Optimum designs for model discrimination and estimation in Binary Response Models

by

Wei-Shan Hsieh

Advisor

Mong-Na Lo Huang

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

July, 2005

Abstract

This paper is concerned with the problem of finding an experimental design for discrimination between two rival models and for model robustness that minimizing the maximum bias simultaneously in binary response experiments. The criterion for model discrimination is based on the T-optimality criterion proposed in Atkinson and Fedorov (1975), which maximizes the sum of squares of deviations between the two rival models while the criterion for model robustness is based on minimizing the maximum probability bias of the two rival models. In this paper we obtain the optimum designs satisfy the above two criteria for some commonly used rival models in binary response experiments such as the probit and logit models etc.

Keywords : Least square estimate (LSE), mean square error, model discrimination, model robustness, symmetric location and scale family.

Contents

1 Introduction

Optimum designs for precise estimation of model parameters have been discussed for quite a long time. However, optimum designs for discrimination between models have had less attention and have been developed only for some cases. There are some researches discussing the design problem especially on how to discriminate between models. See Chambers and Cox (1967), Atkinson and Fedorov (1975), Yanagisawa (1988, 1990) and Müller and Ponce de Leon (1996) etc. Atkinson and Fedorov (1975) proposed the T-optimum design criterion for model discrimination between two rival models which maximizes the ordinary sum of squared deviations at support points for two rival response models. Following their idea, we look for the T-optimum designs for discriminating two binary response models.

First, our main goal is to find the T-optimum designs with binary response models. On the other hand, we would also like to consider the optimum design for prediction with for model robustness in mind for binary response models as introduced by Huang and Hwang (2004). It would be interesting to see how the T-optimum design in binary response models performs in the sense of model robustness in estimation that minimizes the maximum deviation between the true and assumed models.

A binary response experiment is that the response variable y takes only one of two possible values, say 1 or 0. The relation between the response variable y and the independent variable x is controlled by another random variable Z in the relation that random variable Z is less than the independent variable x if and only if the response variable y is 1. We can only observe the result that whether the event $\{Z < x\}$ happens or not. The mathematical expression is that

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

We do not know exactly what the distribution of Z is, but Z is conventionally assumed to be 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.

There are several possible models discussed in many literatures, for examples, the probit, logit, double exponential and the double reciprocal families. Moreover, the probit and logit models are used most often. There are a lot of researches discussing the optimum designs for binary response experiments with a given model under the assumption that Z is from a symmetric location and scale family. The four families we have just mentioned above are presented here,

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

where μ and σ are unknown parameters.

1.1 Literatures Review

As mentioned before, there have been some researches discussing the design problem about discrimination between binary response models. For different cases and purposes, the results are different. Results in Chambers and Cox (1967), Atkinson and Fedorov (1975), Yanagisawa (1988) and Müller and Ponce de Leon (1996) are briefly described below.

First, about optimum designs for model discrimination in general models Atkinson and Fedorov (1975) proposed the T-optimality criterion related to maximizing the noncentrality parameter of the χ^2 distribution of residual sum of squares, which is equivalent to maximizing the power of the F test for departures from the assumed model. They summarized the properties of T-optimum designs in a theorem analogous to the celebrated general equivalence theorem of Kiefer and Wolfowitz for D-optimum designs.

For logit and probit binary models, a first attempt to tackle the problem of model discrimination was made by Chambers and Cox (1967). They considered the experiments with only three dose levels and found the power of a significance test for the null hypothesis that the response curve is logistic against the alternative that it is normal, and vice versa. From this they gave a suitable spacing of dose levels for discrimination. They stated that approximately 1000 observations are necessary for even modest sensitivity.

For binary data models, the Pearson chi-squared statistic and the log likelihood ratio statistic are two measures which can be used for testing model adequacy. It is well known (cf. Agresti(1990)) that both of them are asymptotically distributed as $\chi^2_{k-m_0}$ under the null hypothesis H_0, and as $\chi^2_{k-m_0,\lambda}$ under the alternative hypothesis H_A where k is the number of observations and m_0 is the number of parameters in the model under the null hypothesis.

The noncentrality parameter λ is given by, respectively

$$\lambda = \sum_{i=1}^{k} n_i \frac{(F_{Ai} - \hat{F_{0i}})^2}{\hat{F_{0i}}(1 - \hat{F_{0i}})} \tag{1.1}$$

for Pearson chi-squared statistic and

$$\lambda = 2\sum_{i=1}^{k} n_i(F_{Ai}log(\frac{F_{Ai}}{\hat{F_{0i}}}) + (1 - F_{Ai})log(\frac{1 - F_{Ai}}{1 - \hat{F_{0i}}})) \qquad (1.2)$$

for log likelihood ratio statistic. Here, subscript 0 refers to the true model and A to the rival one, the hat terms denote the eatimates, k refers to the number of experimental units at each of which n_i observations are taken, and m_0 to the munber of parameters in the H_0 model.

Yanagisawa (1988) extended the work of Chambers and Cox (1967) and proposed a test statistic, the weighted sum of squares, for discrimination between alternative binary response models which is asymptotically equivalent to the log likelihood ratio statistic and Pearson's goodness of fit statistic. The result of Yanagisawa (1988), say T_{PS}-optimum design, is based on maximizing the noncentrality parameter of the Pearson's chi-squared statistic (1.1). They also presented procedure for finding the optimal designs. Under certain conditions they proved that the maximum value of the power can be obtained when the degrees of freedom of the test statistic is one ,i.e. the number of support points is $m_0 + 1$. Several mathematical properties of the incomplete gamma function ratio and the non-central chi-squared distribution are required in the discussion and have been established by them.

Based on maximizing the noncentrality parameter of the log likelihood ratio statistic (1.2) Müller and Ponce de Leon (1996) tried to find the corresponding optimum design, say T_{LR}-optimum design. They proposed a sequential procedure to design optimum experiments for discriminating between two binary data models. To be able to specify the problem explicitly, not only the model link functions need to be provided but also their associated linear predictor structures. Further more, they supposed that one of the models is true although it is not known which of them. Under those assumptions, the procedure consists of making sequential choices of single experimental units to discriminate between the rival models as efficiently as possible. A simulation study for the classical case of probit versus logit model was presented.

1.2 Maxmin and minimax criteria

The purpose of this work is to design an experiment for discrimination between two rival models and for estimation of the distribution function with model robustness property at the same time. The maxmin criterion for model discrimination is based on the T-optimality criterion proposed by Atkinson and Fedorov (1975). To distinguish from the T_{PS}- and T_{LR}-optimum design criteria, we call it a T_{LS}-optimum design criterion. To be more explicitly, suppose the possible models of Z are from two symmetric location and scale families with standard distributions $F(\frac{x-\mu_0}{\sigma_0})$ and $G(\frac{x-\mu}{\sigma})$ where μ_0, σ_0, μ and σ are unknown parameters, and we do not know which one is the true model. In the following,we firstly introduce the maxmin criterion.

A design is a set of distinct support points at $x_1, \ldots, x_n$ with corresponding weights $w_1, \ldots, w_n$, $w_i > 0$, $\sum w_i = 1$, denoted by

$$\xi = \left\{ \begin{array}{ccc} x_1 & \ldots & x_n \\ w_1 & \ldots & w_n \end{array} \right\}.$$

Without loss of generality, we assume F is the true model. A least square minimizer $(\hat{\mu}, \hat{\sigma})$, called LSM of the parameters in the second model, G, is a solution of the equation

$$\int_{\chi} \|F(\frac{x-\mu_0}{\sigma_0}) - G(\frac{x-\hat{\mu}}{\hat{\sigma}})\|^2 \xi(dx) = \min_{\mu,\sigma} \int_{\chi} \|F(\frac{x-\mu_0}{\sigma_0}) - G(\frac{x-\mu}{\sigma})\|^2 \xi(dx)$$

where ξ stands for the given design. A design is more favorable if it can yield as large a value as possible of the sum of squares for lack of fit of the second model, which is equivalent to maximizing

$$\begin{aligned} T_{LS}(\xi) &= \int_{\chi} \|F(\frac{x-\mu_0}{\sigma_0}) - G(\frac{x-\hat{\mu}}{\hat{\sigma}})\|^2 \xi(dx) \\ &= \min_{\mu,\sigma} \int_{\chi} \|F(\frac{x-\mu_0}{\sigma_0}) - G(\frac{x-\mu}{\sigma})\|^2 \xi(dx). \end{aligned}$$

Thus, the design ξ^* which satisfies

$$T_{LS}(\xi^*) = \max_{\xi} \min_{\mu,\sigma} \int_{\chi} \|F(\frac{x-\mu_0}{\sigma_0}) - G(\frac{x-\mu}{\sigma})\|^2 \xi(dx)$$

is defined to be a T_{LS}-optimum design.

Note that the T_{LS}-criterion can be thought of as a maximin type of criterion. Huang and Hwang (2004) considered the model robustness criterion to minimize the maximum distance between the distribution functions under the true and assumed models which can be thought of as a minimax type of criterion that is if ξ^* is a minimax design then it satisfies.

$$mB(\xi^*) = \min_{\xi} \max_{x \in \mathbb{R}} |F(\frac{x-\mu_0}{\sigma_0}) - G(\frac{x-\hat{\mu}}{\hat{\sigma}})|,$$

where $(\hat{\mu}, \hat{\sigma})$ are minimizers of (μ, σ) of the absolute deviation above.

Now we look for the T_{LS}-optimum design and then verify that it is also a minimax bias design. In the process of finding the desired T_{LS}-optimum design, we need to use the equivalence theorem of Atkinson and Fedorov (1975) to obtain our result. Now, we introduce the equivalence theorem of Atkinson and Fedorov (1975) in Theorem 1.1 as follows.

Theorem 1.1 Assume that

(a) the design region χ is compact and $F(x;\theta)$ and $G(x;\theta)$ are continuous for x$\in \chi$.

(b) $F(x;\theta)$ and $G(x;\theta)$ are differentiable functions of θ.

(c) the least square minimizers, $(\hat{\mu}, \hat{\sigma})$, of (μ, σ) is unique.

Given the preceding assumptions:

(i) a necessary and sufficient condition for a design ξ^* to be T_{LS}-optimum is fulfillment of the inequality

$$\psi_G(x, \xi^*) \leq \max_{\xi} \int_{\chi} \| F(\frac{x-\mu_0}{\sigma_0}) - G(\frac{x-\hat{\mu}}{\hat{\sigma}}) \|^2 \xi(dx) = T_{LS}(\xi^*),$$

where $$\psi_G(x, \xi^*) = \{F(\tfrac{x-\mu_0}{\sigma_0}) - G(\tfrac{x-\hat{\mu}}{\hat{\sigma}})\}^2;$$

(ii) at the support points of the optimum design $\psi_G(x, \xi^*)$ achieves its upper bound.

