_q - \mu_0}{\sigma_0})| \\
&= \sup_{q\in(0,1)} |G(\frac{\tilde{x}_q - \mu_1}{\sigma_1}) - F(\frac{\tilde{x}_q - \mu_0}{\sigma_0})| = \sup_{q\in(0,1)} |G(\frac{\tilde{x}_q - \mu_0}{\sigma_0/\beta_{F,G}(p)}) - F(\frac{\tilde{x}_q - \mu_0}{\sigma_0})| \\
&= \sup_{x\in\mathbf{R}} |G(\frac{x - \mu_0}{\sigma_0/\beta_{F,G}(p)}) - F(\frac{x - \mu_0}{\sigma_0})| = \sup_{x\in\mathbf{R}} |G(\beta_{F,G}(p)x) - F(x)| \\
&= \sup_{x\in\mathbf{R}} |F(x) - G(\beta_{F,G}(p)x)| = \sup_{x\in\mathbf{R}} |D_{F,G}(\beta_{F,G}(p), x)|.
\end{aligned}$$

If we care all quantile estimations fitted or not, we would like to find a symmetric two-points design $\xi^*_{F,G}$ which assigns equal weight on the $(1 - p^*_{F,G})^{th}$ and $(p^*_{F,G})^{th}$ quantiles such that the maximum difference between probabilities at x_q and $\tilde{x}_q$ for all $q \in (0, 1)$ may be minimized. That is, to find the design $\xi^*_{F,G}$ expressed as

$$\xi^*_{F,G} = \left\{ \begin{array}{cc} x_{1-p^*_{F,G}} & x_{p^*_{F,G}} \\ \frac{1}{2} & \frac{1}{2} \end{array} \right\}$$

where

$$p^*_{F,G} = \arg \inf_{p \in (0,1)} \sup_{x \in \mathbf{R}} |D_{F,G}(\beta_{F,G}(p), x)|.$$

Next, if the true model is with distribution $G(x; \mu_1, \sigma_1)$ but the assumed model is with $F(x; \mu_0, \sigma_0)$ where μ_0 and σ_0 are unknown, then we would like to find a symmetric two-points design $\xi^*_{G,F}$ with support points $\{x_{1-p^*_{G,F}}, x_{p^*_{G,F}}\}$ where

$$p^*_{G,F} = \arg \inf_{p \in (0,1)} \sup_{x \in \mathbf{R}} |D_{G,F}(\beta_{G,F}(p), x)|.$$

Since

$$D_{G,F}(\beta_{G,F}(p), x) = G(x) - F(\beta_{G,F}(p)x) = G(x) - F(\frac{x}{\beta_{F,G}(p)}) \quad \forall p \in (0,1),$$

therefore

$$\begin{aligned} \sup_{x \in \mathbf{R}} |D_{G,F}(\beta_{G,F}(p), x)| &= \sup_{x \in \mathbf{R}} |G(x) - F(\frac{x}{\beta_{F,G}(p)})| \\ &= \sup_{x \in \mathbf{R}} |G(\beta_{F,G}(p)x) - F(x)| = \sup_{x \in \mathbf{R}} |D_{F,G}(\beta_{F,G}(p), x)| \quad \forall p \in (0,1). \end{aligned}$$

It implies that $p^*_{G,F} = p^*_{G,F} = p^*$. That is, whenever F or G is the true model, the design which minimizes the maximum difference between the true model and the assumed model is the same. We denote this design $\xi^* = \xi^*_{F,G} = \xi^*_{G,F}$ to be the minimum bias two-points design (here it is called the mB_2 design in short) for models from symmetric location and scale families with standard distributions F and G.

After the introduction about the mB_2 designs above, this paper is organized as follows. In Section 2, we exhibit some min-max results concerning about model estimation and related design problems for the two models. We will study some characteristics of the mB_2 design points for probit-logit case, and later use them to give a numerical method to find the mB_2 design for any two possible models which are from symmetric location and scale families. In section 3, numerical and simulation results are presented under the following two cases, the probit-logit case with mean 0 and variance 1 and the F_1-F_4 case with location parameter 0 and scale parameter 1. D- and A-efficiencies of the mB_2 design are evaluated under an assumed model. Furthermore, when the assumed model is incorrect, the biases and the mean square errors in evaluating the true probabilities with the estimated quantiles are computed and compared with that by using the D- and A-optimal designs for the wrongly assumed model. Finally, we give some discussions and conclusions in section 4.

2 The min-max results for two models

As mentioned earlier, F and G are two standard distributions assumed to be from location and scale families, the distance function $D_{F,G}(b,x)$ and the scale function $\beta_{F,G}(p)$ have some good properties, like symmetry and continuously differentiable, which are expressed in Appendix B. According to these properties, the mB_2 design points is $\{x_{1-p^*_{F,G}}, x_{p^*_{F,G}}\}$, where

$$p^*_{F,G} = \inf_{p\in[\frac{1}{2},1)} \sup_{x>0} D_{F,G}(\beta_{F,G}(p), x).$$

Particularly, by the special properties for probit-logit case which are presented in Appendix B, the mB_2 design points for this case is $\{x_{1-p^*}, x_{p^*}\}$, where $\beta_{F_1,F_2}(x_{p^*}) = b^*$ and

$$b^* = \inf_{b>\sqrt{8/\pi}} \sup_{x>0} D_{F_1,F_2}(b,x).$$

Since $\lim_{x\to\infty} D_{F,G}(b,x) = 0$ and $\lim_{x\to 0^+} D_{F,G}(b,x) = 0$ for all $b > 0$, therefore the maximizer M_b of $|D_{F,G}(b,x)|$ must be a critical point, i.e.

$$\frac{\partial}{\partial x} D(b,x)|_{x=M_b} = 0.$$

We first study some characteristics of the critical points of $D(b,x)$ for all $b > \sqrt{\frac{8}{\pi}}$ for the probit-logit case, and then we use these to find the mB_2 design for probit and logit models. For the general cases, we only provide a numerical method to find the mB_2 designs when the standard distributions are two of the four distributions, i.e. F_1, F_2, F_3, and F_4, as mentioned in Section 1.

2.1 The probit and logit case

In this subsection, we will search for all critical points of $D(b,x)$ for all $b > \sqrt{\frac{8}{\pi}}$ in probit-logit case, then find the infimum of $\{\sup_{x>0} |D(b,x)| : b \geq \sqrt{\frac{8}{\pi}}\}$. It is observed according Figure 2 in Appendix C that there are two extreme points of $D(\beta(p), x)$ on $\{x : x > 0\}$ for almost $p \in (0,1)$, the 1^{st} extreme point is the minimum, the 2^{nd} one is the maximum, and they are both decreasing as b increases. The following three lemmas provide theoretical verifications of the above results.

Lemma 2.1.

For all $b > \sqrt{\frac{8}{\pi}}$, there are two critical points of $D(b, x)$ on $\{x : x > 0\}$. For $b_0 = \sqrt{\frac{8}{\pi}}$, there is one critical point of $D(b_0, x)$ on $\{x : x > 0\}$.

Proof.

For all $b \geq \sqrt{\frac{8}{\pi}}$, there is 1 or 2 critical points of $D(b, x) = F_1(x) - F_2(bx)$ on $\{x : x > 0\}$ if and only if there is 1 or 2 roots of $\frac{\partial}{\partial x} D(b, x) = f_1(x) - bf_2(bx)$ on $\{x : x > 0\}$. Now, we will solve this equation on $\{x : x > 0\}$ for all $b \geq \sqrt{\frac{8}{\pi}}$.

$$
\begin{aligned}
& \frac{1}{\sqrt{2\pi}} e^{-x^2/2} - \frac{be^{bx}}{(1+e^{bx})^2} = 0 \\
\Leftrightarrow \quad & (1+e^{bx})^2 = \sqrt{2\pi} b e^{bx} e^{x^2/2} \\
\Leftrightarrow \quad & 2\ln(1+e^{bx}) = \ln(b) + \ln(2\pi)/2 + x^2/2 + bx \\
\Leftrightarrow \quad & \psi(b, x) = \frac{x^2}{2} - bx + \ln(b) + \frac{\ln(2\pi)}{2} - 2\ln(1+e^{-bx}) = 0.
\end{aligned} \tag{2.1}
$$

By (2.1),

$$
\begin{aligned}
\frac{\partial}{\partial x}\psi(b, x) &= x - b + \frac{2b}{1+e^{-bx}}, \\
\frac{\partial^2}{\partial x^2}\psi(b, x) &= 1 - \frac{2b^2 e^{bx}}{(1+e^{-bx})^2}.
\end{aligned} \tag{2.2}
$$

By (2.2), we let $x_{b,1} = \frac{1}{b}\ln(b^2 - 1 + b\sqrt{b^2-2})$ for all fixed $b \geq \sqrt{\frac{8}{\pi}}$, then

$(A1)$ $\dfrac{\partial}{\partial x}\psi(b,x)$ is strictly decreasing for all $x \in (0, x_{b,1})$ since $\dfrac{\partial^2}{\partial x^2}\psi(b,x) < 0$ there.

$(A2)$ $\psi(x_{b,1})$ is the minimum since $\dfrac{\partial^2}{\partial x^2}\psi(b,x)|_{x=x_{b,1}} = 0$.

$(A3)$ $\dfrac{\partial}{\partial x}\psi(b,x)$ is strictly increasing for all $x > x_{b,1}$ since $\dfrac{\partial^2}{\partial x^2}\psi(b,x) > 0$ there.

Since $\frac{\partial}{\partial x}\psi(b,x) = 0$ as $x \to 0^+$ and $(A1)$, therefore $\frac{\partial}{\partial x}\psi(b,x_{b,1}) < 0$ for all $b \geq \sqrt{\frac{8}{\pi}}$. With $(A1)$ to $(A3)$ and by the fact that $\frac{\partial}{\partial x}\psi(b,x) \to \infty$ as $x \to \infty$, it can be obtained that:

$(B1)$ There exists a unique $x_{b,2} \in (x_{b,1}, \infty)$ such that $\dfrac{\partial}{\partial x}\psi(b,x)|_{x=x_{b,2}} = 0$.

$(B2)$ $x_{b,2}$ is the minimizer of $\psi(b,x)$ since $\dfrac{\partial^2}{\partial x^2}\psi(b,x)|_{x=x_{b,2}} > 0$.