2 Desired T_{LS}-optimum design

In this section, we find the T_{LS}-optimum design in binary response models. In the following, we would like to present some necessary properties of the T_{LS}-optimum design and use these properties to find the desired T_{LS}-optimum design.

2.1 Properties of T_{LS}-optimum designs

First note that for symmetric binary response models there exists a T_{LS}-optimum design which is also symmetric. This property can be established by showing the concavity and the symmetry properties of the mapping $\xi \to T_{LS}(\xi)$, when the true and assumed models are from symmetric location and scale family. For all symmetric designs, the only LSM of the mean parameter, μ, in the assumed models is the mean parameter, μ_0, in the true model. Any of the two- and three-point symmetric designs is not a T_{LS}-optimum for binary response models. Last, for all symmetric designs with odd numbers, $2n+1$, of support points are not T_{LS}-optimum designs. We summarize these properties of the mapping, $\xi \rightarrow T_{LS}(\xi)$, and the T_{LS}-optimum designs in two Lemmas below and the proofs are delayed to appendix.

Lemma 2.1. *For an arbitrary design ξ_1, if ξ_2 is the reflected design of ξ with respect to μ_0, let $\tilde{\xi} = \alpha\xi_1 + (1-\alpha)\xi_2$ where $0 < \alpha < 1$, then*

(i)$T_{LS}(\tilde{\xi}) \geq \alpha T_{LS}(\xi_1) + (1-\alpha)T_{LS}(\xi_2)$,

(ii)$T_{LS}(\xi_1) = T_{LS}(\xi_2)$.

Lemma 2.2. *Three properties of T_{LS}-optimum designs are as follows*

(i) For all symmetric designs, the only LSM of μ is μ_0.

(ii) Any of two- and three-point symmetric designs is not a T_{LS}-optimum design.

(iii) Any symmetric design with odd numbers of support points is not T_{LS}-optimum.

2.2 T_{LS}-optimum design for binary response models

In previous subsections, we have shown that two- and three-point symmetric designs are not T_{LS}-optimum designs. Before we discuss the symmetric four-point designs, we have to define some notation first.

Let σ^* satisfy

$$\max_{x>0}\{F(\frac{x-\mu_0}{\sigma_0}) - G(\frac{x-\mu_0}{\sigma^*})\} = -\min_{x>0}\{F(\frac{x-\mu_0}{\sigma_0}) - G(\frac{x-\mu_0}{\sigma^*})\}, \tag{2.1}$$

$$a = arg\min_{x>0}\{F(\frac{x-\mu_0}{\sigma_0}) - G(\frac{x-\mu_0}{\sigma^*})\} - \mu_0, \tag{2.2}$$

$$b = arg\max_{x>0}\{F(\frac{x-\mu_0}{\sigma_0}) - G(\frac{x-\mu_0}{\sigma^*})\} - \mu_0 \tag{2.3}$$

and

$$w = \frac{b \cdot g(\frac{b}{\sigma^*})}{2[b \cdot g(\frac{b}{\sigma^*}) + a \cdot g(\frac{a}{\sigma^*})]} \quad \text{where } \mathrm{g}(\cdot) = G'(\cdot). \tag{2.4}$$

Now consider the symmetric four-point design ξ^* with supports $\mu_0 - b$, $\mu_0 - a$, $\mu_0 + a$, $\mu_0 + b$ and corresponding weights $\frac{1}{2} - w, w, w, \frac{1}{2} - w$ respectively, where σ^* a, b and w are defined in (2.1) - (2.4), then the design ξ^* is defined as

$$\xi^* = \left\{ \begin{array}{cccc} \mu_0 - b & \mu_0 - a & \mu_0 + a & \mu_0 + b \\ \frac{1}{2} - w & w & w & \frac{1}{2} - w \end{array} \right\}.$$

If the true model is with distribution $F(\frac{x-\mu_0}{\sigma_0})$ but the assumed model is with distribution $G(\frac{x-\mu}{\sigma})$ where μ and σ are unknown, then $(\hat{\mu}_N, \hat{\sigma}_N)$, the estimated LSM of (μ, σ), is proved to converge to (μ_0, σ^*) as the number of observations $N \to \infty$, that is,

$$\lim_{N\to\infty} \hat{\mu}_N = \mu_0, \lim_{N\to\infty} \hat{\sigma}_N = \sigma^*,$$

the details of the proof is delayed to Appendix B.

Under the assumption that σ^* is the only LSM of σ for ξ^*, we can use the result of Atkinson and Fedorov (1975) to claim that the design ξ^* defined above is a T_{LS}-optimal design. Since (μ_0, σ^*) is also the resulting minimizers of the minimum bias design that minimizes the maximum distance between the two distributions, we can say that the T_{LS}-optimum design ξ^* is also a model robust design that minimizes the maximum bias. In the following we present results for a special case when the assumed model is with distribution of the probit family.

In the following, we prove that the design ξ^* defined above is a T_{LS}-optimal design for the case when the assumed model is with distribution of probit family, by first proving that σ^* is the only LSE of σ for ξ^* and then use the result of Atkinson and Fedorov (1975) to verify it. In the following, we illustrate that σ^* is the only LSE of σ for ξ^*. Let us recall the representation of least square minimizers of the parameters in the second model, G, are the solutions of the equation

$$\int_{\chi} \|F(\frac{x-\mu_0}{\sigma_0}) - G(\frac{x-\hat{\mu}}{\hat{\sigma}})\|^2 \xi(dx) = \min_{\mu,\sigma} \int_{\chi} \|F(\frac{x-\mu_0}{\sigma_0}) - G(\frac{x-\mu}{\sigma})\|^2 \xi(dx)$$

and define

$$\begin{aligned} h(\mu,\sigma,\xi^*) &= w\{[F(\frac{a}{\sigma_0}) - G(\frac{a-(\mu_0-\mu)}{\sigma})]^2 + [F(\frac{a}{\sigma_0}) - G(\frac{a+(\mu_0-\mu)}{\sigma})]^2\} \\ &\quad +(\frac{1}{2}-w)\{[F(\frac{b}{\sigma_0}) - G(\frac{b-(\mu_0-\mu)}{\sigma})]^2 + [F(\frac{b}{\sigma_0}) - G(\frac{b+(\mu_0-\mu)}{\sigma})]^2\}. \end{aligned}$$

Since we have shown that μ_0 is the only LSM of μ, replacing $\hat{\mu}$ with μ_0, then

$$h(\mu_0,\sigma,\xi^*) = 2w[F(\frac{a}{\sigma_0}) - G(\frac{a}{\sigma})]^2 + 2(\frac{1}{2}-w)[F(\frac{a}{\sigma_0}) - G(\frac{a}{\sigma})]^2 \tag{2.5}$$

and

$$\frac{\partial}{\partial\sigma}h(\mu_0,\sigma,\xi^*) = 4w[F(\frac{a}{\sigma_0}) - G(\frac{a}{\sigma})]g(\frac{a}{\sigma})\frac{a}{\sigma^2} + 4(\frac{1}{2}-w)[F(\frac{b}{\sigma_0}) - G(\frac{b}{\sigma})]g(\frac{b}{\sigma})\frac{b}{\sigma^2}.$$

In order to determine the uniqueness of the LSE of σ , a sufficient condition is to check the sign of the partial derivative for σ of the h function. Let $D(\sigma) = g(\frac{a}{\sigma^*})g(\frac{b}{\sigma}) - g(\frac{a}{\sigma})g(\frac{b}{\sigma^*})$ where $g(\cdot)$ is the probability density function of the assumed model. We verify that the sign of $\frac{\partial}{\partial\sigma}h(\mu_0,\sigma,\xi^*)$ is the same as $D(\sigma)$ in Lemma 2.3 below and the proof is delayed to the appendix.

Lemma 2.3. *Let the h function be defined as (2.5), then the sign of $\frac{\partial}{\partial\sigma}h(\mu_0,\sigma,\xi^*)$ is the same as $D(\sigma)$.*

In what follows, we consider a case that the assumed model is probit, that is, $g(\cdot)$ in (2.5) is with the normal probability density function. We verify that σ^* is the only

LSE of σ in Lemma 2.4 and find a candidate of the T_{LS}-optimum design then prove it in Theorem 2.5 to be T_{LS}-optimal when the assumed model is probit.

Lemma 2.4. *Let σ^* be defined as in (2.1)-(2.4) and the assumed model $g(\cdot)$ is with the normal probability density function, then σ^* is the only LSM of σ.*

Proof. Consider

$$\begin{aligned} D(\sigma) &= g(\frac{a}{\sigma^*})g(\frac{b}{\sigma}) - g(\frac{a}{\sigma})g(\frac{b}{\sigma^*}) \\ &= \frac{1}{2\pi}exp\{-\frac{a^2}{2\sigma^{*2}} - \frac{b^2}{2\sigma^2}\} - \frac{1}{2\pi}exp\{-\frac{a^2}{2\sigma^2} - \frac{b^2}{2\sigma^{*2}}\} \\ &= \frac{1}{2\pi}exp\{-\frac{a^2}{2\sigma^{*2}} - \frac{a^2}{2\sigma^2}\}[exp\{-\frac{b^2-a^2}{2\sigma^2}\} - exp\{-\frac{b^2-a^2}{2\sigma^{*2}}\}] \end{aligned}$$

since the exponential function is positive for all real value and it is also a strictly increasing function hence $D(\sigma)$ equals to zero only as $\sigma = \sigma^*$. That is to say, σ^* is the only LSE of σ. □

Since we have shown that (μ_0, σ^*) is the only pair of the LSE of (μ, σ), we may say that the design ξ^* defined above is exactly a T_{LS}-optimum design by using the equivalence theorem of Atkinson and Fedorov (1975) to verify it.

Theorem 2.5. *Let a, b, σ^* and w be defined as above, then when the assumed model is probit ξ^* is a desired T_{LS}-optimum design, where*

$$\xi^* = \left\{ \begin{array}{cccc} \mu_0 - b & \mu_0 - a & \mu_0 + a & \mu_0 + b \\ \frac{1}{2} - w & w & w & \frac{1}{2} - w \end{array} \right\}.$$

As the LSE of (μ, σ), (μ_0, σ^*), coincide with the result of Huang and Hwang (2004),

$$\xi^* = arg \min_{\xi} \max_{x \in \mathbb{R}} |F(\frac{x-\mu}{\sigma}) - G(\frac{x-\hat{\mu}}{\hat{\sigma}})|,$$

is also minimax when the assumed model G is probit. It is interesting to find that ξ^* is not only a T_{LS}-optimum design but also minimax.

2.3 General cases

In the previous section, it is shown that σ^* is the LSE of σ for the design ξ^* defined before. However, in order to use the main result of Atkinson and Fedorov (1975) to say that the design ξ^* defined above is the desired T_{LS}-optimum design, it needs to be showed that σ^* is the unique LSE of σ for the design ξ^*. Here we can not say that they possess this property for all binary response models. However, we may discuss those models considered most often in many literatures.

The cases that the assumed models are with double-exponential or double-reciprocal models can be proved that σ^* is the unique LSM of σ for the design ξ^* by using similar argument as above (see Appendix C). However, when the assumed model is with distributions of logit family, it is difficult to show that σ^* is the only LSM of σ. Although we do not prove that the LSM of σ is unique analytically, but we see from numerical computation result that the LSM is unique numerically. For general cases, only when the assumption that the LSM of σ is unique is satisfied, the same method can be used to find the T_{LS}-optimum design.

Table 1 shows the numerical results of the T_{LS}-optimum designs for discrimination and with minimum biases. F_1, F_2, F_3 and F_4 stands for the probit, logit, double-exponential and double-reciprocal models, respectively and σ, a, b and w are defined as before.

Table 1: Results of the T_{LS}-optimum design with true parameters (0,1)

Models		Supports	Weights	
True , Assumed	σ^*	a, b	w	Maximum bias
F_1, F_2	.5876	.5710 , 2.0430	.1716	.00946
F_1, F_3	.8744	.3769 , 1.9104	.2337	.02820
F_1, F_4	.4770	.3157 , 2.2431	.1882	.07523
F_2, F_3	1.495	.3201 , 1.7750	.2571	.01978
F_2, F_4	.8194	.2945 , 2.3520	.1873	.06702
F_3, F_4	.5317	.3270 , 3.3381	.1673	.05095

3 Efficiency and bias comparisons

In this section, we are going to illustrate the relationship between T_{LS}-optimum design for model discrimination and minimum bias design for model robustness. In the process of searching for the desired T_{LS}-optimum design, results of Huang and Hwang (2004) are adopted where the maximum bias is minimized. Hence, the design we propose now not only can discriminate between models with reasonable power but also has the model robustness property that minimizes the maximum bias of the assumed model.

For binary data models, there are two useful measures for testing model adequacies are available, the Pearson's chi-squared statistic and the log likelihood ratio statistic. The result of Yanagisawa (1988), say T_{PS}-optimum design, is based on maximizing the noncentrality parameter of the Pearson's chi-squared statistic (1.1), while that of Müller and Ponce de Leon (1996) is based on maximizing the log likelihood ratio statistic (1.2) to find the T_{LR}-optimum design. We compare performances of the T_{LS}-optimum design, with theirs. Table 2 and Table 3 report $N(.5)$, the minimum number of observations necessary to achieve a power of 50% at a significance level 5%, and the corresponding efficiency and maximum bias when the true model is probit(0,1) and the assumed model is logitand vice versa.

The design we propose may not be the most powerful with respect to the Pearson's goodness of fit test or the likelihood ratio teat, but it possesses the good property that minimizes the maximum bias of the assumed model. In other words, if we make the wrong choice of the true model, the estimated probability at any quantile would not be far from the true one and the maximum bias would be bounded by a small value anywhere. The efficiencies of T_{LS}-optimum design w.r.t. T_{PS}- and T_{LR}-optimum design are 71.5%($=\frac{1629}{2280}$) and 64.6%, respectively when the true model is from probit model and are 52.4% and 71.5%, respectively when the true model is from logit model.

Table 2: Comparison of efficiencies for H_0:logit(μ, σ) v.s. H_A:probit(0,1) case

criteria	(μ, σ)	$N(.5)$		max|bias|	efficiency
T_{LS}-	(0, 0.5876)	2280	2049	0.0095	
T_{PS}-	(0, 0.5553)	1629	–	0.0221	$\text{eff}_{PS}(\text{T}_{LS}) = 71.5\%$
T_{LR}-	(0, 0.5528)	–	1323	0.0221	$\text{eff}_{LR}(\text{T}_{LS}) = 64.6\%$

Table 3: Comparison of efficiencies for H_0:probit(μ, σ) v.s. H_A:logit(0,1) case

criteria	(μ, σ)	$N(.5)$		max|bias|	efficiency
T_{LS}-	(0, 1.7018)	1614	1827	0.0095	
T_{PS}-	(0, 1.9311)	845	–	0.0364	$\text{eff}_{PS}(\text{T}_{LS}) = 52.4\%$
T_{LR}-	(0, 1.8332)	–	1036	0.0250	$\text{eff}_{LR}(\text{T}_{LS}) = 71.5\%$

3.1 Numerical computation of the corresponding bias and MSE

Some numerical computation and simulations are presented here. First, we compute the theoretical value of the probability bias of q^{th} quantile as the number of observations $N \to \infty$, i.e.

$$bias_{\infty}(q) = q - F(G^{-1}(q; \mu_0, \sigma^*); \mu_0, \sigma_0),$$

where $F(\cdot;\mu_0,\sigma_0)$ is the true model, and $G(\cdot;\mu_0,\sigma^*)$ is the assumed model. We also point out the maximum and minimum of $bias_\infty(q)$ for all q >0.5, say q_M and q_m from each design respectively to compare the maximum probability bias. Then, we use 2000 observations per simulation run to compute the simulated probability biases and mean square errors of some q^{th} quantiles and the number of simulation runs is 1000. The reason we use 2000 observations per run is that about 2000 observations are needed for T_{LS}-optimum design to attain power 50% at significance level 5%. For $q \in (\frac{1}{2}, 1)$, the simulated 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 $\tilde{x}_{q,n} = G^{-1}(q;\hat{\mu},\hat{\sigma})$, $\hat{\mu}$ and $\hat{\sigma}$ are the $LSEs$.

3.2 The probit and logit case

When the true model is probit with mean 0 and variance 1 but the assumed model is logit, the T_{LS}-, T_{PS}- and T_{LR}-optimum design are compared and the results are presented in Table 4 - Table 6, respectively. We can see that the maximum bias from the T_{LS}-optimum design is smaller than that of the others but the mean square errors are not outstanding over all design region, but for region with quantiles near the one with the maximum bias are smaller than those of other designs. This is not surprising as each design should perform quite well near the quantiles where the design supports are.

The case when the true model is logit with mean 0 and variance 1 but the assumed model is probit, the T_{LS}-, T_{PS}- and T_{LR}-optimum design are also compared and the results are presented in Table 7 - Table 9, respectively. The results are similar to those of the previous cases that the maximum bias of T_{LS}-optimum design is the smallest in the three designs and the mean square errors of the T_{LS}-optimum design are smaller than those of the other designs near the quantile with minimum bias.

3.3 Simulation of small sample size in the probit and logit case

In Table 2, we can see that $N(.5)$, the minimum number of observations necessary to achieve a power of 50% at a significance level 5% of the three optimum designs respectively, is greater than a thousand when the true model is with probit model. When the true model is with logit model, $N(.5)$ is greater than 800 which is also a large number of observations. Since in realistic situation, it is usually difficult to carry out a experiment with more than a thousand observations, we do some small sample size simulations to compare the performance of the T_{PS}- and the T_{LS}-optimum designs. The number of simulation sample size is set to be 30, 50 and 80 respectively. The simulation procedure is the same as the previous simulation which is with large sample size but the simulation run is 10,000 for each case. Besides the mean squared errors and the biases of the probability, we also present the mean squared errors and the biases of quantile. The results are presented in Table 10 -15.

The estimated value of parameters are also recorded and presnted in Table 16. We can see that for location parameter μ, the estimation of the T_{LS}-optimum design is better than that of the T_{PS}-optimum design, which means the estimator of μ in T_{LS}-optimum design is closer to the true parameter than the estimator of μ in T_{PS}-optimum design.

Table 4: Bias and MSE on T_{LS}-optimum design under probit model with misspecified logit link function ($var \times 10^{-5}$)

q	x_q	$\tilde{x}_q$	$F(\tilde{x}_q)$	$bias_\infty$	$\overline{\tilde{x}_q}$	$\overline{F(\tilde{x}_q)}$	$bias$	var	$\sqrt{MSE}$
0.51	0.0251	0.0235	0.5094	0.0006	0.0257	0.5102	-0.0002	31.3	0.0177
0.55	0.1257	0.1179	0.5469	0.0031	0.1197	0.5476	0.0024	31.4	0.0179
0.6	0.2533	0.2383	0.5942	0.0058	0.2397	0.5946	0.0054	31.5	0.0185
0.7	0.5244	0.4979	0.6907	0.0093	0.4984	0.6907	0.0093	31.4	0.0200
0.7255_{q_M}	0.5993	0.5711	0.7160	0.0095	0.5714	0.7159	0.0096	31.1	0.0200
0.8	0.8416	0.8146	0.7924	0.0076	0.8140	0.7918	0.0082	28.7	0.0188
0.9	1.2816	1.2912	0.9017	-0.0017	1.2889	0.9006	-0.0006	17.9	0.0134
0.9700_{q_m}	1.8808	2.0427	0.9795	-0.0095	2.0378	0.9786	-0.0086	3.2	0.0103
0.99	2.3264	2.7002	0.9965	-0.0065	2.6931	0.9962	-0.0062	0.2	0.0064

Table 5: Bias and MSE on T_{PS}-optimum design under probit model with misspecified logit link function

q	x_q	$\tilde{x}_q$	$F(\tilde{x}_q)$	$bias_\infty(q)$	$\text{var}^{asy.}$	$\sqrt{MSE}$
0.51	0.0251	0.0222	0.5089	0.0012	0.000224	0.0150
0.55	0.1257	0.1114	0.5444	0.0057	0.000224	0.0160
0.6	0.2533	0.2252	0.5891	0.0109	0.000225	0.0186
0.7	0.5244	0.4705	0.6810	0.0190	0.000225	0.0242
0.7713_{q_M}	0.7431	0.6751	0.7502	0.0221	0.000222	0.0258
0.8	0.8416	0.7698	0.7793	0.0207	0.000205	0.0252
0.9	1.2816	1.2202	0.8888	0.0112	0.000128	0.0159
0.9853_{q_m}	2.1793	2.3369	0.9903	-0.0049	0.000023	0.0069
0.99	2.3264	2.5518	0.9946	-0.0046	0.000002	0.0048