$(B3)$ $\psi(b,x)$ is strictly decreasing $\forall x \in (0, x_{b,2})$ and is strictly increasing $\forall x > x_{b,2}$.

Since $\psi(\sqrt{\frac{8}{\pi}}, x) = 0$ and $\psi(b,x) > 0$ as $x \to 0^+$ for all $b > \sqrt{\frac{8}{\pi}}$, $\psi(b,x) \to \infty$ as $x \to \infty$ for all $b \geq \sqrt{\frac{8}{\pi}}$, and $(B1)$ to $(B3)$, if we can show that $\psi(b, x_{b,2}) < 0$, then it implies that

$(C1)$ For all $b > \sqrt{\frac{8}{\pi}}$, $\psi(b,x)$ has a unique root in $(0, b_{b,2})$ and a unique root in $(b_{b,2}, \infty)$.

$(C2)$ For $b = \sqrt{\frac{8}{\pi}}$, $\psi(b,x)$ has no roots in $(0, b_{b,2})$ and a unique root in $(b_{b,2}, \infty)$.

Although we do not know what $x_{b,2}$ is for all $b \geq \sqrt{\frac{8}{\pi}}$, if we can show that there is a $x_{b,3} > 0$ for all $b \geq \sqrt{\frac{8}{\pi}}$ which satisfies $\psi(b, x_{b,3}) < 0$, then we have proved that $\psi(b, x_{b,2}) \leq \psi(b, x_{b,3}) < 0$.

Let $x_{b,3} = b$ for all $b \geq \sqrt{\frac{8}{\pi}}$, then it is easily shown that

$$\begin{aligned}
\psi(b, x_{b,3}) &= \psi(b,b) = \frac{b^2}{2} - b^2 + \ln(b) + \frac{\ln(2\pi)}{2} - 2\ln(1+e^{-b^2}) \\
&= \frac{b^2}{2} + b^2 + \ln(b) + \frac{\ln(2\pi)}{2} - 2\ln(1+e^{b^2}) \\
&\leq \frac{b^2}{2} + \ln(b) + \frac{\ln(2\pi)}{2} - 2\ln(e^{b^2}) \\
&= \frac{-b^2}{2} + \ln(b) + \frac{\ln(2\pi)}{2} < 0 \quad \text{for all } b \geq \sqrt{\tfrac{8}{\pi}}.
\end{aligned}$$

Then we have proved that there exists a point $r_{b,1} \in (0, x_{b,2})$ for all $b > \sqrt{\frac{8}{\pi}}$ and a point $r_{b,2} \in (x_{b,2}, \infty)$ for all $b \geq \sqrt{\frac{8}{\pi}}$ such that

$$f(r_{b,1}) - bg(br_{b,1}) = 0 \quad \text{and} \quad f(r_{b,2}) - bg(br_{b,2}) = 0. \tag{2.3}$$

That is, there are exact two extreme points $r_{b,1}$ and $r_{b,2}$ for $D(b,x)$ on $\{x : x > 0\}$ for all $b > \sqrt{\frac{8}{\pi}}$ and there is a unique extreme point $r_{b,2}$ for $D(\sqrt{\frac{8}{\pi}}, x)$. □

Lemma 2.2.

Let $R_1(b) = r_{b,1}$ for all $b > \sqrt{\frac{8}{\pi}}$ and $R_2(b) = r_{b,2}$ for all $b \geq \sqrt{\frac{8}{\pi}}$, where $r_{b,1}$ and $r_{b,2}$ are defined in the proof of Theorem 3.1. Then R_1 and R_2 are continuously differentiable on $(\sqrt{\frac{8}{\pi}}, \infty)$, and R_2 is continuous on $[\sqrt{\frac{8}{\pi}}, \infty)$.

Proof.

Since

$$\psi(b,x) = \frac{x^2}{2} - bx + \ln(b) + \frac{\ln(2\pi)}{2} - 2\ln(1+e^{-bx})$$

is a continuously differentiable function on $\{(b,x) : b \geq \sqrt{\frac{8}{\pi}}, x > 0\} \subseteq \mathbf{R}^2$. Let

$$\psi_1(b,x) = (b, \psi(b,x)) \quad \text{on} \quad D_1 = \{(b,x) : b > \sqrt{\frac{8}{\pi}}, x \in (0, x_{b,2})\}$$

$$\psi_2(b,x) = (b, \psi(b,x)) \quad \text{on} \quad D_2 = \{(b,x) : b \geq \sqrt{\frac{8}{\pi}}, x > x_{b,2}\}$$

where $x_{b,2}$ is defined in Lemma 3.1.

First we will prove that $R_1(b)$ is continuously differentiable on its domain. Since the Jacobian determinant of $\psi_1(b,x)$

$$|J_1| = \begin{vmatrix} 1 & \frac{\partial}{\partial b}\psi_1(b,x) \\ 0 & \frac{\partial}{\partial x}\psi_1(b,x) \end{vmatrix} = \frac{\partial}{\partial x}\psi(b,x) < 0$$

for all $b, x \in D_1$, and D_1 is an open set in $\mathbf{R}^2$. Therefore, by the Inverse Function Theorem, ψ_1^{-1} exists and is continuously differentiable on $\psi_1(D_1)$. Since $\{(b,0) : b > \sqrt{\frac{8}{\pi}}\} \subseteq \psi_1(D_1)$, therefore $\psi_1^{-1}(b,0) = (b, R_1(b))$ is continuously differentiable on $\psi_1(D_1)$, and then we get that $R_1(b) = r_{b,1}$ is a continuously differentiable function on $\{b : b > \sqrt{\frac{8}{\pi}}\}$.

Similarly, since the Jacobian determinant of $\psi_2(b,x)$

$$|J_2| = \begin{vmatrix} 1 & \frac{\partial}{\partial b}\psi_1(b,x) \\ 0 & \frac{\partial}{\partial x}\psi_1(b,x) \end{vmatrix} = \frac{\partial}{\partial x}\psi(b,x) > 0$$

for all $b, x \in D_2'$, where $D_2' = \{(b,x) : b > \sqrt{\frac{8}{\pi}}, x > x_{b,2}\}$ is the interior set of D_2. Therefore, by the Inverse Function Theorem, ψ_2^{-1} exists and is continuously differentiable on $\psi_2(D_2')$. Since $\{(b,0) : b > \sqrt{\frac{8}{\pi}}\} \subseteq \psi_2(D_2')$, therefore $\psi_2^{-1}(b,0) = (b, R_2(b))$ is continuously differentiable on $\psi_2(D_2')$, and then we obtain that $R_2(b) = r_{b,2}$ is a continuously differentiable function on $\{b : b > \sqrt{\frac{8}{\pi}}\}$.

Finally, since $\psi_2 : D_2 \to \mathbf{R}^2$ is a continuous function. If $\psi_2 : D_2 \to \mathbf{R}^2$ is a one-to-one function, then $\psi_2^{-1} : \psi_2(D_2) \to \mathbf{R}^2$ exists and is continuous on D_2. If $\psi_2(b_1,x_1) = \psi_2(b_2,x_2)$, then $b_1 = b_2$, and $\psi(b_1,x_1) = \psi(b_1,x_2)$. Since for all fixed $b \geq \sqrt{(\frac{8}{\pi})}$, $\frac{\partial}{\partial x}\psi(b,x) > 0$ for all $x > x_{b,2}$, and $x_1, x_2 > x_{b_1,2}$. Therefore $\psi(b_1,x_1) = \psi(b_1,x_2)$ if and only if $x_1 = x_2$. It implies that $\psi_2 : D_2 \to \mathbf{R}^2$ is one-to-one and then we get that $\psi_2^{-1} : \psi_2(D_2) \to \mathbf{R}^2$ exists and is continuous. Since $\{(b,0) : b \geq \sqrt{\frac{8}{\pi}}\} \subseteq \psi_2(D_2)$, therefore $\psi_2^{-1}(b,0) = (b, R_2(b))$ is continuous on $\psi_2(D_2)$, and then we get that $R_2(b) = r_{b,2}$ is a continuous function on $\{b : b \geq \sqrt{\frac{8}{\pi}}\}$. $\square$

Lemma 2.3.

$D(b, R_1(b))$ is negative and strictly decreasing on $\{b : b > \sqrt{\frac{8}{\pi}}\}$. $D(b, R_2(b))$ is positive and strictly decreasing on $\{b : b \geq \sqrt{\frac{8}{\pi}}\}$.

Proof.

For all $b > \sqrt{\frac{8}{\pi}}$,

$$\begin{aligned}
\frac{\partial}{\partial b} D(b, R_1(b)) &= \frac{\partial}{\partial b}[F_1(R_1(b)) - F_2(bR_1(b))] \\
&= f_1(R_1(b)) \times \frac{\partial R_1(b)}{\partial b} - f_2(bR_1(b)) \times (R_1(b) + b\frac{\partial R_1(b)}{\partial b}) \\
&= \frac{\partial R_1(b)}{\partial b}[f_1(R_1(b)) - bf_2(bR_1(b))] - f_2(bR_1(b))R_1(b) \quad \text{(by (2.3))} \\
&= -f_2(bR_1(b))R_1(b) < 0,
\end{aligned}$$

and similarly by (2.3),

$$\frac{\partial}{\partial b} D(b, R_2(b)) < 0.$$

Since $\frac{\partial}{\partial b} D(b, R_1(b)) < 0$ and $\frac{\partial}{\partial b} D(b, R_2(b)) < 0$ for all $b > \sqrt{\frac{8}{\pi}}$, therefore $D(b, R_1(b))$ and $D(b, R_2(b))$ are strictly decreasing on $\{b : b > \sqrt{\frac{8}{\pi}}\}$. Since $D(b, R_2(b))$ is continuous on $\{b : b \geq \sqrt{\frac{8}{\pi}}\}$, therefore $D(b, R_2(b))$ is strictly decreasing there.