Table 6: Bias and MSE on T_{LR}-optimum design under probit model with misspecified logit link function

q	x_q	$\tilde{x}_q$	$F(\tilde{x}_q)$	$bias_\infty(q)$	$\text{var}^{asy.}$	$\sqrt{MSE}$
0.51	0.0251	0.0221	0.5088	0.0012	0.000203	0.0143
0.55	0.1257	0.1109	0.5442	0.0058	0.000203	0.0154
0.6	0.2533	0.2241	0.5887	0.0113	0.000203	0.0182
0.7	0.5244	0.4684	0.6802	0.0198	0.000203	0.0244
0.7739_{q_M}	0.7518	0.6802	0.7518	0.0221	0.000201	0.0263
0.8	0.8416	0.7663	0.7783	0.0217	0.000185	0.0256
0.9	1.2816	1.2146	0.8877	0.0123	0.000116	0.0163
0.9862_{q_m}	2.2029	2.3600	0.9909	-0.0047	0.000021	0.0065
0.99	2.3264	2.5401	0.9965	-0.0045	0.000002	0.0047

Table 7: Bias and MSE on T_{LS}-optimum design ($var \times 10^{-5}$)

q	x_q	$\tilde{x}_q$	$F(\tilde{x}_q)$	$bias_\infty(q)$	$\overline{\tilde{x}_q}$	$\overline{F(\tilde{x}_q)}$	$bias$	var	$\sqrt{MSE}$
0.51	0.0221	0.0235	0.5107	-0.0007	0.0238	0.5108	-0.0008	32.5	0.0180
0.55	0.1106	0.1179	0.5533	-0.0033	0.1177	0.5531	-0.0031	32.4	0.0183
0.7	0.4671	0.4920	0.7094	-0.0094	0.4896	0.7082	-0.0081	33.0	0.0199
0.7160_{q_m}	0.5098	0.5357	0.7255	-0.0095	0.5311	0.7241	-0.0081	32.9	0.0198
0.8	0.7643	0.7897	0.8073	-0.0073	0.7855	0.8055	-0.0055	29.6	0.0181
0.9	1.2114	1.2024	0.8985	0.0015	1.1959	0.8967	0.0034	18.7	0.0141
0.9795_{q_M}	2.1318	1.9174	0.9700	0.0095	1.9066	0.9688	0.0107	4.5	0.0126
0.99	2.5334	2.1827	0.9813	0.0087	2.1704	0.9803	0.0097	2.4	0.0108

Table 8: Bias and MSE on T_{PS}-optimum design for logit$(0, \frac{\sqrt{3}}{\pi})$ and probit model

q	x_q	$\tilde{x}_q$	$F(\tilde{x}_q)$	$bias_\infty(q)$	var$^{asy.}$	$\sqrt{MSE}$
. 0.51	0.0221	0.0267	0.5121	-0.0012	0.000170	0.0131
0.55	0.1106	0.1338	0.5604	-0.0104	0.000170	0.0142
0.7	0.4671	0.5583	0.7335	-0.0335	0.000173	0.0231
0.7651_{q_m}	0.6510	0.7696	0.8015	-0.0364	0.000172	0.0249
0.8	0.7643	0.8961	0.8355	-0.0355	0.000155	0.0242
0.9	1.2114	1.3644	0.9224	-0.0224	0.000098	0.0150
0.99	2.5334	2.4768	0.9889	0.0011	0.000024	0.0069
0.9965_{q_M}	3.1158	2.8713	0.9946	0.0019	0.000013	0.0058

Table 9: Bias and MSE on T_{LR}-optimum design for logit$(0, \frac{\sqrt{3}}{\pi})$ and probit model

q	x_q	$\tilde{x}_q$	$F(\tilde{x}_q)$	$bias_\infty(q)$	var$^{asy.}$	$\sqrt{MSE}$
0.51	0.0221	0.0253	0.5115	-0.0015	0.000232	0.0153
0.55	0.1106	0.1270	0.5573	-0.0073	0.000232	0.0163
0.7	0.4671	0.5300	0.7234	-0.0234	0.000236	0.0251
0.7557_{q_m}	0.6226	0.6999	0.7807	-0.0250	0.000235	0.0269
0.8	0.7643	0.8506	0.8239	-0.0239	0.000212	0.0261
0.9	1.2114	1.2952	0.9129	-0.0129	0.000134	0.0169
0.99	2.5334	2.3512	0.9861	0.0039	0.000032	0.0074
0.9862_{q_M}	2.6934	2.4584	0.9886	0.0040	0.000017	0.0061

In the case of sample size 80 and in the sense of mean squared error, the performance of the T_{LS}-optimum design is better than the T_{PS}-optimum design over where the

Table 10: Bias and MSE on T_{LS}-optimum design under probit model with misspecified logit link function with sample size 30 ($var \times 10^{-5}$)

q	$\sqrt{MSE(q)}$	$bias_{30}(q)$	$var_{30}(q)$	$\overline{MSE(\tilde{x}_q)}$	$\overline{bias(\tilde{x}_q)}$	$var(\tilde{x}_q)$
.01	.25658	-.13456	.04773	1.45951	-.38179	1.9844
.03	.26042	-.13964	.04832	1.17724	-.41005	1.2178
.1	.24835	-.13192	.04428	.85196	-.35231	.6017
.2	.22008	-.10541	.03732	.64824	-.25574	.3548
.2745	.19860	-.08138	.03282	.55441	-.18884	.2717
.3	.19179	-.07267	.03150	.52936	-.16674	.2524
.4	.17056	-.03710	.02771	.46053	-.08277	.2052
.45	.16456	-.01879	.02673	.44312	-.04179	.1946
.5	.16243	-.00034	.02638	.43713	-.00110	.1911
.55	.16438	.01810	.02669	.44267	.03959	.1944
.6	.17021	.03642	.02765	.45968	.08057	.2048
.7	.19123	.07199	.03139	.52783	.16454	.2515
0.7255	.19801	.08069	.03269	.55274	.18663	.2707
.8	.21944	.10475	.03718	.64625	.25354	.3534
.9	.24769	.13135	.04409	.84969	.35011	.5994
0.97	.25967	.13926	.04803	1.17492	.40785	1.2141
.99	.25567	.13428	.04734	1.45727	.37959	1.9796

Table 11: Bias and MSE on T_{PS}-optimum design under probit model with misspecified logit link function with sample size 30

q	$\sqrt{MSE(q)}$	$bias_{30}(q)$	$var_{30}(q)$	$\overline{MSE(\tilde{x}_q)}$	$\overline{bias(\tilde{x}_q)}$	$var(\tilde{x}_q)$
.01	.253841	-.108778	.052603	1.7895	-.03032	3.20138
.03	.260781	-.111648	.055541	1.4799	-.12239	2.17509
.1	.252199	-.103259	.052942	1.1371	-.13758	1.27409
.2	.230926	-.077098	.047383	.9277	-.08725	.85301
.2745	.216298	-.053116	.043964	.8275	-.04398	.68286
.3	.211889	-.044480	.042919	.7991	-.02899	.63765
.4	.198464	-.009882	.039290	.7074	.02979	.49953
.45	.194333	.007280	.037712	.6714	.05908	.44725
.5	.191979	.023960	.036282	.6411	.08833	.40323
.55	.191318	.039873	.035013	.6166	.11758	.36636
.6	.192171	.054720	.033936	.5981	.14688	.33609
.7	.197392	.079850	.032588	.5817	.20565	.29612
0.7255	.199251	.084984	.032479	.5829	.22064	.29110
.8	.205253	.095997	.032913	.6032	.26392	.29422
.9	.213461	.098250	.035913	.6980	.31424	.38844
0.97	.217659	.089125	.039432	.9292	.29905	.77396
.99	.211433	.084390	.037582	1.1799	.20698	1.34920

Table 12: Bias and MSE on T_{LS}-optimum design under probit model with misspecified logit link function with sample size 50 ($var \times 10^{-5}$)

q	$\sqrt{MSE(q)}$	$bias_{50}(q)$	$var_{50}(q)$	$\overline{MSE(\tilde{x}_q)}$	$\overline{bias(\tilde{x}_q)}$	$var(\tilde{x}_q)$
.01	.150881	-.054650	.0197783	1.049610	-.044518	1.09969
.03	.159039	-.063857	.0212158	.835982	-.153991	.67515
.1	.164593	-.071215	.0220191	.609238	-.189062	.33542
.2	.155993	-.062719	.0204003	.472543	-.151348	.20039
.2745	.146007	-.050166	.0188012	.410594	-.114518	.15547
.3	.142596	-.045131	.0182968	.394267	-.101466	.14515
.4	.131764	-.023078	.0168290	.350544	-.049558	.12042
.45	.128898	-.011191	.0164894	.340287	-.023449	.11524
.5	.128255	.000936	.0164484	.337693	.002680	.11403
.55	.129934	.013075	.0167118	.342730	.028810	.11663
.6	.133771	.024998	.0172696	.355312	.054919	.12323
.7	.146187	.047192	.0191436	.403021	.106826	.15101
.7255	.149915	.052278	.0197415	.420202	.119878	.16219
.8	.160639	.065006	.0215791	.484295	.156708	.20998
.9	.169808	.073771	.0233926	.623242	.194423	.35063
.97	.163948	.066446	.0224639	.851235	.159351	.69920
.99	.155013	.056990	.0207811	1.064880	.049878	1.13149

Table 13: Bias and MSE on T_{PS}-optimum design under probit model with misspecified logit link function with sample size 50

q	$\sqrt{MSE(q)}$	$bias_{50}(q)$	$var_{50}(q)$	$\overline{MSE(\tilde{x}_q)}$	$\overline{bias(\tilde{x}_q)}$	$var(\tilde{x}_q)$
.01	.262793	-.105654	.0578972	1.509840	-.095504	2.270490
.03	.265775	-.110033	.0585292	1.261560	-.199595	1.551690
.1	.255472	-.110112	.0531412	.980360	-.228516	.908886
.2	.232912	-.093568	.0454931	.796843	-.186901	.600026
.2745	.214648	-.075497	.0403740	.702645	-.148078	.471783
.3	.208372	-.068706	.0386984	.674687	-.134427	.437132
.4	.184520	-.040629	.0323968	.579064	-.080394	.328852
.45	.173416	-.026355	.0293785	.537791	-.053300	.286379
.5	.163080	-.012341	.0264429	.500188	-.026205	.249502
.55	.153623	.001117	.0235986	.466174	.000889	.217317
.6	.145104	.013719	.0208669	.435996	.027983	.189309
.7	.130895	.035009	.0159080	.390201	.082015	.145530
.7255	.127854	.039314	.0148012	.382685	.095667	.137296
.8	.120324	.048330	.0121420	.375519	.134490	.122927
.9	.114014	.049138	.0105846	.428619	.176105	.152702
.97	.112601	.039370	.0111290	.614037	.147184	.355378
.99	.109452	.035372	.0107286	.831216	.043093	.689063

Table 14: Bias and MSE on T_{LS}-optimum design under probit model with misspecified logit link function with sample size 80 ($var \times 10^{-5}$)

q	$\sqrt{MSE(q)}$	$bias_{80}(q)$	$var_{80}(q)$	$\overline{MSE(\tilde{x}_q)}$	$\overline{bias(\tilde{x}_q)}$	$var(\tilde{x}_q)$
.01	.069131	-.0169319	.0044924	.782416	.1483130	.590178
.03	.083326	-.0257290	.0062813	.605440	-.0095499	.366466
.1	.105157	-.0388276	.0095503	.443330	-.0999234	.186557
.2	.111039	-.0399478	.0107338	.351861	-.0972761	.114344
.2745	.108557	-.0340691	.0106239	.310045	-.0783654	.089987
.3	.107205	-.0312209	.0105183	.298866	-.0707019	.084322
.4	.101888	-.0173906	.0100787	.268057	-.0379008	.070418
.45	.100122	-.0094748	.0099346	.260137	-.0206477	.067245
.5	.099393	-.0012785	.0098775	.257209	-.0031952	.066146
.55	.099806	.0069323	.0099133	.259307	.0142572	.067037
.6	.101266	.0148924	.0100329	.266440	.0315103	.069997
.7	.106014	.0289041	.0104035	.295936	.0643114	.083442
.7255	.107229	.0318197	.0104856	.306853	.0719749	.088978
.8	.109328	.0379501	.0105123	.348088	.0908856	.112905
.9	.102951	.0373201	.0092060	.439345	.0935329	.184276
.97	.080881	.0247384	.0059298	.602385	.0031594	.362858
.99	.066750	.0162072	.0041929	.780603	-.1547030	.585408

Table 15: Bias and MSE on T_{PS}-optimum design under probit model with misspecified logit link function with sample size 80

q	$\sqrt{MSE(q)}$	$bias_{80}(q)$	$var_{80}(q)$	$\overline{MSE(\tilde{x}_q)}$	$\overline{bias(\tilde{x}_q)}$	$var(\tilde{x}_q)$
.01	.146622	-.0339929	.0203425	.920978	.102371	.837721
.03	.150370	-.0408726	.0209405	.756414	-.040194	.570546
.1	.154778	-.0516065	.0212931	.590264	-.113086	.335623
.2	.151835	-.0478080	.0207683	.485226	-.099353	.225573
.2745	.145711	-.0378339	.0198003	.431960	-.074778	.180997
.3	.143215	-.0336105	.0193810	.416430	-.065411	.169135
.4	.132787	-.0148207	.0174128	.365415	-.026570	.132822
.45	.127706	-.0047980	.0162857	.345071	-.006517	.119032
.5	.122986	.0051818	.0150988	.327985	.013678	.107387
.55	.118727	.0147975	.0138771	.314228	.033873	.097591
.6	.114926	.0237219	.0126452	.303991	.053926	.089502
.7	.108162	.0380422	.0102517	.295322	.092767	.078609
.7255	.106423	.0406214	.0096757	.295956	.102134	.077158
.8	.100536	.0445719	.0081208	.305805	.126710	.077461
.9	.088093	.0376753	.0063410	.347267	.140442	.100871
.97	.073216	.0199454	.0049628	.451355	.067550	.199159
.99	.066708	.0130310	.0042801	.593637	-.075014	.346777

Table 16: Estimated value of parameters

	T_{LS}		T_{PS}	
Sample size	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\mu}$	$\hat{\sigma}$
30	.00109999	.423418	-.0883313	.480444
50	-.00268018	.495994	.0262055	.491184
80	.00319525	.539236	-.0136781	.525566

probability is smaller than 0.7. Although the performance of the T_{PS}-optimum design is better than that of the T_{LS}-optimum design, they are not different very much and the explanation of this phenomenon is one of the support points of the T_{PS}-optimum design is allocated at extreme quantile which makes the performance better over there. In the case of sample size 30 and 50, the performances of the two kinds optimum designs are not far away form each other in the sense of mean squared error. However, in the sense of mean squared error of quantile, the T_{LS}-optimum design performs better than the T_{PS}-optimum design whatever the sample size is.

In Table 16, the simulated estimation parameters are presented. We can see that in small sample size case, the estimator of location parameter of the T_{LS}-optimum design is closer to the true location parameter than the estimator of location parameter of the T_{PS}-optimum design. We guess that this phenomenon may be the reason which cause the smaller mean squared error of the T_{PS}-optimum design than that of the T_{LS}-optimum design. There are more simulation cases could be carry out to compare the performances of the two different kinds of optimum designs in the future.

4 Discussions and conclusions

In Table 2 and Table 3, the efficiencies of the T_{LS}-optimum design with respect to the T_{PS}-optimum design and T_{LR}-optimum design are presented. Both Yanagisawa (1988) and Ponce de Leon (1996) stated that it is easier to discriminate between models when the true model is logit and the result of T_{LS}-optimum design is in accordance with their findings.

The T_{LS}-optimum designs not only can discriminate between models but also possess the model robust property that minimizes the maximum bias. In fact, the A- and D-optimum designs do not possess the ability of discrimination between models. Table 2 and Table 3 also show the maximum bias under each design, respectively. In Table 3 we can see that although the efficiency of the T_{LS}-optimum design with respect to Yanagisawa's optimum design under Pearson's goodness of fit test is about 52.4%, but it would be with higher risk if one uses the Yanagisawa's design and chooses the wrong model.

The maximum bias is about 3.64% in the above situation which is relatively large to that of ours which is less than 1%. We can also see from the two tables that when the rival models are probit and logit models the maximum bias of each case is larger than 2.2% which might be somewhat risky if a slight difference of the probability may cause serious consequences, for example, the pressure applied to the explosive or poisonous substances. In this situation, one might rather to bound the probability bias but not to care what exactly the true model is.

Optimum designs for discrimination have been discussed in several articles under different criteria. Although T_{LS}-optimum designs are not most powerful for model discrimination with respect to the Pearson's chi-squared goodness of fit test or the log likelihood ratio test, the T_{LS}-optimum designs still possess some advantages. We suggest to use T_{LS}-optimum designs when one is in the situation that it is difficult to discriminate between the two rival models and a slight bias would cause serious consequence that one may want to control the maximum probability bias.

Since the T_{LS}-criterion here is based on minimizing the ordinary sum of squared deviations, another statistic can be considered to test the null hypothesis, that is, use the difference between the sum of squares of the two fitted models,

$$dSSE = SSE_0 - SSE_A$$

and one would reject the null hypothesis if the value of $dSSE$ is large. At present, we do not know the exact or asymptotic distribution of the statistic, $dSSE$. About SSE_0 and SSE_A, we speculate that both of them might be distributed as linear combinations of chi-squared distributions. We will investigate the theoretical properties more in the future.

To understand the distribution of $dSSE$, a simulation study has been done to examine the results for the case when the true model is with probit(0,1) and the assumed model is logit. The observations per simulation run is 2,000 and the total simulation runs is 10,000. We suggest to reject the null hypothesis at the significance level 5% when $dSSE > 0.181$. The power of the test statistic, $dSSE$, at the significance level 5% is about 62.1%. However, the critical value of the statistic, $dSSE$, is related to the scale parameter of the true model, it would be of interest to find the null distribution of statistic proposed with respect to the scale parameter of the null distribution.

References

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

[2] Chambers, E. A. and Cox, D.R. (1967). Discrimination between alternative binary response models. *Biometrika*, 54, No.3/4, 573-578.

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

[4] Huang, M-N. L. and Hwang, S-H. (2004). Model robust designs for binary response experiments. *Department of Applied Mathematics, National Sun Yat-sen University*, Manuscript.

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

[6] Müller, 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.

[7] 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.

[8] 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.

[9] Uciński, D. and Bogacka, B. (2004). T-optimum designs for discrimination between two multiresponse dynamic models. *Journal of the Royal Statistical Society Series B*, 67, No.1, 3-18.

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

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

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

[13] Yanagisawa, Y. (1990). Designs for discrimination between bivariate binary response models. *Biometrical Journal*, 32, 25-34.

A Appendix

Proof of Lemma 2.1. We would first prove that for each asymmetric design ξ_1, we can find a symmetric design $\tilde{\xi}$, which is a linear combination of designs ξ_1 and ξ_2 with equal weight, $\frac{1}{2}$, where ξ_2 is the reflected design of ξ_1 with respect to the true mean parameter μ_0. The performance of $\tilde{\xi}$ would be better than ξ_1 in a sense of T_{LS}-optimality. The mathematical expression is

$$T_{LS}(\tilde{\xi}) = T_{LS}(\frac{\xi_1+\xi_2}{2}) \geq \frac{T_{LS}(\xi_1)+T_{LS}(\xi_2)}{2} = T_{LS}(\xi_1), \quad \text{where} \quad \tilde{\xi} = \frac{\xi_1+\xi_2}{2},$$

$$\xi_1 = \left\{ \begin{array}{ccc} \mu_0 + x_1 & ... & \mu_0 + x_n \\ w_1 & ... & w_n \end{array} \right\} \quad \text{and} \quad \xi_2 = \left\{ \begin{array}{ccc} \mu_0 - x_1 & ... & \mu_0 - x_n \\ w_1 & ... & w_n \end{array} \right\},$$

$$0 < w_i < 1 \quad \text{and} \quad \sum_{i=1}^{n} w_i = 1.$$

The last equality holds if $T_{LS}(\xi_1) = T_{LS}(\xi_2)$ and the inequality holds if the mapping $\xi \to T_{LS}(\xi)$ is concave. In the following we will prove the following two properties, concavity and symmetry, of the mapping, $\xi \to T_{LS}(\xi)$.