Next, we will show that $D(b, R_1(b))$ is negative and $D(b, R_1(b))$ is positive. By the definition of $\psi(b, x)$,

$$\begin{aligned}
\psi(b, x) > 0 \quad &\Leftrightarrow \quad \ln(\frac{bf_2(bx)}{f_1(x)}) > 0 \Leftrightarrow \frac{bf_2(bx)}{f_1(x)} > 1 \\
&\Leftrightarrow \quad \frac{\partial}{\partial x} D(b, x) = f_1(x) - bf_2(bx) < 0,
\end{aligned}$$

$\frac{\partial}{\partial x}\psi(b,x) < 0$ for all $x < R_1(b) < x_{b,2}$, and $\psi(b, R_1(b)) = 0$, we can deduce that $\psi(b,x) > 0$ for all $x \in (0, R_1(b))$. So we know that $\frac{\partial}{\partial x}D(b,x) < 0$ for all $x \in (0, R_1(b))$. Since $\lim_{x\to 0^+} D(b,x) = 0$ for all $b > \sqrt{\frac{8}{\pi}}$, therefore we have proved that

$$D(b, R_1(b)) < \lim_{x\to 0^+} D(b,x) = 0, \quad \forall b > \sqrt{\frac{8}{\pi}}.$$

Similarly, since $\frac{\partial}{\partial x}\psi(b,x) > 0$ for all $x > R_2(b) > x_{b,2}$ and $\psi(b, R_2(b)) = 0$, therefore $\psi(b,x) > 0$ and $\frac{\partial}{\partial x}D(b,x) < 0$ for all $x > R_2(b)$. Since $\lim_{x\to\infty} D(b,x) = 0$ for all $b \geq \sqrt{\frac{8}{\pi}}$, therefore we have proved that

$$D(b, R_2(b)) > \lim_{x\to\infty} D(b,x) = 0, \quad \forall b \geq \sqrt{\frac{8}{\pi}}.$$

□

By Lemma 2.1 and Lemma 2.3, it is easy to know that :

$$\sup_{x>0} |D(b,x)| = \max\{|D(b, R_1(b))|, |D(b, R_2(b))|\} \quad \forall b > \sqrt{\frac{8}{\pi}}, \tag{2.4}$$

$$\sup_{x>0} |D(\sqrt{\frac{8}{\pi}}, x)| = |D(\sqrt{\frac{8}{\pi}}, R_2(\sqrt{\frac{8}{\pi}}))|, \tag{2.5}$$

for probit-logit case. According to (2.4) and (2.5), Table 1 is obtained by numerical computation to get more information about the maximum and the maximizer of $D(\beta(p), x)$ for all $p \geq 0.5$.

According to Table 1, the maximizer is $R_1(b)$ for some b and is $R_2(b)$ for the others. We can see that $\max_{x>0} |D(b,x)|$ is decreasing at first and then increasing, and the maximizer is $R_2(b)$ at first and then is $R_1(b)$. By the property of $\max_{x>0} D|(b,x)|$ above, we have the following theorem to tell us where the mB_2 design points are.

Table 1: $\max_{x>0} |D(\beta(p), x)|$ for some p

p	$b = \beta(p)$	$R_1(b)$	$\lvert D(b, R_1(b))\rvert$	$R_2(b)$	$\lvert D(b, R_2(b))\rvert$	maximum	maximizer
0.5	1.5958	–	–	1.7318	0.01767	0.01767	$R_2(1.5958)$
0.55	1.5969	0.0724	$< 10^{-5}$	1.7350	0.01756	0.01756	$R_2(1.5969)$
0.7	1.6158	0.2912	0.00095	1.7881	0.01581	0.01581	$R_2(1.6158)$
0.8	1.6472	0.4397	0.00362	1.8797	0.01319	0.01319	$R_2(1.6472)$
0.88	1.6957	0.5605	0.00877	2.0249	0.00982	0.00982	$R_2(1.6957)$
0.89	1.7046	0.5758	0.00978	2.0517	0.00929	0.00978	$R_1(1.7046)$
0.9	1.7145	0.5911	0.01092	2.0814	0.00872	0.01092	$R_1(1.7145)$
0.95	1.7901	0.6681	0.01984	2.3058	0.00530	0.01984	$R_1(1.7901)$
0.99	1.9753	0.7292	0.04145	2.8223	0.00139	0.04145	$R_1(1.9753)$
0.9999	2.4765	0.7115	0.09185	4.0519	0.00002	0.09185	$R_1(2.4765)$

Theorem 2.4.

There exists a unique $b^* > \sqrt{\frac{8}{\pi}}$ *such that* $|D(b^*, R_1(b^*))|$ *is equal to* $|D(b^*, R_2(b^*))|$ *and*

$$\sup_{x>0} |D(b^*, x)| = \inf_{b \geq \sqrt{8/\pi}} \sup_{x>0} |D(b, x)|$$

Proof.

First, we prove that there exists a unique b^* uniquely. As $D(b^*, R_1(b^*))$ and $D(b^*, R_2(b^*))$ are strictly decreasing on $\{b : b > \sqrt{\frac{8}{\pi}}\}$, $D(b^*, R_1(b^*)) < 0$, and $D(b^*, R_2(b^*)) > 0$ for all $b > \sqrt{\frac{8}{\pi}}$, therefore ,by the proof of Lemma 3.3, we have that

$$|D(b, R_1(b))| - |D(b, R_2(b))| = -D(b, R_1(b)) - D(b, R_2(b))$$

is a strictly increasing function. Since

$$|D(1.6, R_1(1.6))| - |D(1.6, R_2(1.6))| < 0 \quad \text{and}$$

$$|D(1.8, R_1(1.8))| - |D(1.8, R_2(1.8))| > 0,$$

therefore there exists a unique $b^* \in (1.6, 1.8)$ such that

$$|D(b^*, R_1(b^*))| - |D(b^*, R_2(b^*))| = 0, \quad \text{i.e. } |D(b^*, R_1(b^*))| = |D(b^*, R_2(b^*))|.$$

and

$$|D(b, R_1(b))| < |D(b, R_2(b))| \quad \forall b \in (\sqrt{\frac{8}{\pi}}, b^*)$$
$$|D(b, R_1(b))| > |D(b, R_2(b))| \quad \forall b \in (b^*, \infty).$$

Finally, we will show that b^* is the minimizer of $\sup_{x>0} |D(b,x)|$ for all $b \geq \sqrt{\frac{8}{\pi}}$. Since

$$|D(b, R_1(b))| < |D(b, R_2(b))| \quad \forall b \in (\sqrt{\frac{8}{\pi}}, b^*),$$
$$|D(b, R_1(b))| > |D(b, R_2(b))| \quad \forall b \in (b^*, \infty),$$
$$|D(b, R_1(b))| \text{ is strictly increasing on } \{b : b > \sqrt{\frac{8}{\pi}}\},$$
$$|D(b, R_2(b))| \text{ is strictly decreasing on } \{b : b \geq \sqrt{\frac{8}{\pi}}\}, \text{ and}$$
$$\sup_{x>0} |D(\sqrt{\frac{8}{\pi}}, x)| = |D(\sqrt{\frac{8}{\pi}}, R_2(\sqrt{8/\pi}))|.$$

Therefore,

$$\sup_{x>0} |D(b,x)| = |D(b, R_2(b))| > |D(b^*, R_2(b^*))| = \max_{x>0} |D(b^*, x)| \quad \forall b \in [\sqrt{\frac{8}{\pi}}, b^*)$$
$$\sup_{x>0} |D(b,x)| = |D(b, R_1(b))| > |D(b^*, R_1(b^*))| = \max_{x>0} |D(b^*, x)| \quad \forall b \in (b^*, \infty)$$

that is,

$$\max_{x>0} |D(b^*, x)| = \inf_{b \geq \sqrt{8/\pi}} \sup_{x>0} |D(b,x)|.$$

$\square$

By the theorem above and after using *Mathematica* for some computation, we obtain that:

$$b^* = 1.7017,$$

$$p^* = 0.8869 \quad \text{which satisfies } \beta(p^*) = b^* \text{ and } p \geq \tfrac{1}{2}, \text{ and}$$

$$\inf_{b \geq \sqrt{8/\pi}} \sup_{x>0} |D(b,x)| = \max_{x>0} |D(b^*, x)| = 0.00946.$$

It means that the mB_2 design for probit and logit models is $\{x_{1-p^*}, x_{p^*}\}$ where x_{p^*} denotes the $(p^*)^{th}$ quantile of the true model. Moreover, if we choose the wrong model, the maximum difference between the probabilities of a quantile and its estimation is 0.00946.

2.2 General cases

For the general cases with the standard distributions being any two of F_1, F_2, F_3, and F_4, we only provide some numerical results. Note that

$$\delta_{F,G}(p) = \sup_{x>0} D_{F,G}(\beta_{F,G}(p), x) + \inf_{x>0} D_{F,G}(\beta_{F,G}(p), x).$$

The figures in Appendix C show the similarities of the properties in these cases. For instance, when $p > \frac{1}{2}$, $\delta(p)$ is strictly monotone, and the maximum of $|D(b, x)|$ is smaller as $\delta(p)$ is closer to 0. Now the question become to find the unique root of $\delta(p)$ for $p > \frac{1}{2}$. In many works, Newton's Method is used to find roots, but it is not useful here since the derivatives of $\delta(p)$ is hard to solved. We provide a numerical bisection method to find the corresponding mB_2 design for arbitrary two models with standard distributions F and G, where F and G are two of F_1, F_2, F_3, or F_4.

Step 1: Find $p_{2,0} > p_{1,0} > \frac{1}{2}$ such that

$$\delta_{F,G}(p_{1,0}) \times \delta_{F,G}(p_{2,0}) \leq 0.$$

Step 2: Let $p_{3,n} = \frac{p_{1,n}+p_{2,n}}{2}$, and

$$(p_{1,n+1}, p_{2,n+1}) = \begin{cases} (p_{1,n}, p_{3,n}) & \text{if } \delta_{F,G}(p_{1,0}) \times \delta_{F,G}(p_{3,0}) \leq 0, \\ (p_{3,n}, p_{2,n}) & \text{otherwise.} \end{cases}$$

Step 3: Repeat Step 2 until $p_{2,n} - p_{1,n}$ is small enough. Let $p_n^* = \frac{p_{2,n}+p_{1,n}}{2}$. Then the approximate mB_2 design is ξ_n^*

$$\xi_n^* = \left\{ \begin{array}{cc} x_{1-p_n^*} & x_{p_n^*} \\ \frac{1}{2} & \frac{1}{2} \end{array} \right\}.$$

By using this method, we find the approximate mB_2 designs in Table 2 when the possible models are from two of the four families with standard distributions F_1, F_2, F_3, or F_4.