We show that the mapping $\xi \to T_{LS}(\xi)$ is concave first. Let $\xi = \alpha\xi_1 + (1-\alpha)\xi_2$, by the definition of the mapping $\xi \to T_{LS}(\xi)$, it can be seen that

$$\begin{aligned}
T_{LS}(\xi) &= \min_{\mu,\sigma} \textstyle\int_\chi \|F(\frac{x-\mu_0}{\sigma_0}) - G(\frac{x-\mu}{\sigma})\|^2 \xi(dx), \\
&= \textstyle\int_\chi \|F(\frac{x-\mu_0}{\sigma_0}) - G(\frac{x-\hat{\mu}}{\hat{\sigma}})\|^2 (\alpha\xi_1 + (1-\alpha)\xi_2)(dx) \\
&= \alpha \textstyle\int_\chi \|F(\frac{x-\mu_0}{\sigma_0}) - G(\frac{x-\hat{\mu}}{\hat{\sigma}})\|^2 \xi_1(dx) + (1-\alpha)\int_\chi \|F(\frac{x-\mu_0}{\sigma_0}) - G(\frac{x-\hat{\mu}}{\hat{\sigma}})\|^2 \xi_2(dx) \\
&\geq \alpha \textstyle\int_\chi \|F(\frac{x-\mu_0}{\sigma_0}) - G(\frac{x-\hat{\mu}_1}{\hat{\sigma}_1})\|^2 \xi_1(dx) + (1-\alpha)\int_\chi \|F(\frac{x-\mu_0}{\sigma_0}) - G(\frac{x-\hat{\mu}_2}{\hat{\sigma}}_2)\|^2 \xi_2(dx) \\
&= \alpha \min_{\mu,\sigma} \textstyle\int_\chi \|F(\frac{x-\mu_0}{\sigma_0}) - G(\frac{x-\mu}{\sigma})\|^2 \xi_1(dx) \\
&\qquad + (1-\alpha) \min_{\mu,\sigma} \textstyle\int_\chi \|F(\frac{x-\mu_0}{\sigma_0}) - G(\frac{x-\mu}{\sigma})\|^2 \xi_2(dx) \\
&= \alpha T_{LS}(\xi_1) + (1-\alpha) T_{LS}(\xi_2).
\end{aligned}$$

Thus, it is proved immediately that the mapping $\xi \rightarrow T_{LS}(\xi)$ is concave and then the special case holds for

$$T_{LS}(\tilde{\xi}) = T_{LS}(\frac{\xi_1 + \xi_2}{2}) \geq \frac{T_{LS}(\xi_1) + T_{LS}(\xi_2)}{2}, \quad \text{where} \quad \tilde{\xi} = \frac{\xi_1 + \xi_2}{2}.$$

We now prove that $T_{LS}(\xi_1) = T_{LS}(\xi_2)$. Once the statement is verified, then we have the result that $T_{LS}(\tilde{\xi}) \geq T_{LS}(\xi_1)$ where $\tilde{\xi}$ and ξ_1 are as defined in Lemma 2.1.

Recall the definition of

$$T_{LS}(\xi) = \min_{\mu,\sigma} \int_{\chi} \|F(\frac{x-\mu_0}{\sigma_0}) - G(\frac{x-\mu}{\sigma})\|^2 \xi(dx),$$

then

$$\begin{aligned} T_{LS}(\xi_1) &= \min_{\mu_1,\sigma_1} \textstyle\int_{\chi} \|F(\frac{x-\mu_0}{\sigma_0}) - G(\frac{x-\mu_1}{\sigma_1})\|^2 \xi_1(dx) \\ &= \min_{\mu_1,\sigma_1} \textstyle\sum_{i=1}^n w_i [F(\frac{x_i}{\sigma_0}) - G(\frac{x_i + (\mu_0 - \mu_1)}{\sigma_1})]^2 \end{aligned}$$

and

$$\begin{aligned} T_{LS}(\xi_2) &= \min_{\mu_2,\sigma_2} \textstyle\int_{\chi} \|F(\frac{x-\mu_0}{\sigma_0}) - G(\frac{x-\mu_2}{\sigma_2})\|^2 \xi_2(dx) \\ &= \min_{\mu_2,\sigma_2} \textstyle\sum_{i=1}^n w_i [F(\frac{-x_i}{\sigma_0}) - G(\frac{-x_i + (\mu_0 - \mu_2)}{\sigma_2})]^2, \end{aligned}$$

where $0 < w_i < 1$ and $\sum_{i=1}^n w_i = 1$ and ξ_2 is the reflected design of ξ_1 with respect to μ_0. More explicitly, it can be expressed as

$$\xi_1 = \left\{ \begin{array}{ccc} \mu_0 + x_1 & \dots & \mu_0 + x_n \\ w_1 & \dots & w_n \end{array} \right\} \qquad \text{and} \qquad \xi_2 = \left\{ \begin{array}{ccc} \mu_0 - x_1 & \dots & \mu_0 - x_n \\ w_1 & \dots & w_n \end{array} \right\}.$$

Let the LSE of μ_1 be denoted by $\hat{\mu}_1$, i.e.

$$\begin{aligned} \hat{\mu_1} &= arg \min_{\mu_1} \int_{\xi} [F(\frac{x-\mu_0}{\sigma_0}) - G(\frac{x-\mu_1}{\sigma_1})]^2 \xi_1(dx) \\ &= arg \min_{\mu_1} \sum_{i=1}^n w_i [F(\frac{x_i}{\sigma_0}) - G(\frac{x_i + (\mu_0 - \mu_1)}{\sigma_1})]^2. \end{aligned}$$

Define $\hat{\mu}_1 = \mu_0 + \mu^*$, that is to say $\mu_0 - \hat{\mu}_1 = -\mu^*$. Then another expression of $\hat{\mu}_1$ is that

$$\hat{\mu}_1 = arg \min_{\mu_1} \sum_{i=1}^n w_i [F(\frac{x_i}{\sigma_0}) - G(\frac{x_i - \mu^*}{\sigma_1})]^2.$$

We show that the relationship between $\hat{\mu}_1$ and $\hat{\mu}_2$ is $\hat{\mu}_2 = \mu_0 - \mu^*$. Since

$$\begin{aligned}
& \min_{\mu_2} \sum_{i=1}^{n} w_i [F(\frac{-x_i}{\sigma_0}) - G(\frac{-x_i + (\mu_0 - \mu_2)}{\sigma_2})]^2 \\
= & \min_{\mu_2} \sum_{i=1}^{n} w_i [1 - F(\frac{x_i}{\sigma_0}) - 1 + G(\frac{x_i - (\mu_0 - \mu_2)}{\sigma_2})]^2 \\
= & \min_{\mu_2} \sum_{i=1}^{n} w_i [F(\frac{x_i}{\sigma_0}) - G(\frac{x_i - (\mu_0 - \mu_2)}{\sigma_2})]^2,
\end{aligned}$$

then $(\mu_0 - \hat{\mu}_2) = \mu^*$, i.e. $\hat{\mu}_2 = \mu_0 - \mu^*$. Let the LSE of σ_1 is $\hat{\sigma}_1$ which can be expressed as

$$\begin{aligned}
\hat{\sigma}_1 &= arg \min_{\sigma_1} \sum_{i=1}^{n} w_i [F(\frac{x_i}{\sigma_0}) - G(\frac{x_i + (\mu_0 - \hat{\mu}_1)}{\sigma_1})]^2, \quad \text{where} \quad \hat{\mu}_1 = \mu_0 + \mu^* \\
&= arg \min_{\sigma_1} \sum_{i=1}^{n} w_i [F(\frac{x_i}{\sigma_0}) - G(\frac{x_i - \mu^*}{\sigma_1})]^2.
\end{aligned}$$

Now we show that $\hat{\sigma}_2 = \hat{\sigma}_1$. If this statement is proved, then it would be easy to see that $T_{LS}(\xi_1) = T_{LS}(\xi_2)$ and $T_{LS}(\tilde{\xi}) \geq T_{LS}(\xi_1)$. We know that from the definition

$$\begin{aligned}
\hat{\sigma}_2 &= arg \min_{\sigma} {}_2 \sum_{i=1}^{n} w_i [F(\frac{-x_i}{\sigma_0}) - G(\frac{-x_i + (\mu_0 - \hat{\mu}_2)}{\sigma_2})]^2, \quad \text{where} \quad \hat{\mu}_2 = \mu_0 - \mu^* \\
&= arg \min_{\sigma_2} \sum_{i=1}^{n} w_i [F(\frac{-x_i}{\sigma_0}) - G(\frac{-x_i + \mu^*}{\sigma_2})]^2 \\
&= arg \min_{\sigma_1} \sum_{i=1}^{n} w_i [F(\frac{x_i}{\sigma_0}) - G(\frac{x_i - \mu^*}{\sigma_1})]^2 \\
&= \hat{\sigma}_1.
\end{aligned}$$

Then, by previous result, we can say that

$$T_{LS}(\xi_2) = T_{LS}(\xi_1) \quad \text{and} \quad T_{LS}(\tilde{\xi}) \geq T_{LS}(\xi_1),$$

where $\tilde{\xi}$, ξ_1 and ξ_2 are as defined in Lemma 2.1. Then by previous statement, we can say that there must exist a symmetric T_{LS}-optimum design. Thus, we may concentrate our attentions on those designs which are symmetric in finding a T_{LS}-optimum design.

Proof of Lemma 2.2. Below we prove that the unique minimizer of μ is μ_0 for symmetric design with respect to μ_0 with four support points. For all symmetric designs with more than five support points, the proof of the uniqueness of minimizer of μ is similar and omitted. Now we consider the design

$$\xi^* = \left\{ \begin{array}{cccc} \mu_0 - b & \mu_0 - a & \mu_0 + a & \mu_0 + b \\ \frac{1}{2} - w & w & w & \frac{1}{2} - w \end{array} \right\},$$

where $a, b > 0$, then the least square estimates of μ is μ_0. Let

$$\begin{aligned}
h(\mu, \sigma) &= \int_{\chi} [F(\frac{x-\mu_0}{\sigma_0}) - G(\frac{x-\mu}{\sigma})]^2 \xi(dx) \\
&= (\frac{1}{2} - w)[F(\frac{-b}{\sigma_0}) - G(\frac{\mu_0 - b - \mu}{\sigma})]^2 + w[F(\frac{-a}{\sigma_0}) - G(\frac{\mu_0 - a - \mu}{\sigma})]^2 \\
&\quad + w[F(\frac{a}{\sigma_0}) - G(\frac{\mu_0 + a - \mu}{\sigma})]^2 + (\frac{1}{2} - w)[F(\frac{b}{\sigma_0}) - G(\frac{\mu_0 + b - \mu}{\sigma})]^2 \\
&= (\frac{1}{2} - w)N[F(\frac{-b}{\sigma_0}) - G(\frac{-b + (\mu_0 - \mu)}{\sigma})]^2 + w[F(\frac{-a}{\sigma_0}) - G(\frac{-a + (\mu_0 - \mu)}{\sigma})]^2 \\
&\quad w[F(\frac{a}{\sigma_0}) - G(\frac{a + (\mu_0 - \mu)}{\sigma})]^2 + (\frac{1}{2} - w)[F(\frac{b}{\sigma_0}) - G(\frac{b + (\mu_0 - \mu)}{\sigma})]^2 \\
&= (\frac{1}{2} - w)[F(\frac{b}{\sigma_0}) - G(\frac{b - (\mu_0 - \mu)}{\sigma})]^2 + w[F(\frac{a}{\sigma_0}) - G(\frac{a - (\mu_0 - \mu)}{\sigma})]^2 \\
&\quad + w[F(\frac{a}{\sigma_0}) - G(\frac{a + (\mu_0 - \mu)}{\sigma})]^2 + (\frac{1}{2} - w)N[F(\frac{b}{\sigma_0}) - G(\frac{b + (\mu_0 - \mu)}{\sigma})]^2 \\
&= w\{[F(\frac{a}{\sigma_0}) - G(\frac{a + (\mu_0 - \mu)}{\sigma})]^2 + [F(\frac{a}{\sigma_0}) - G(\frac{a - (\mu_0 - \mu)}{\sigma})]^2\} \\
&\quad + (\frac{1}{2} - w)\{[F(\frac{b}{\sigma_0}) - G(\frac{b - (\mu_0 - \mu)}{\sigma})]^2 + [F(\frac{b}{\sigma_0}) - G(\frac{b + (\mu_0 - \mu)}{\sigma})]^2\}.
\end{aligned}$$

First we want to show that μ_0 is a minimizer of the h function with respect to μ, more explicitly, to show that $h(\mu, \sigma) \geq h(\mu_0, \sigma)$, $\forall \sigma > 0$, i.e. we want to prove that

$$[F(\frac{b}{\sigma_0}) - G(\frac{b + (\mu_0 - \mu)}{\sigma})]^2 + [F(\frac{b}{\sigma_0}) - G(\frac{b + (\mu_0 - \mu)}{\sigma})]^2 \geq 2[F(\frac{b}{\sigma_0}) - G(\frac{b}{\sigma})]^2,$$

$\forall\ b > 0$, $\mu \in \mathbf{R}$, $\sigma > 0$, which is equivalent to proving that

$$[F(y) - G(x-c)]^2 + [F(y) - G(x+c)]^2 \geq 2[F(y) - G(x)]^2$$
$$\forall y > 0, x > 0, c > 0, \quad \text{where} \quad y = \frac{b}{\sigma_0}, x = \frac{b}{\sigma} \quad \text{and} \quad c = \frac{\mu_0 - \mu}{\sigma}.$$

$$\begin{aligned}
& [F(y) - G(x-c)]^2 + [F(y) - G(x+c)]^2 \\
= \; & [F(y) - G(x) + G(x) - G(x-c)]^2 + [F(y) - G(x) + G(x) - G(x+c)]^2 \\
= \; & \{[F(y) - G(x)]^2 + [G(x) - G(x-c)]^2 + 2[F(y) - G(x)][G(x) - G(x-c)]\} \\
= \; & +\{[F(y) - G(x)]^2 + [G(x) - G(x+c)]^2 + 2[F(y) - G(x)][G(x) - G(x+c)]\} \\
= \; & [F(y) - G(x)]^2 + 2[F(y) - G(x)][2G(x) - G(x-c) - G(x+c)] \\
= \; & +[G(x) - G(x-c)]^2 + [G(x) - G(x+c)]^2 + [F(y) - G(x)]^2 \\
= \; & \{[F(y) - G(x)]^2 + 2[F(y) - G(x)][2G(x) - G(x-c) - G(x+c)] \\
= \; & +[2G(x) - G(x-c) - G(x+c)]^2\} - 2[G(x) - G(x-c)][G(x) - G(x+c)] + [F(y) - G(x)]^2 \\
= \; & \{[F(y) - G(x)] + [2G(x) - G(x-c) - G(x+c)]\}^2 \\
= \; & +2[G(x) - G(x-c)][G(x+c) - G(x)] + [F(y) - G(x)]^2 \\
= \; & \geq [F(y) - G(x)]^2 + [F(y) - G(x)]^2 = 2[F(y) - G(x)]^2, \forall y \geq 0, x \geq 0, c \geq 0.
\end{aligned}$$

if $[2G(x) - G(x-c) - G(x+c)] \geq 0, \forall x \geq 0, c \geq 0$, which will be established in the following. First, for $x > c \geq 0, [2G(x) - G(x-c) - G(x+c)] \geq 0$,

since $G(x)$ is concave function as $x \geq 0$ and the inequality holds only when c=0.

For $c \geq x > 0$,

$$\begin{aligned}
& 2G(x) - G(x-c) - G(x+c) \\
= \; & [G(x) - G(x-c)] - [G(x+c) - G(x)]
\end{aligned}$$

$$
\begin{aligned}
&= \int_{x-c}^{x} g(t)dt - \int_{x}^{x+c} g(t)dt, \forall c > x \geq 0 \\
&= \int_{x-c}^{0} g(t)dt + \int_{0}^{x} g(t)dt - \int_{x}^{x+c} g(t)dt \\
&= \int_{0}^{c-x} g(t)dt + \int_{0}^{x} g(t)dt - \int_{x}^{x+c} g(t)dt \\
&> \int_{x}^{c} g(t)dt + \int_{0}^{x} g(t)dt - \int_{x}^{x+c} g(t)dt \\
&= \int_{0}^{c} g(t)dt - \int_{x}^{x+c} g(t)dt > 0.
\end{aligned}
$$

The inequalities hold because $g(x)$ is symmetric with respect to x=0 and strictly decreasing as x >0, i.e.

$$\int_{a}^{a+x} g(t)dt > \int_{b}^{b+x} g(t)dt, \forall b > a \geq 0, x > 0.$$

Now, we can say that

$$\min_{\mu} h(\mu,\sigma) = h(\mu_0,\sigma), \forall \sigma > 0, \quad \text{i.e.} \quad \mu_0 = arg \min_{\mu} h(\mu,\sigma).$$

That is to say, μ_0 is the unique solution that $h(\mu,\sigma) = \min_{\mu} h(\mu,\sigma)$.

Second, we show that symmetric two-point design is not T_{LS}-optimum design. It can be proved by the definition of T_{LS}-optimality criterion directly.

$$T_{LS}(\xi) = \min_{\mu,\sigma} \int_{\chi} \|F(\frac{x-\mu_0}{\sigma_0}) - G(\frac{x-\mu}{\sigma})\|^2 \xi(dx)$$

$$= \min_{\mu,\sigma} \{w * [F(\frac{a-\mu_0}{\sigma_0}) - G(\frac{a-\mu}{\sigma})]^2 + (1-w) * [F(\frac{a-\mu_0}{\sigma_0}) - G(\frac{a-\mu}{\sigma})]^2\}$$
$$= 0.$$

If μ and σ are such that $F(\frac{a-\mu_0}{\sigma_0}) - G(\frac{a-\mu}{\sigma})$ and $F(\frac{-a-\mu_0}{\sigma_0}) - G(\frac{-a-\mu}{\sigma})$ where
$\xi = \left\{ \begin{array}{cc} -a & a \\ 1-w & w \end{array} \right\}$,∀ a>0 and $1 > w > 0$.

Similarly, symmetric three-point design, $\xi^* = \left\{ \begin{array}{ccc} -a & 0 & a \\ w & 1-2w & w \end{array} \right\}$,$\forall$ a>0 and $1 > w > 0$, is not a T_{LS}-optimum design either.

Proof of Lemma 2.3 For convenience, we define the function of σ, $D(\sigma)$=$g(\frac{a}{\sigma^*})g(\frac{b}{\sigma}) - g(\frac{a}{\sigma})g(\frac{b}{\sigma^*})$ and we now prove that the sign of $\frac{\partial}{\partial\sigma}h(\mu_0, \sigma, \xi^*)$ is the same as $D(\sigma)$.

$$
\begin{aligned}
& \frac{\partial}{\partial\sigma}h(\mu_0, \sigma, \xi^*) \\
= \; & 4w[F(\frac{a}{\sigma_0}) - G(\frac{a}{\sigma})]g(\frac{a}{\sigma})\frac{a}{\sigma^2} + 4(\frac{1}{2} - w)[F(\frac{b}{\sigma_0}) - G(\frac{b}{\sigma})]g(\frac{b}{\sigma})\frac{b}{\sigma^2} \\
= \; & \frac{4}{\sigma^2}\{aw[F(\frac{a}{\sigma_0}) - G(\frac{a}{\sigma})]g(\frac{a}{\sigma}) + b(\frac{1}{2} - w)[F(\frac{b}{\sigma_0}) - G(\frac{b}{\sigma})]g(\frac{b}{\sigma})\}.
\end{aligned}
$$

Replacing w by $\dfrac{b \cdot g(\frac{b}{\sigma^*})}{2[b \cdot g(\frac{b}{\sigma^*}) + a \cdot g(\frac{a}{\sigma^*})]}$ yields

$$
\begin{aligned}
& \frac{2}{\sigma^2}\{\frac{a \cdot g(\frac{a}{\sigma^*}) \cdot b \cdot [F(\frac{b}{\sigma_0}) - G(\frac{b}{\sigma})]g(\frac{b}{\sigma})}{[b \cdot g(\frac{b}{\sigma^*}) + a \cdot g(\frac{a}{\sigma^*})]} + \frac{b \cdot g(\frac{b}{\sigma^*}) \cdot a \cdot [F(\frac{a}{\sigma_0}) - G(\frac{a}{\sigma})]g(\frac{a}{\sigma})}{[b \cdot g(\frac{b}{\sigma^*}) + a \cdot g(\frac{a}{\sigma^*})]}\} \\
= \; & \frac{2}{\sigma^2} \cdot \frac{ab}{[b \cdot g(\frac{b}{\sigma^*}) + a \cdot g(\frac{a}{\sigma^*})]}\{g(\frac{a}{\sigma^*})[F(\frac{b}{\sigma_0}) - G(\frac{b}{\sigma})]g(\frac{b}{\sigma}) + g(\frac{b}{\sigma^*})[F(\frac{a}{\sigma_0}) - G(\frac{a}{\sigma})]g(\frac{a}{\sigma})\} \\
= \; & \frac{2}{\sigma^2} \cdot \frac{ab}{[b \cdot g(\frac{b}{\sigma^*}) + a \cdot g(\frac{a}{\sigma^*})]}\{g(\frac{a}{\sigma^*})g(\frac{b}{\sigma})[F(\frac{b}{\sigma_0}) - G(\frac{b}{\sigma})] + g(\frac{b}{\sigma^*})g(\frac{a}{\sigma})[F(\frac{a}{\sigma_0}) - G(\frac{a}{\sigma})]\}.
\end{aligned}
$$

Since $\frac{2}{\sigma^2} \cdot \frac{ab}{[b \cdot g(\frac{b}{\sigma^*}) + a \cdot g(\frac{a}{\sigma^*})]} > 0$, the sign of $\frac{\partial}{\partial\sigma}h(\mu_0, \sigma, \xi^*)$ is the same as $g(\frac{b}{\sigma^*})g(\frac{a}{\sigma})[F(\frac{a}{\sigma_0}) - G(\frac{a}{\sigma})] + g(\frac{a}{\sigma^*})g(\frac{b}{\sigma})[F(\frac{b}{\sigma_0}) - G(\frac{b}{\sigma})]$.