Table 2: The mB_2 designs for some cases

Distributions	p^*	$b^* = \beta(p^*)$	$\max_{x\in\mathbf{R}} D(b^*, x)$
F_1, F_2	0.8869	1.7017	0.00946
F_1, F_3	0.8386	1.1437	0.02821
F_1, F_4	0.8362	2.0963	0.07524
F_2, F_3	0.8126	0.6690	0.01978
F_2, F_4	0.8293	1.2204	0.06702
F_3, F_4	0.8423	1.8808	0.05095

The method above is just a way to find the mB_2 design. If the choices of $p_{1,0}$ and $p_{2,0}$ are both close to p^*, then this method is quite efficient. From the numerical computations, it indicates that p^* is usually in $(0.8, 0.9)$. If there are any other two distributions which have similar properties as the above cases, it is advised to choose $b_{1,0} = 0.8$ and $b_{2,0} = 0.9$ at first.

3 Efficiencies and biases comparisons

Suppose that there is a design ξ,

$$\xi = \left\{ \begin{array}{cccc} x_1 & x_2 & \dots & x_n \\ p_1 & p_2 & \dots & p_n \end{array} \right\} \quad \text{, where } p_i > 0 \ \forall i = 1, 2, \dots n \text{ and } \sum_{i=1}^{n} p_i = 1,$$

and the model is from a symmetric location and scale family with standard distribution F. Then the information matrix for estimating μ and σ is

$$M = \sum_{i=1}^{n} \frac{p_i}{F_i(1-F_i)} \begin{pmatrix} (\frac{\partial F_i}{\partial \mu})^2 & (\frac{\partial F_i}{\partial \mu})(\frac{\partial F_i}{\partial \sigma}) \\ (\frac{\partial F_i}{\partial \mu})(\frac{\partial F_i}{\partial \sigma}) & (\frac{\partial F_i}{\partial \sigma})^2 \end{pmatrix} \quad \text{, where } F_i = F(\frac{x_i - \mu}{\sigma}).$$

Note that the D- and A-efficiencies of design ξ is

$$D\text{-efficiency}(\xi) = \frac{det(M_\xi)}{det(M_D)}, A\text{-efficiency}(\xi) = \frac{tr(M_A^{-1})}{tr(M_\xi^{-1})},$$

where M_D and M_A are the information matrices of the D- and A-optimal designs respectively under the true model.Let ξ_{D_i} and ξ_{A_i} are the D- and A-optimal designs respectively, and D_i- and A_i-efficiencies are D- and A-efficiencies respectively when the true model is with standard distribution F_i, $i = 1 \dots 4$. We compare the mB_2 design with the D- and A- optimal designs under the D- and A-optimality criteria. Next, we do some numerical works and simulations. First, we compute the probability bias of q^{th} quantile as the number of observations $N \to \infty$, i.e.

$$bias_\infty(q) = q - F(G^{-1}(q; \mu_0; \frac{\sigma_0}{\beta_{F,G}(p)}); \mu_0, \sigma_0),$$

where $F(\ \cdot\ ; \mu_0, \sigma_0)$ is the true model, and $G(\ \cdot\ ; \mu_0; \frac{\sigma_0}{\beta_{F,G}(p)})$ is the assumed model, where $\{x_{1-p}, x_p\}$ are the design points. Also, we point out the maximum and the minimum of $bias_\infty(q)$ for all $q > 0.5$, say q_M and q_m respectively, from each design to compare the maximum probability bias. Next, we simulate 1000 times with 1000 observations each time to compute the probability biases and mean square errors of some q^{th} quantiles. That is, for $q \in (\frac{1}{2}, 1)$, the bias and MSE are computed by

$$bias_{1000}(q) = q - \frac{1}{1000}\sum_{n=1}^{1000} F(\tilde{x}_{q,n}),$$
$$MSE_{1000}(q) = \frac{1}{1000}\sum_{n=1}^{1000}(q - F(\tilde{x}_{q,n}))^2,$$

where $\tilde{x}_{q,n} = \hat{\mu} + \hat{\sigma}G^{-1}(q)$ and $\hat{\mu}$ and $\hat{\sigma}$ are the MLEs.

3.1 The probit and logit case

Suppose that the true model is probit or logit with mean 0 and variance 1. The efficiencies of mB_2 design, D-, and A-optimal designs for these two models are presented in Table 3. We can see that the mB_2 design is not bad under the assumed model. The efficiencies of the mB_2 design are greater than 98% under the probit model, and are greater than 76% under the logit model.

Table 3: Comparison of efficiencies for probit-logit case

design	p	$weight$	D_1-eff.	A_1-eff.	D_2-eff.	A_2-eff.
mB_2	(0.113, 0.887)	(0.5, 0.5)	99.17%	98.35%	85.15%	76.79%
D_1	(0.128, 0.872)	(0.5, 0.5)	100%	99.73%	91.35%	82.84%
A_1	(0.138, 0.862)	(0.5, 0.5)	99.63%	100%	94.70%	86.61%
D_2	(0.176, 0.824)	(0.5, 0.5)	92.85%	96.65%	100%	96.62%
A_2	(0.214, 0.786)	(0.5, 0.5)	80.53%	87.85%	95.52%	100%

When the true model is logit but the assumed model is probit, mB_2, D_1-, and A_1-optimal designs are compared and the results are presented in Table 3 - Table 5. We can see that the maximum probability bias from the mB_2 design is smaller than that of the others but the mean square errors are not outstanding.

Table 4: Bias and MSE on the mB_2 design under logit model with misspecified probit link function

q	x_q	$\tilde{x}_q$	$F_2(\tilde{x}_q)$	$bias_\infty(q)$	$\overline{\tilde{x}_q}$	$\overline{F_2(\tilde{x}_q)}$	$bias_{1000}$	$\sqrt{MSE_{1000}}$
0.51	0.0221	0.0235	0.5107	-0.0007	0.0258	0.5117	-0.0017	0.0227
0.55	0.1106	0.1179	0.5533	-0.0033	0.1200	0.5541	-0.0041	0.0228
0.6	0.2235	0.2377	0.6061	-0.0061	0.2395	0.6067	-0.0067	0.0229
0.7160_{q_m}	0.5098	0.5358	0.7255	-0.0095	0.5370	0.7255	-0.0095	0.0216
0.8	0.7643	0.7896	0.8072	-0.0072	0.7903	0.8069	-0.0069	0.0179
0.9	1.2114	1.2024	0.8985	0.0015	1.2023	0.8979	0.0021	0.0116
0.9795_{q_M}	2.1318	1.9169	0.9700	0.0095	1.9153	0.9696	0.0099	0.0111
0.99	2.5334	2.1826	0.9813	0.0087	2.1805	0.9809	0.0091	0.0097

Table 5: Bias and MSE on the D_1-optimal design under logit model with misspecified probit link function

q	x_q	$\tilde{x}_q$	$F_2(\tilde{x}_q)$	$bias_\infty(q)$	$\overline{\tilde{x}_q}$	$\overline{F_2(\tilde{x}_q)}$	$bias_{1000}$	$\sqrt{MSE_{1000}}$
0.51	0.0221	0.0234	0.5106	-0.0006	0.0238	0.5108	-0.0008	0.0207
0.55	0.1106	0.1171	0.5529	-0.0029	0.1174	0.5529	-0.0029	0.0209
0.6	0.2235	0.2360	0.6054	-0.0054	0.2361	0.6053	-0.0053	0.0211
0.7085_{q_m}	0.4896	0.5115	0.7166	-0.0081	0.5113	0.7162	-0.0076	0.0205
0.8	0.7643	0.7840	0.8056	-0.0056	0.7835	0.8050	-0.0050	0.0173
0.9	1.2114	1.1938	0.8971	0.0029	1.1928	0.8963	0.0037	0.0125
0.9776_{q_M}	2.0818	1.8690	0.9674	0.0102	1.8672	0.9668	0.0108	0.0121
0.99	2.5334	2.1670	0.9807	0.0093	2.1648	0.9803	0.0097	0.0104

Table 6: Bias and MSE on the A_1-optimal design under logit model with misspecified probit link function

q	x_q	$\tilde{x}_q$	$F_2(\tilde{x}_q)$	$bias_\infty(q)$	$\overline{\tilde{x}_q}$	$\overline{F_2(\tilde{x}_q)}$	$bias_{1000}$	$\sqrt{MSE_{1000}}$
0.51	0.0221	0.0232	0.5105	-0.0005	0.0235	0.5106	-0.0006	0.0206
0.55	0.1106	0.1165	0.5526	-0.0026	0.1170	0.5528	-0.0027	0.0207
0.6	0.2235	0.2349	0.6049	-0.0049	0.2356	0.6051	-0.0051	0.0209
0.7032_{q_m}	0.4756	0.4947	0.7104	-0.0072	0.4959	0.7105	-0.0073	0.0204
0.8	0.7643	0.7804	0.8046	-0.0046	0.7821	0.8046	-0.0046	0.0172
0.9	1.2114	1.1884	0.8962	0.0038	1.1909	0.8960	0.0040	0.0127
0.9763_{q_M}	2.0500	1.8392	0.9656	0.0107	1.8429	0.9654	0.0109	0.0124
0.99	2.5334	2.1572	0.9804	0.0096	2.1616	0.9802	0.0098	0.0105

On the other hand, when the true model is probit but the assumed model is logit, mB_2, D_2-, and A_2-optimal designs are compared and the results are presented in Appendix D. The outcomes are similar that the maximal probability bias of the mB_2 design is the smallest and the mean square errors from that is not outstanding.

3.2 The probit and double reciprocal case

The F_1-F_4 case is considered in this subsection since the min-max difference between these two models is larger than that in the other 5 cases. Suppose that the true model is probit or double reciprocal with location parameter 0 and scale parameter 1. The efficiencies of mB_2 design, D-, and A-optimal designs for these two models are presented in Table 7. The D_4- and A_4-efficiencies of mB_2, D_1-, and A_1-optimal designs are poor, and the D_1- and A_1-efficiencies from D_4- and A_4-optimal designs are not acceptable.

Table 7: Comparison of efficiencies for F_1-F_4 case

design	p	*weight*	D_1-eff.	A_1-eff.	D_4-eff.	A_4-eff.
mB_2	(0.164, 0.836)	(0.5, 0.5)	95.79%	98.40%	7.94%	30.90%
D_1	(0.128, 0.872)	(0.5, 0.5)	100%	99.73%	3.29%	15.48%
A_1	(0.138, 0.862)	(0.5, 0.5)	99.63%	100%	4.34%	19.28%
D_4	(0.207, 0.5, 0.793)	(0.262, 0.478, 0.262)	49.28%	59.90%	100%	78.00%
A_4	(0.207, 0.5, 0.793)	(0.401, 0.198, 0.401)	70.28%	79.64%	71.80%	100%

In many cases, we use probit link function when we do not know what the true model is. In the case when the true model is double reciprocal, mB_2, D_1-, and A_1-optimal designs are compared and the results are presented in Appendix E. We can see that although the maximum probability bias from mB_2 design is the best one, but it is still poor. It would be better to discriminate the model first.

4 Discussions and conclusions

In many binary experiments, we may not know exactly what the true model is. If the assumed model is incorrect, the true q^{th} quantile may be distant from its estimation for some q. In this work, we introduce a criterion evaluating the closeness between two distributions by the CDF differences between them. We use this criterion to define the mB_2 design for two possible models. In other words, the maximum probability bias of q^{th} quantile estimation for all $q \in (0, 1)$ of the mB_2 design reaches the minimum as the number of observations N goes to infinity. A numerical method is also given to find the mB_2 design points in the general cases.

For probit-logit case, it is observed that when the assumed model is the correct one, the efficiencies of the mB_2 design is not bad. The D- and A-efficiencies of mB_2 design are more than 98% when the true model is the probit and are more than 76% when the true model is the logit. If the model is incorrect, the maximal probability bias of the mB_2 design as $N \to \infty$ is smaller than those by other designs, but the mean square errors by the mB_2 design are not outstanding. For F_1-F_4 case, the efficiencies of each of the design discussed here is not good if the design is not correctly specified, but the efficiencies of the mB_3 design, which is discussed in Appendix F, are acceptable for each model, the D_i- and A_i-efficiencies are more than 78% for $i = 1, 4$. If the true model is F_4 but we use F_1, then the maximum probability bias of mB_2 or mB_3 design as $N \to \infty$ is the smallest, but is still greater than 7%. This indicates the seriousness of using a misspecified link model in quantile estimation.

Since we can not discriminate the models discussed here by any two-points designs or symmetric three-points designs, therefore the probability bias of the quantile estimations may be very large, for instance, when the true model is F_4 but we use F_1. For avoiding this kind of mistakes in quantile estimation, a procedure is recommended as follows.

Step 1. Using model discrimination designs first, sequential design approach following that by Muller and Ponce de Leon (1996) can be considered.

Step 2. If there is a model clearly classified to be suitable, then we may perform some further experiments using the optimal design under that model later. Otherwise, it would be better to use the mB_2 or the mB_3 design for the smallest maximal probability bias depending on the results from the first step.

In the future, it would be of interest to find the minimum mean square error design with model uncertainty in mind. Moreover, sometimes we only care whether the extreme quantiles are estimated with high accuracy, such as in the pyrotechnics experiments. We may try to find an optimal design which minimizes the maximum bias of the probabilities with interval restrictions. Finally, for some experiments, the possible models may not be symmetric, the minimum bias designs for these cases can also be considered.

References

[1] Atkinson, A. C., and Fedorov, V. V. (1975). Optimal design: Experiments for discriminating between several models. *Biometrika*, 62, 289-304.

[2] Chao, M. T., and Fuh, C. D. (1999). On Bruceton test for pyrotechnics sensitivity analysis. *17 E&PS 5.1-5.29, Proceedings of the 17th Symposium on Explosives and Pyrotechnic, 1999.*

[3] Dette, H., and Sahm, M. (1997). Standardized optimal designs for binary response experiments. *South African Statistical Journal*, 31, 271-298.

[4] Khan, M. K., and Yazdi, A. A. (1988). On D-optimal designs for binary data. *Journal of Statistical Planning and Inference*, 18, 83-91.

[5] Mathew, T., and Sinha, B. K. (2001). Optimal designs for binary data under logistic regression. *Journal of Statistical Planning and Inference*, 93 (1-2), 295-307.

[6] Minkin, S. (1987). Optimal designs for binary data. *Journal of the American Statistical Association*, 82, 1098-1103.

[7] Muller, W. G., and Ponce de Leon, A. C. M. (1996). Discrimination between two binary data models: Sequentially designed experiments. *Journal of Statistical Computation and Simulation*, 55 , 87-100.

[8] Roussas, G. G. (1997). A course in mathematical statistics. (pp.199)

[9] Sitter, R. R., and Wu, C. F. J. (1993). Optimal designs for binary response experiments: Fieller, D, and A criteria. *Scandinavian Journal of Statistics*, 20, 329-341.

[10] Sitter, R. R., and Fainaru, I. (1997). Optimal designs for the logit and probit models for binary data. *The Canadian Journal of Statistics*, 25, 175-190.

[11] Wu, C. F. J. (1985). Efficient sequential designs with binary data. *Journal of the American Statistical Association*, 80, 974-984.

[12] Wu, C. F. J. (1988). Optimal design for percentile estimation of a quantal response curve. *Optimal Design and Analysis of Experiments*, 213-224.

[13] Yanagisawa, Y. (1988). Designs for discrimination between binary response models. *Journal of Statistical Planning and Inference*, 19 , 31-41.

[14] Yanagisawa, Y. (1990). Designs for discrimination between bivariate binary response models. Biometrical Journal. *Journal of Mathematical Methods in Biosciences*, 32, 25-34.

A The convergence of MLEs for two-points designs with a misspecified link model

Let the true model be $F(\,\cdot\,;\mu_0,\sigma_0)$ and the assumed model be $G(\,\cdot\,;\mu_1,\sigma_1)$ where μ_i and σ_i are unknown parameters, i=0 or 1. Consider a two-points design ξ with sample size N, and supports at x_1, x_2 corresponding weights $\frac{N_1}{N}$, $\frac{N_2}{N}$ respectively.

$$\xi = \left\{ \begin{array}{cc} x_1 & x_2 \\ \frac{N_1}{N} & \frac{N_2}{N} \end{array} \right\}, \text{ where } N_1,\, N_2 \in \mathbf{N}, \text{ and } N_1 + N_2 = N.$$

Let S_i be the number of responses observing 1 at x_i in N_i runs, $i = 1$ or 2. Then

$$S_i \sim B(N_i, F(\frac{x_i - \mu_0}{\sigma_0})),\ i = 1 \text{ or } 2.$$

The log likelihood function of the assumed model is

$$\ln L = S_1 \ln G_1 + (N_1 - S_1)\ln(1 - G_1) + S_2 \ln G_2 + (N_2 - S_2)\ln(1 - G_2),$$

where $G_1 = G(\frac{x_1-\mu_1}{\sigma_1})$ and $G_2 = G(\frac{x_2-\mu_1}{\sigma_1})$. Then by the following equations,

$$\begin{aligned}
\frac{\partial}{\partial \mu_1}\ln L &= \frac{g_1}{\sigma_1}(\frac{1 - S_1/N_1}{1 - G_1} - \frac{S_1/N_1}{G_1}) + \frac{g_2}{\sigma_1}(\frac{1 - S_2/N_2}{1 - G_2} - \frac{S_2/N_2}{G_2}),\\
\frac{\partial}{\partial \sigma_1}\ln L &= \frac{g_1(x_1 - \mu_1)}{\sigma_1^2}(\frac{1 - S_1/N_1}{1 - G_1} - \frac{S_1/N_1}{G_1}) + \frac{g_2(x_2 - \mu_1)}{\sigma_1^2}(\frac{1 - S_2/N_2}{1 - G_2} - \frac{S_2/N_2}{G_2}).
\end{aligned}$$

where $g_i = G_i'$. The MLEs of μ_1 and σ_1, say $\hat{\mu}_1$ and $\hat{\sigma}_1$, should satisfy the following equations,

$$\left\{ \begin{array}{l} \frac{1-S_1/N_1}{1-G_1} - \frac{S_1/N_1}{G_1} = 0 \\ \frac{1-S_2/N_2}{1-G_2} - \frac{S_2/N_2}{G_2} = 0 \end{array} \right.$$

which in turn implies:

$$\left\{ \begin{array}{l} \frac{S_1}{N_1} = G(\frac{x_1-\hat{\mu}_1}{\hat{\sigma}_1}) \\ \frac{S_2}{N_2} = G(\frac{x_1-\hat{\mu}_1}{\hat{\sigma}_1}) \end{array} \right.$$

and

$$\left\{ \begin{array}{l} \hat{\mu}_1 = \frac{x_1 G^{-1}(S_2/N_2) - x_2 G^{-1}(S_1/N_1)}{G^{-1}(S_2/N_2) - G^{-1}(S_1/N_1)} \\ \hat{\sigma}_1 = \frac{x_2 - x_1}{G^{-1}(S_2/N_2) - G^{-1}(S_1/N_1)} \end{array} \right.$$

According to the following Theorem in Roussas (1997), we could show that $\hat{\mu}_1$ and $\hat{\sigma}_1$ would be convergent almost surely.

Theorem A.1. *If for* $j = 1, \ldots, k, X_n^{(j)}, n \geq 1$, *and* $X^{(j)}$ *are r.v.'s, and* $g : \mathbf{R}^k \to \mathbf{R}$ *is continuous, so that* $g(X_n^{(1)}, \ldots, X_n^{(k)})$ *and* $g(X^{(1)}, \ldots, X^{(k)})$ *are r.v.'s, then* $X_n^{(j)} \to X^j$ *almost surely for* $j = 1, \ldots, k$ *imply* $g(X_n^{(1)}, \ldots, X_n^{(k)}) \to g(X^{(1)}, \ldots, X^{(k)})$ *almost surely.*

Let $F(\frac{x_i - \mu_0}{\sigma_0}) = p_i$ for $i = 1$ or 2. By Theorem A.1., since $\frac{S_i}{N_i} \to p_i$ almost surely by the Strong Law of Large Number and $x_i = \mu_0 + \sigma_0 F^{-1}(p_i)$ for $i = 1$ or 2, therefore

$$\begin{aligned}
\hat{\mu}_1 &\to \frac{[\mu_0 + \sigma_0 F^{-1}(p_1)]G^{-1}(p_2) - [\mu_0 + \sigma_0 F^{-1}(p_2)]G^{-1}(p_1)}{G^{-1}(p_2) - G^{-1}(p_1)} \\
&= \mu_0 + \frac{\sigma_0[F^{-1}(p_1)G^{-1}(p_2) - F^{-1}(p_2)G^{-1}(p_1)]}{G^{-1}(p_2) - G^{-1}(p_1)} \quad \text{and} \\
\hat{\sigma}_1 &\to \frac{[\mu_0 + \sigma_0 F^{-1}(p_2)] - [\mu_0 + \sigma_0 F^{-1}(p_1)]}{G^{-1}(p_2) - G^{-1}(p_1)} = \sigma_0 \frac{F^{-1}(p_2) - F^{-1}(p_1)}{G^{-1}(p_2) - G^{-1}(p_1)}
\end{aligned}$$

almost surely as N_1, $N_2 \to \infty$.