Consider the function

$$\begin{aligned}
& g(\frac{a}{\sigma})g(\frac{b}{\sigma^*})[F(\frac{a}{\sigma_0}) - G(\frac{a}{\sigma})] + g(\frac{a}{\sigma^*})g(\frac{b}{\sigma})[F(\frac{b}{\sigma_0}) - G(\frac{b}{\sigma})] \\
= \ & g(\frac{a}{\sigma})g(\frac{b}{\sigma^*})[F(\frac{a}{\sigma_0}) - G(\frac{a}{\sigma^*}) + G(\frac{a}{\sigma^*}) - G(\frac{a}{\sigma})] \\
& + g(\frac{a}{\sigma^*})g(\frac{b}{\sigma})[F(\frac{b}{\sigma_0}) - G(\frac{b}{\sigma^*}) + G(\frac{b}{\sigma^*}) - G(\frac{b}{\sigma})] \\
= \ & [F(\frac{b}{\sigma_0}) - G(\frac{b}{\sigma^*})][g(\frac{a}{\sigma^*})g(\frac{b}{\sigma}) - g(\frac{a}{\sigma})g(\frac{b}{\sigma^*})] \\
& + g(\frac{a}{\sigma})g(\frac{b}{\sigma^*})[G(\frac{a}{\sigma^*}) - G(\frac{a}{\sigma})] + g(\frac{a}{\sigma^*})g(\frac{b}{\sigma})[G(\frac{b}{\sigma^*}) - G(\frac{b}{\sigma})].
\end{aligned}$$

Since $[F(\frac{b}{\sigma_0}) - G(\frac{b}{\sigma^*})] > 0$, $g(\cdot) > 0$ and $[G(\frac{x}{\sigma^*}) - G(\frac{x}{\sigma})] \geq 0$ for $x \geq 0$ and $\sigma \geq \sigma^*$, then

$$\frac{\partial}{\partial\sigma} h(\mu_0, \sigma, \xi^*) \geq 0 \quad \text{for} \quad \sigma \geq \sigma^* \quad \text{if} \quad g(\frac{a}{\sigma^*})g(\frac{b}{\sigma}) - g(\frac{a}{\sigma})g(\frac{b}{\sigma^*}) \geq 0 \quad \text{for} \quad \sigma \geq \sigma^*,$$

i.e. The sign of $\frac{\partial}{\partial\sigma} h(\mu_0, \sigma, \xi^*)$ is the same as $g(\frac{a}{\sigma^*})g(\frac{b}{\sigma}) - g(\frac{a}{\sigma})g(\frac{b}{\sigma^*})$.

Lemma A.1. *Let $\{\hat{\mu}_N, \hat{\sigma}_N\}$ be the least squared estimator of $\{\mu, \sigma\}$ and N is the number of total observations, then*

$$\lim_{N\to\infty} \hat{\mu}_N = \mu_0, \quad \textit{and} \quad \lim_{N\to\infty} \hat{\sigma}_N = \sigma^*.$$

We have shown that $\lim_{N\to\infty} \hat{\mu}_N = \mu_0$ in the proof of Lemma 2.2, and we will show that $\lim_{N\to\infty} \hat{\sigma}_N = \sigma^*$ now. Recall the h function,

$$\begin{aligned}
h(\mu, \sigma) = & \tfrac{N_1}{N}\{[F(\tfrac{a}{\sigma_0}) - G(\tfrac{a+(\mu_0-\mu)}{\sigma})]^2 + [F(\tfrac{a}{\sigma_0}) - G(\tfrac{a-(\mu_0-\mu)}{\sigma})]^2\} \\
& + \tfrac{N_2}{N}\{[F(\tfrac{b}{\sigma_0}) - G(\tfrac{b-(\mu_0-\mu)}{\sigma})]^2 + [F(\tfrac{b}{\sigma_0}) - G(\tfrac{b+(\mu_0-\mu)}{\sigma})]^2\}.
\end{aligned}$$

Since we have shown that $\hat{\mu} = \mu_0$, then

$$h(\mu_0, \sigma, \xi^*) = 2\tfrac{N_1}{N}[F(\tfrac{a}{\sigma_0}) - G(\tfrac{a}{\sigma})]^2 + 2\tfrac{N_2}{N}[F(\tfrac{a}{\sigma_0}) - G(\tfrac{a}{\sigma})]^2,$$

and the LSE of σ, $\hat{\sigma}$, must satisfy the equation

$$\frac{\partial}{\partial\sigma}h(\mu_0,\sigma,\xi^*)|_{\sigma=\hat{\sigma}} = 0$$
$$4\frac{N_1}{N}[F(\frac{a}{\sigma_0}) - G(\frac{a}{\sigma})]g(\frac{a}{\sigma})\frac{a}{\sigma^2} + 4\frac{N_2}{N}[F(\frac{b}{\sigma_0}) - G(\frac{b}{\sigma})]g(\frac{b}{\sigma})\frac{b}{\sigma^2}|_{\sigma=\hat{\sigma}} = 0$$
$$\frac{4}{\sigma^2}\{a\frac{N_1}{N}[F(\frac{a}{\sigma_0}) - G(\frac{a}{\sigma})]g(\frac{a}{\sigma}) + b\frac{N_2}{N}[F(\frac{b}{\sigma_0}) - G(\frac{b}{\sigma})]g(\frac{b}{\sigma})\}|_{\sigma=\hat{\sigma}} = 0.$$

Replacing $\frac{N_1}{N}$ by $w = \frac{b\cdot g(\frac{b}{\sigma^*})}{2[b\cdot g(\frac{b}{\sigma^*}) + a\cdot g(\frac{a}{\sigma^*})]}$ and $\frac{N_2}{N} = \frac{1}{2} - \frac{N_1}{N}$,

then the final equality becomes

$$\{\frac{a\cdot g(\frac{a}{\sigma^*})\cdot b\cdot[F(\frac{b}{\sigma_0}) - G(\frac{b}{\sigma})]g(\frac{b}{\sigma})}{[b\cdot g(\frac{b}{\sigma^*}) + a\cdot g(\frac{a}{\sigma^*})]} + \frac{b\cdot g(\frac{b}{\sigma^*})\cdot a\cdot[F(\frac{a}{\sigma_0}) - G(\frac{a}{\sigma})]g(\frac{a}{\sigma})}{[b\cdot g(\frac{b}{\sigma^*}) + a\cdot g(\frac{a}{\sigma^*})]}\}|_{\sigma=\hat{\sigma}} = 0$$
$$\{g(\frac{a}{\sigma^*})[F(\frac{b}{\sigma_0}) - G(\frac{b}{\sigma})]g(\frac{b}{\sigma}) + g(\frac{b}{\sigma^*})[F(\frac{a}{\sigma_0}) - G(\frac{a}{\sigma})]g(\frac{a}{\sigma})\}|_{\sigma=\hat{\sigma}} = 0$$
$$\{g(\frac{a}{\sigma^*})g(\frac{b}{\hat{\sigma}})[F(\frac{b}{\sigma_0}) - G(\frac{b}{\hat{\sigma}})] + g(\frac{b}{\sigma^*})g(\frac{a}{\hat{\sigma}})[F(\frac{a}{\sigma_0}) - G(\frac{a}{\hat{\sigma}})]\} = 0 \qquad (\star)$$

Since $[F(\frac{a}{\sigma_0}) - G(\frac{a}{\sigma^*})] = -[F(\frac{b}{\sigma_0}) - G(\frac{b}{\sigma^*})]$, and by equation $(\star)$ we know that σ^* is a LSE of σ.

Lemma A.2. *σ^* is the only LSE of σ when the assumed model is with the double-exponential or double-reciprocal distributions, and the true model is from one of the other three models we introduced in the beginning.*

By Lemma 2.3, to prove that σ^* is the only LSE of σ, we have only to check that σ^* is the only solution of the function $D(\sigma)$=0, that is, $g(\frac{a}{\sigma^*})g(\frac{b}{\sigma}) - g(\frac{a}{\sigma})g(\frac{b}{\sigma^*}) = 0$. We first prove that $D(\sigma^*)$=0 when the probability function $g(\cdot)$ of the assumed model is with double-exponential distribution, that is, $g(x;\mu,\sigma) = \frac{1}{2}e^{-|\frac{x-\mu}{\sigma}|}$, and $a > 0$, $b > 0$ and

$\sigma^* > 0$ are defined as in (2.1)-(2.4), then

$$\begin{aligned} D(\sigma_1) &= g(\frac{a}{\sigma^*})g(\frac{b}{\sigma}) - g(\frac{a}{\sigma})g(\frac{b}{\sigma^*}) \\ &= \frac{1}{2}exp\{-|\frac{a}{\sigma^*}| - |\frac{b}{\sigma}|\} - \frac{1}{2}exp\{-|\frac{a}{\sigma}| - |\frac{b}{\sigma^*}|\} \\ &= \frac{1}{2\pi}exp\{-\frac{a}{\sigma^*} - \frac{a}{\sigma}\}[exp\{-\frac{b-a}{\sigma}\} - exp\{-\frac{b-a}{\sigma^*}\}], \end{aligned}$$

since $exp\{\cdot\} > 0$ and $exp\{\frac{x}{\sigma}\} - exp\{\frac{x}{\sigma^*}\} = 0$ only as $\sigma = \sigma^*$, then it implies σ^* is the only LSE of σ.

We now prove that $D(\sigma^*)$=0 when the probability function $g(\cdot)$ of the assumed model is with double-reciprocal distribution, that is, $g(x;\mu,\sigma) = \frac{1}{2(1+|\frac{x-\mu}{\sigma}|)^2}$, and $a > 0$, $b > 0$ and $\sigma^* > 0$ are defined as before, then

$$\begin{aligned} D(\sigma) &= g(\frac{a}{\sigma^*})g(\frac{b}{\sigma}) - g(\frac{a}{\sigma})g(\frac{b}{\sigma^*}) \\ &= \frac{1}{2(1+|\frac{a}{\sigma^*}|)^2} \cdot \frac{1}{2(1+|\frac{b}{\sigma}|)^2} - \frac{1}{2(1+|\frac{a}{\sigma}|)^2} \cdot \frac{1}{2(1+|\frac{b}{\sigma^*}|)^2} \\ &= \frac{1}{4(1+\frac{a}{\sigma^*})^2(1+\frac{a}{\sigma})^2}[\left(\frac{1+\frac{a}{\sigma}}{1+\frac{b}{\sigma}}\right)^2 - \left(\frac{1+\frac{a}{\sigma^*}}{1+\frac{b}{\sigma^*}}\right)^2] \\ &= \frac{1}{4(1+\frac{a}{\sigma^*})^2(1+\frac{a}{\sigma})^2}[\left(\frac{\sigma+a}{\sigma+b}\right)^2 - \left(\frac{\sigma^*+a}{\sigma^*+b}\right)^2], \end{aligned}$$

since

$$[\left(\frac{\sigma+a}{\sigma+b}\right)^2 - \left(\frac{\sigma^*+a}{\sigma^*+b}\right)^2] = 0 \quad \text{only as} \quad \sigma = \sigma^*,$$

σ^* is the only LSE of σ.