Particularly, when x_1 and x_2 are equal to x_{1-p} and x_p respectively for some $p \neq \frac{1}{2}$ and $N_1 = N_2 = \frac{N}{2}$, then

$$\hat{\mu}_1 \to \mu_0 \quad \text{and} \quad \hat{\sigma}_1 \to \sigma_0 \frac{F^{-1}(p)}{G^{-1}(p)}$$

almost surely as $N \to \infty$.

B Properties of the scale function $\beta_{F,G}(p)$ and the distance function $D_{F,G}(b,x)$

In Appendix B, the main goal is to present some properties about $\beta_{F,G}(p)$ and $D_{F,G}(b,x)$ where F and G are standard distributions from certain symmetric location and scale families. According to Figure 1, β function appears to be symmetric and continuously differentiable, and the maximum value of $|D(b,x)|$ for all $x \in \mathbf{R}$ seems to be equal to the maximum value of $|D(b,x)|$ for $x > 0$. We will verify these theoretically in this section. More specifically, we discuss some specific properties for the probit-logit case. In general, these two functions for any two of the possible models from the four families given above have similar properties.

In the following, we assume all of the possible models are from symmetric location and scale families.

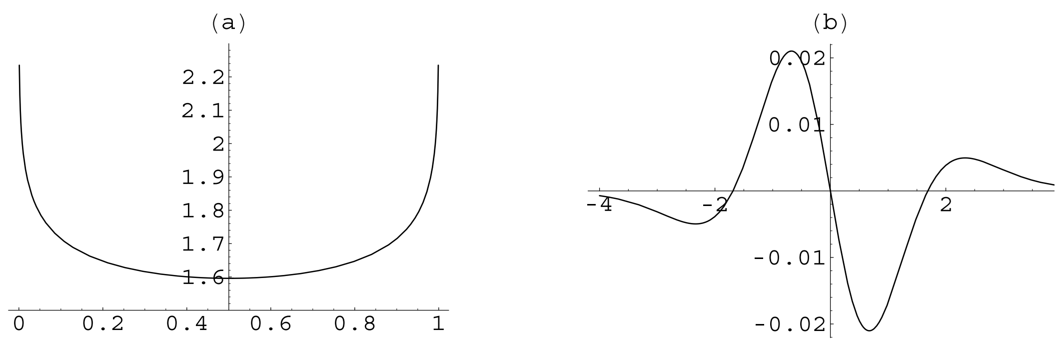

Figure 1: (a):graph of $\beta_{F_1,F_2}(p)$. (b):graph of $D_{F_1,F_2}(1.8,x)$.

Lemma B.1.

Suppose F and G are two standard distributions of possible models. Then for all p in $(0,1)$, $\beta_{F,G}(p)$ is equal to $\beta_{F,G}(1-p)$.

Proof.

It is trivial when $p = \frac{1}{2}$. Since

$$F^{-1}(p) = -F^{-1}(1-p),\ G^{-1}(p) = -G^{-1}(1-p), \text{ and } F^{-1}(p) \neq 0.$$

for all p in $(0,1)$ and $p \neq \frac{1}{2}$. Therefore,

$$\beta_{F,G}(p) = \frac{G^{-1}(p)}{F^{-1}(p)} = \frac{G^{-1}(1-p)}{F^{-1}(1-p)} = \beta_{F,G}(1-p).$$

□

Theorem B.2.

$\beta_{F,G}(p)$ is continuously differentiable on $(0,1)$ except $\frac{1}{2}$. If the densities of F and G are continuous and $\lim_{p\to 1/2} \beta'_{F,G}(p)$ exists, $\lim_{p\to 1/2} \beta'_{F,G}(p) = 0$ and $\beta_{F,G}(p)$ is continuously differentiable on $(0,1)$.

Proof.

First, we show that if $\lim_{p\to 1/2} \beta'_{F,G}(p)$ exists then it is equal to 0. Suppose $\lim_{p\to 1/2} \beta'_{F,G}(p)$ exists and equal to $t \in \mathbf{R}$. By Lemma 2.1, we conclude that $\beta_{F,G}(p) = \beta_{F,G}(1-p)$, so $\beta_{F,G}(\frac{1}{2}+h) = \beta_{F,G}(\frac{1}{2}-h)$ for all $h \in (0, \frac{1}{2})$. It implies

$$\begin{aligned}\beta'_{F,G}(\frac{1}{2}+h) &= \lim_{k\to 0} \frac{\beta_{F,G}(1/2+h+k) - \beta_{F,G}(1/2+h)}{k} \\ &= \lim_{k\to 0} \frac{\beta_{F,G}(1/2-h-k) - \beta_{F,G}(1/2-h)}{k} \\ &= -\lim_{k\to 0} \frac{\beta_{F,G}(1/2-h-k) - \beta_{F,G}(1/2-h)}{-k} = -\beta'_{F,G}(\frac{1}{2}-h).\end{aligned}$$

Therefore,

$$t = \lim_{h\to 0^+} \beta'_{F,G}(\frac{1}{2}+h) = \lim_{p\to 1/2} \beta'_{F,G}(p) = \lim_{h\to 0^-} \beta'_{F,G}(\frac{1}{2}-h) = -t.$$

It means that t must be 0.

Next, we show the continuously differentiation. Since $F'(x)$ and $G'(x)$ are continuous and greater than 0 for all $x \in \mathbf{R}$, by Inverse Function Theorem, $F^{-1}(p)$ and $G^{-1}(p)$ are continuously differentiable on $(0, 1)$. Also, $F^{-1}(p) \neq 0$ for all $p \in (0, 1)$ except $\frac{1}{2}$, therefore $\beta_{F,G}(p) = \frac{G^{-1}(p)}{F^{-1}(p)}$ is continuously differentiable for all $p \in (0, 1)$ except $\frac{1}{2}$.

By the definition of the differentiation,

$$\beta'_{F,G}(\frac{1}{2}) = \lim_{h \to 0} \frac{\beta_{F,G}(\frac{1}{2} + h) - b_0}{h},$$

where $b_0 = \lim_{x \to 1/2}$. Since

$$\lim_{h \to 0} h = 0 \quad \text{and} \quad \lim_{h \to 0} [\ \beta_{F,G}(\frac{1}{2} + h) - b_0\ ] = b_0 - b_0 = 0.$$

Therefore,

$$\beta'_{F,G}(\frac{1}{2}) = \lim_{h \to 0} \frac{\beta'_{F,G}(\frac{1}{2} + h) - 0}{1} = \lim_{p \to \frac{1}{2}} \beta'_{F,G}(p) \quad \text{by L'Hospital's Rule.}$$

It implies that $\beta_{F,G}(p)$ is continuously differentiable at $\frac{1}{2}$. So $\beta_{F,G}(p)$ is continuously differentiable on $(0, 1)$. $\square$

Corollary B.3.

Let $\beta(p) = \beta_{F_1,F_2}(p)$, *where* F_1 *and* F_2 *are standard normal and logistic distributions respectively. Then* $\beta'(p)$ *is continuously differentiable on* $(0, 1)$.

Proof.

Since the densities of F_1 and F_2 are continuous, therefore we only need to show that $\lim_{p \to 1/2} \beta'(p)$ exists. Let $x_p = F^{-1}(p)$,

$$\beta'(p) = \frac{F_2^{-1}(p)'}{F_1^{-1}(p)} - \frac{F_2^{-1}(p) F_1^{-1}(p)'}{(F_1^{-1}(p))^2} = \frac{1}{p(1-p)f_1(x_p)} \eta_0(p),$$

for all p in $(0,1)$ except $\frac{1}{2}$, where

$$\eta_0(p) = \frac{x_p f_1(x_p) - (\ln \frac{p}{1-p})p(1-p)}{x_p^2}.$$

Since $\lim_{p\to\frac{1}{2}} \frac{1}{p(1-p)f_1(x_p)} = 4\sqrt{2\pi} \in \mathbf{R}$, therefore $\lim_{p\to\frac{1}{2}} \beta'(p)$ exists if $\lim_{p\to\frac{1}{2}} \eta_0(p)$ exists.

$$\begin{aligned}
\lim_{p\to\frac{1}{2}} \eta_0(p) &= \lim_{p\to\frac{1}{2}} \frac{x_p f_1(x_p) - (\ln \frac{p}{1-p})p(1-p)}{x_p^2} \\
&= \lim_{x\to 0} \frac{x f_1(x) - (\ln \frac{F_1(x)}{1-F_1(x)})F_1(x)[1-F_1(x)]}{x^2} \quad (\frac{0}{0}) \\
&\quad \text{(by L'Hospital's Rule)} \\
&= \lim_{x\to 0} \frac{-x^2 f_1(x) - (\ln \frac{F_1(x)}{1-F_1(x)})[f_1(x) - 2f_1(x)F_1(x)]}{2x} \quad (\frac{0}{0}) \\
&\quad \text{(by L'Hospital's Rule)} \\
&= \lim_{x\to 0} \frac{1}{2}\{-2xf_1(x) + x^3 f_1(x) - (\ln \frac{F_1(x)}{1-F_1(x)})f_1(x)[x - 2f_1(x)^2 - 2xf_1(x)F_1(x)] \\
&\quad - \frac{f_1(x)}{F_1(x)(1-F_1(x))} f_1(x)[1-2F_1(x)]\} = 0.
\end{aligned}$$

It implies that $\lim_{p\to\frac{1}{2}} \eta_0(p)$ exists and $\beta(p)$ is continuously differentiable on $(0,1)$. □

Corollary B.4.

Let $\beta(p) = \beta_{F_1,F_2}(p)$, where F_1 and F_2 are standard normal and logistic distributions. Then $\beta(p)$ is a strictly increasing function on $p \in [\frac{1}{2},1)$, and $\beta : [\frac{1}{2},1) \to [\sqrt{\frac{8}{\pi}},\infty)$ is a one-to-one and onto function.

Proof.

For all p in $(\frac{1}{2},1)$,

$$\beta'(p) = \frac{F_2^{-1}(p)'}{F_1^{-1}(p)} - \frac{F_2^{-1}(p)F_1^{-1}(p)'}{(F_1^{-1}(p))^2} = \frac{F_2^{-1}(p)'}{(F_1^{-1}(p))^2}\eta_1(p),$$

where

$$\eta_1(p) = F_1^{-1}(p) - \frac{F_2^{-1}(p)F_1^{-1}(p)'}{F_2^{-1}(p)'}.$$

Let $f_1 = F_1'$ $x_p = F_1^{-1}(p)$. According to Inverse Function Theorem, we have

$$\eta_1(p) = x_p - [\ \ln(p) - \ln(1-p)\ ]\frac{p(1-p)}{f_1(x_p)}.$$

Because $\frac{F_2^{-1}(p)'}{(F_1^{-1}(p))^2} > 0$ for all p in $(\frac{1}{2}, 1)$, $\beta(p)$ is strictly increasing on $p \in (\frac{1}{2}, 1)$ if $\eta_1(p) > 0$.

If we can show that the infimum of $\eta_1(p)$ on $p \in (\frac{1}{2}, 1)$ is greater than 0, then this property is proved. It is easy to show that

$$\begin{aligned} \eta_1'(p) &= \frac{1}{f_1(x_p)} - (\frac{1}{p} - \frac{1}{1-p})\frac{p(1-p)}{f_1(x_p)} + \frac{(\ln(p) - \ln(1-p))(1 - 2p - (p - p^2)x_p)}{f_1(x_p)} \\ &= \frac{(\ln(p) - \ln(1-p))(1 - 2p - (p - p^2)x_p)}{f_1(x_p)} \\ &\geq \frac{(\ln(p) - \ln(1-p))}{f_1(x_p)} \times \eta_2(p) \end{aligned}$$

for all $p \in (\frac{1}{2}, 1)$, where

$$\eta_2(p) = 1 - 2p - (1-p)x_p.$$

Because $\frac{\ln(p) - \ln(1-p)}{f_1(x_p)} > 0$ and $\lim_{p \to \frac{1}{2}} \eta_1(p) = 0$, now we should show that $\eta_2(p) > 0$ for all $p \in (\frac{1}{2}, 1)$, so that $\eta_1(p) > 0$.

Let

$$\eta_3(u) = \eta_2(F_1(u)), \quad \forall u > 0.$$

Since there exists a $u > 0$ such that $F_1(u) = p$ for all $p \in (\frac{1}{2}, 1)$, we have $\eta_2(p) > 0$ for all $p \in (\frac{1}{2}, 1)$ if we claim that $\eta_3(u) > 0$ for all $u > 0$. By the definition of $\eta_3(u)$,

$$\begin{aligned} \eta_3'(u) &= 2f_1(u) - 1 + F_1(u) + f_1(u)u, \\ \eta_3''(u) &= f_1(u)(-u^2 - 2u + 2). \end{aligned}$$

Let $u_0 = \sqrt{3} - 1$, then

$$\begin{aligned} \eta_3''(u) &> 0 \text{ and } \eta_3'(u) \text{ is increasing when } u \in (0, u_0), \\ \eta_3''(u) &= 0 \text{ and } \eta_3'(u_0) \text{ is the maximum}, \\ \eta_3''(u) &< 0 \text{ and } \eta_3'(u) \text{ is decreasing when } u \in (u_0, \infty). \end{aligned}$$

Since $\lim_{u \to 0^+} \eta_3'(u) \doteq 0.297885$, and $\lim_{u \to \infty} \eta_3'(u) = 0$, therefore $\eta_3'(u) > 0$ for all $u > 0$. So we have that

$$\eta_3(u) > \eta_3(0) = 0, \text{ for all } u > 0.$$

Now, we have proved that $\beta(p)$ is strictly increasing on $p \in (\frac{1}{2}, 1)$. Because of the continuity of $\beta(p)$ on $p \in [\frac{1}{2}, 1)$, $\beta(p)$ is strictly increasing on $p \in [\frac{1}{2}, 1)$.

Moreover, because $\beta(\frac{1}{2}) = \sqrt{\frac{8}{\pi}}$, $\lim_{p \to \infty} \beta(p) = \infty$ and $\beta(p)$ is continuous and strictly increasing on $[\frac{1}{2}, 1)$, $\beta : [\frac{1}{2}, 1) \to [\sqrt{\frac{8}{\pi}}, \infty)$ is a one-to-one and onto function. $\square$

Theorem B.5.

Suppose F and G are standard distributions of possible models. Then for all $b > 0$, $\sup_{x \in \mathbf{R}} |D_{F,G}(b, x)| = \sup_{x > 0} |D_{F,G}(b, x)|$.

Proof.

Since F and G are symmetric at 0, therefore $F(x) = 1 - F(-x)$ and $G(bx) = 1 - G(-bx)$ for all $b > 0$ and $x \in \mathbf{R}$. It implies that

$$|D_{F,G}(b, x)| = |F(x) - G(bx)| = |F(-x) - G(-bx)| = D_{F,G}(p, -x).$$

So $\sup_{x \in \mathbf{R}} |D_{F,G}(p, x)| = \sup_{x \geq 0} |D_{F,G}(p, x)|$.

Since $F(0)$ =G(0)= $\frac{1}{2}$, therefore $|D_{F,G}(b, 0)| = 0$ for all $b > 0$, and $|D(b, x)| \geq 0$ for all

$b > 0$ and $x > 0$, So

$$\sup_{x \in \mathbf{R}} |D_{F,G}(b,x)| = \sup_{x \geq 0} |D_{F,G}(b,x)| = \sup_{x > 0} |D_{F,G}(b,x)|$$

$\square$

C Figures of difference between two models

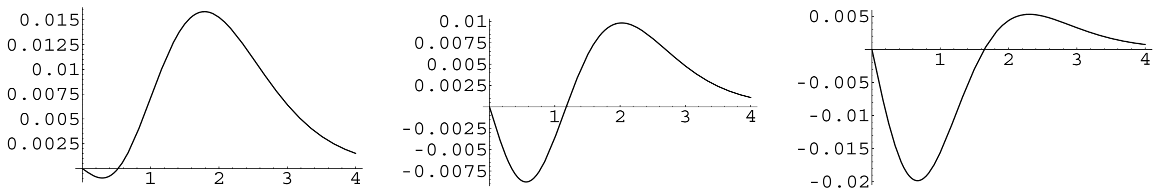

Figure 2: $F_1(x) - F_2(\beta_{F_1,F_2}(p)x)$ for $p = 0.7$, $p = 0.88$, and $p = 0.95$

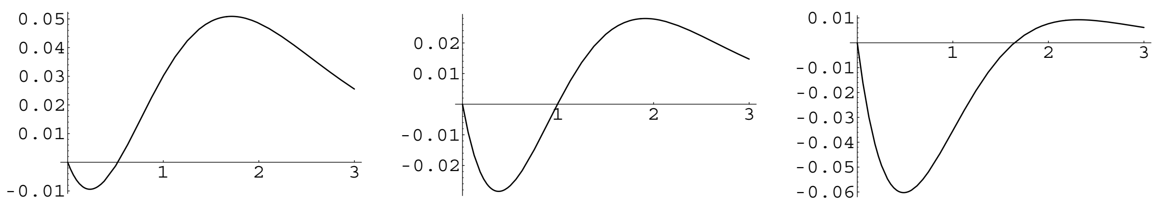

Figure 3: $F_1(x) - F_3(\beta_{F_1,F_3}(p)x)$ for $p = 0.7$, $p = 0.84$, and $p = 0.95$

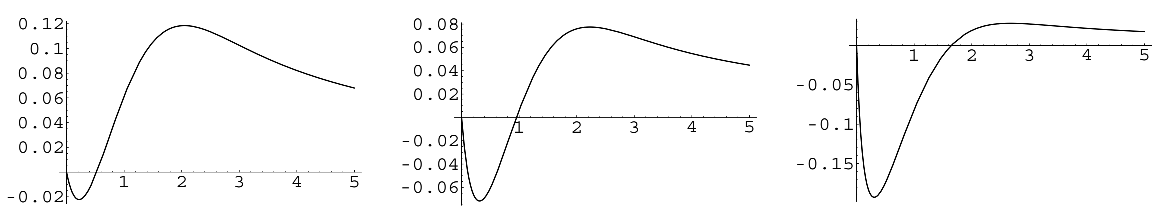

Figure 4: $F_1(x) - F_4(\beta_{F_1,F_4}(p)x)$ for $p = 0.7$, $p = 0.83$, and $p = 0.95$

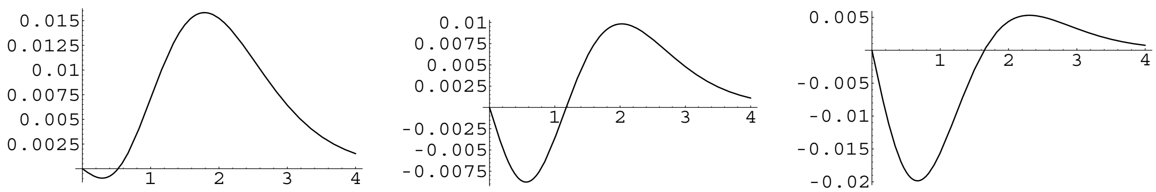

Figure 5: $F_2(x) - F_3(\beta_{F_2,F_3}(p)x)$ for $p = 0.7$, $p = 0.81$, and $p = 0.95$

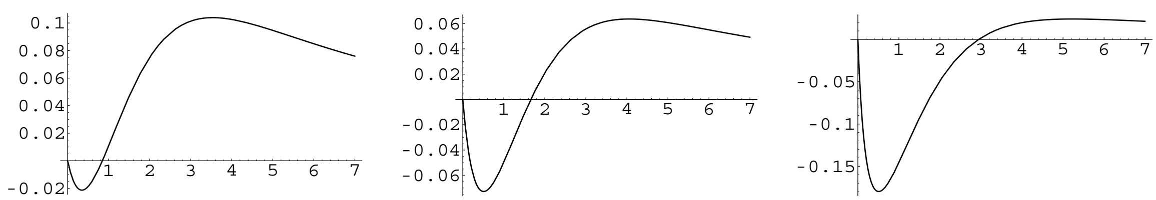

Figure 6: $F_2(x) - F_4(\beta_{F_2,F_4}(p)x)$ for $p = 0.7$, $p = 0.86$, and $p = 0.95$

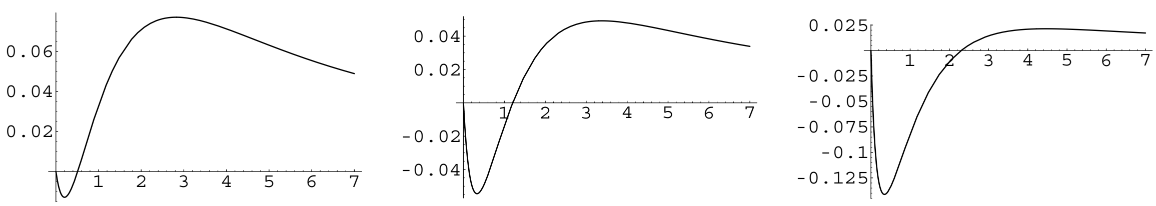

Figure 7: $F_3(x) - F_4(\beta_{F_3,F_4}(p)x)$ for $p = 0.75$, $p = 0.88$, and $p = 0.95$

D Tables for probit being the true model with logit link function

Table 8: Bias and MSE on the mB_2 design for F_1 and F_2

q	x_q	$\tilde{x}_q$	$F_1(\tilde{x}_q)$	$bias_\infty(q)$	$\overline{\tilde{x}_q}$	$\overline{F_1(\tilde{x}_q)}$	$bias_{1000}$	$\sqrt{MSE_{1000}}$
0.51	0.0251	0.0235	0.5094	0.0006	0.0246	0.5098	0.0002	0.0230
0.55	0.1257	0.1179	0.5469	0.0031	0.1187	0.5472	0.0028	0.0230
0.6	0.2533	0.2383	0.5942	0.0058	0.2387	0.5942	0.0058	0.0234
0.7255_{q_M}	0.5993	0.5711	0.7160	0.0095	0.5705	0.7154	0.0100	0.0238
0.8	0.8416	0.8146	0.7924	0.0076	0.8134	0.7914	0.0086	0.0217
0.9	1.2816	1.2912	0.9017	-0.0017	1.2885	0.9004	-0.0004	0.0150
0.9700_{q_m}	1.8808	2.0431	0.9795	-0.0095	2.0382	0.9785	-0.0085	0.0105
0.99	2.3264	2.7002	0.9965	-0.0065	2.6934	0.9962	-0.0062	0.0064

Table 9: Bias and MSE on the D_2-optimal design

q	x_q	$\tilde{x}_q$	$F_1(\tilde{x}_q)$	$bias_\infty(q)$	$\overline{\tilde{x}_q}$	$\overline{F_1(\tilde{x}_q)}$	$bias_{1000}$	$\sqrt{MSE_{1000}}$
0.51	0.0251	0.0241	0.5096	0.0004	0.0239	0.5095	0.0005	0.0198
0.55	0.1257	0.1210	0.5482	0.0018	0.1208	0.5480	0.0020	0.0200
0.6	0.2533	0.2445	0.5966	0.0034	0.2443	0.5964	0.0036	0.0203
0.6876_{q_M}	0.4891	0.4757	0.6829	0.0047	0.4755	0.6825	0.0051	0.0207
0.8	0.8416	0.8358	0.7984	0.0016	0.8356	0.7978	0.0022	0.0192
0.9	1.2816	1.3248	0.9074	-0.0074	1.3245	0.9065	-0.0065	0.0159
0.9598_{q_m}	1.7484	1.9136	0.9722	-0.0124	1.9133	0.9713	-0.0115	0.0137
0.99	2.3264	2.7705	0.9972	-0.0072	2.7702	0.9969	-0.0069	0.0071

Table 10: Bias and MSE on the A_2-optimal design using logit model

q	x_q	$\tilde{x}_q$	$F_1(\tilde{x}_q)$	$bias_\infty(q)$	$\overline{\tilde{x}_q}$	$\overline{F_1(\tilde{x}_q)}$	$bias_{1000}$	$\sqrt{MSE_{1000}}$
0.51	0.0251	0.0244	0.5097	0.0003	0.0251	0.5100	-1.6×10^{-5}	0.0186
0.55	0.1257	0.1223	0.5487	0.0013	0.1229	0.5489	0.0011	0.0187
0.6	0.2533	0.2470	0.5976	0.0024	0.2476	0.5977	0.0023	0.0190
0.6653_{q_M}	0.4270	0.4186	0.6622	0.0031	0.4190	0.6622	0.0031	0.0195
0.8	0.8416	0.8446	0.8008	-0.0008	0.8447	0.8003	-0.0003	0.0190
0.9	1.2816	1.3387	0.9097	-0.0097	1.3383	0.9087	-0.0087	0.0172
0.9551_{q_m}	1.6965	1.8624	0.9687	-0.0137	1.8616	0.9677	-0.0126	0.0152
0.99	2.3264	2.7997	0.9974	-0.0074	2.7980	0.9971	-0.0071	0.0072

E Tables for double-reciprocal being the true model with probit link

Table 11: Bias and MSE on the mB_2 design for F_1 and F_4

q	x_q	$\tilde{x}_q$	$F_1(\tilde{x}_q)$	$bias_\infty(q)$	$\overline{\tilde{x}_q}$	$\overline{F_1(\tilde{x}_q)}$	$bias_{1000}$	$\sqrt{MSE_{1000}}$
0.51	0.0204	0.0526	0.5250	-0.0150	0.0536	0.5231	-0.0131	0.0448
0.55	0.1111	0.2635	0.6043	-0.0543	0.2648	0.6021	-0.0521	0.0614
0.6238_{q_m}	0.3291	0.6614	0.6991	-0.0752	0.6635	0.6982	-0.0744	0.0769
0.7	0.6667	1.0995	0.7618	-0.0618	1.1022	0.7615	-0.0615	0.0628
0.8	1.5000	1.7646	0.8191	-0.0191	1.7685	0.8190	-0.0190	0.0208
0.9	4.0000	2.6869	0.8644	0.0356	2.6924	0.8643	0.0357	0.0361
0.95	9.0000	3.4486	0.8876	0.0624	3.4554	0.8876	0.0624	0.0626
0.9876_{q_M}	40.152	4.7217	0.9124	0.0752	5.7147	0.9124	0.0752	0.0754

Table 12: Bias and MSE on the D_1-optimal design

q	x_q	$\tilde{x}_q$	$F_1(\tilde{x}_q)$	$bias_\infty(q)$	$\overline{\tilde{x}_q}$	$\overline{F_1(\tilde{x}_q)}$	$bias_{1000}$	$\sqrt{MSE_{1000}}$
0.51	0.0204	0.0643	0.5302	-0.0202	0.0603	0.5249	-0.0149	0.0563
0.55	0.1111	0.3224	0.6219	-0.0719	0.3185	0.6167	-0.0667	0.0781
0.6286_{q_m}	0.3463	0.8420	0.7286	-0.1000	0.8383	0.7264	-0.0978	0.1001
0.7	0.6667	1.3456	0.7868	-0.0868	1.3421	0.7856	-0.08562	0.0868
0.8	1.500	2.1596	0.8418	-0.0418	2.1564	0.8411	-0.0411	0.0420
0.9	4.0000	3.2885	0.8834	0.0166	3.2857	0.8831	0.0169	0.0179
0.95	9.0000	4.2207	0.9042	0.0458	4.2183	0.9040	0.0460	0.0462
0.9902_{q_M}	49.761	5.9839	0.9284	0.0617	5.9822	0.9282	0.0619	0.0620

Table 13: Bias and MSE on the A_1-optimal design

q	x_q	$\tilde{x}_q$	$F_1(\tilde{x}_q)$	$bias_\infty(q)$	$\overline{\tilde{x}_q}$	$\overline{F_1(\tilde{x}_q)}$	$bias_{1000}$	$\sqrt{MSE_{1000}}$
0.51	0.0204	0.0604	0.5285	-0.0185	0.0694	0.5291	-0.0191	0.0529
0.55	0.1111	0.3030	0.6163	-0.0663	0.3112	0.6154	-0.0654	0.0747
0.6276_{q_m}	0.3426	0.7848	0.7199	-0.0923	0.7915	0.7195	-0.0919	0.0940
0.7	0.6667	1.2644	0.7792	-0.0792	1.2695	0.7789	-0.0789	0.0800
0.8	1.5000	2.0292	0.8349	-0.0349	2.0319	0.8347	-0.0347	0.0356
0.9	4.0000	3.0900	0.8777	0.0223	3.0892	0.8775	0.0225	0.0232
0.95	9.0000	3.9658	0.8993	0.0507	3.9624	0.8991	0.0509	0.0511
0.9894_{q_M}	46.170	5.5560	0.9237	0.0657	5.5475	0.9235	0.0659	0.0660

F Some further works about mB_3 design for the probit and double reciprocal case

Since D_4- and A_4-optimal designs have 3 points and they seems to be more efficient than others. We guess that the mB_3 design,which is adding $x_{\frac{1}{2}}$ into the mB_2 design, would more efficient. Let mB_3 design is

$$\xi_3^* = \left\{ \begin{array}{ccc} x_{1-p} & x_{\frac{1}{2}} & x_p \\ q^* & 1-2q^* & q^* \end{array} \right\},$$

$$\text{where } q^* = \max_{q \leq 0.5} \min\{D_i\text{-efficiency}, A_i\text{-efficiency}, i = 1, 4\}.$$

By some numerical computation, we can find $q^* = 0.3663$ and the efficiencies of mB_3 design for F_1-F_4 case. By comparison of Table 7 and Table 14, mB_3 design is better than others.

Table 14: Efficiencies of mB_3 deisgn

design	D_1-eff.	A_1-eff.	D_4-eff.	A_4-eff.
mB_3	78.22%	86.62%	95.97%	78.22%