

SFB
823

Optimal discrimination designs for semi-parametric models

Holger Dette, Roman Guchenko,
Viatcheslav Melas, Weng Kee Wong

Nr. 59/2016

Discussion Paper



Optimal discrimination designs for semi-parametric models

Holger Dette

Ruhr-Universität Bochum

Fakultät für Mathematik

44780 Bochum, Germany

e-mail: holger.dette@ruhr-uni-bochum.de

Roman Guchenko, Viatcheslav Melas

St. Petersburg State University

Faculty of Mathematics and Mechanics

St. Petersburg, Russia

e-mail: romanguchenko@ya.ru, vbmelas@post.ru

Weng Kee Wong

Department of Biostatistics

UCLA Fielding School of Public Health

Los Angeles, CA 90095-1772

e-mail: wkwong@ucla.edu

October 28, 2016

Abstract

Much of the work in the literature on optimal discrimination designs assumes that the models of interest are fully specified, apart from unknown parameters in some models. Recent work allows errors in the models to be non-normally distributed but still requires the specification of the mean structures. This research is motivated by the interesting work of Otsu (2008) to discriminate among semi-parametric models by generalizing the KL-optimality criterion proposed by López-Fidalgo et al. (2007) and Tommasi and López-Fidalgo (2010). In our work we provide further important insights in this interesting optimality criterion. In particular, we propose a practical strategy for finding optimal discrimination designs among semi-parametric models that can also be verified using an equivalence theorem. In addition, we study properties of such optimal designs and identify important cases where the proposed semi-parametric optimal discrimination designs coincide with the celebrated T -optimal designs.

Keywords and Phrases: continuous design, discrimination design, equivalence theorem, semi-parametric model, T -optimality, variational calculus.

1 Introduction

Finding optimal discrimination designs dates back to 1970's and probably earlier. Early references are Stigler (1971) who considered nested models and Atkinson and Fedorov (1975a,b), where they proposed T -optimal designs to discriminate among models when errors are normally distributed. The first of the two last named papers concerns discrimination between two models and the second paper generalized the problem to multiple models. T -optimality for discriminating between parametric models assumes that we have a known null model and we wish to test whether a rival model with unknown parameter holds. A likelihood ratio test is then used to discriminate between the two rival models and under local alternatives, the non-centrality parameter of the chi-square distribution of the test statistic, which contains the unknown parameters from the alternative model, is proportional to the T -optimality criterion (Wiens, 2009). Since a larger non-centrality parameter provides a more powerful test, the T -optimal design strategy is to find a design that maximizes the minimum value of the non-centrality parameter, where the minimum is taken over all possible values of the parameters in the model under the alternative hypothesis. Consequently, T -optimality criterion is a type of maximin optimality criterion and designs that maximize the minimum of the non-centrality parameter are called T -optimal designs.

The two seminal papers of Atkinson and Fedorov (1975a,b) contain illustrative examples. Because the T -optimality criterion is not differentiable, calculations to determine the optimal design can be involved even for relatively simple models; see for example, Dette et al. (2012) and Dette et al. (2016b) for some explicit non-trivial results. For the same reason, the construction of efficient algorithms for finding T -optimal designs is a challenging problem. Some recent progress in this direction includes Braess and Dette (2013); Dette et al. (2015, 2016a) and Tommasi et al. (2016). An interactive software tool for finding optimal discrimination designs is available in Stegmaier et al. (2013).

There is notable progress in tackling optimal discrimination design problems on various fronts; the progress appeared periodic a few decades ago and seems to have picked up in recent years. Advances include alternative problem formulations and broader applications of such designs in cognitive science (Covagnaro et al., 2010), psychology (Myung and Pitt, 2009) and chemical engineering (Alberton et al., 2011), to name a few. Different and more flexible optimality criteria have also been proposed. In particular, advances in the construction of optimal discrimination designs include the following: (i) the frequently criticized unrealistic assumption in the T -optimality criterion that requires a known model in the null hypothesis is now removed (Jamsen et al., 2013); (ii) the class of models of interest now includes generalized linear models, where outcomes can be discrete (Waterhouse et al., 2008), (iii) optimal designs for discriminating multivariate models (Yanagisawa, 1990; Ucinski and Bogacka, 2005); (iv) Bayesian optimal designs for model discrimination (Felsenstein, 1992; Tommasi and López-Fidalgo, 2010; Dette et al., 2015), (v) dual-objective optimal designs for model discrimination (Atkinson et al., 1998; Ng and Chick, 2004; Atkinson, 2008; Alberton et al., 2011; Abd El-Monsef and Seyam, 2011), (vi) discriminating models with correlated errors (Campos-Barreiro and Lopez-Fidalgo, 2016), (vii) adaptive designs for model discrimination (Myung and Pitt,

2009; Donckels et al., 2009) and (viii) optimal discrimination designs for dynamic models (Ucinski and Bogacka, 2005). Other important references that describe alternative approaches and properties of optimal discrimination designs are López-Fidalgo et al. (2007); Dette and Titoff (2009); Dette et al. (2015), among others.

All references cited so far require a parametric specification of the conditional distribution of the response, which raises some robustness issues in the application of the T -optimality criterion for constructing optimal discrimination designs. Robustness properties of optimal discrimination designs with respect to various model assumptions have been considered by Wiens (2009); Ghosh and Dutta (2013) and Dette et al. (2013). Otsu (2008) proposed a new optimality criterion for discriminating between models, which is similar in spirit to the classical T -optimality criterion and its extensions except that it does not require an exact specification of the conditional distribution. Optimal discrimination designs were found using the duality relationships in entropy-like minimization problems (Borwein and Lewis, 1991) and the resulting optimal designs are called semi-parametric optimal discrimination designs.

The present paper provides a more careful analysis of the novel approach proposed by Otsu (2008). We propose a more practical strategy for finding semi-parametric optimal discrimination designs and derive several important properties of such optimal designs. In Section 2, we define semi-parametric optimal discrimination designs and derive several auxiliary results that substantially simplify the calculation of optimal designs. In Section 3, we provide equivalence theorems to characterize semi-parametric optimal discrimination designs and demonstrate that all designs derived by Otsu (2008) are in fact not optimal. Section 4 describes the relation between semi-parametric optimal discrimination designs and optimal discrimination designs under some criteria discussed in the literature. In particular, we identify cases, where the semi-parametric optimal discrimination designs proposed by Otsu (2008) coincide with the T -optimal discrimination designs and KL-optimal designs introduced by Atkinson and Fedorov (1975a,b) and López-Fidalgo et al. (2007), respectively. Section 5 presents numerical results and all technical details are deferred to the appendix in Section 6.

2 Semi-parametric discrimination designs

We model the response variable Y as a function of a vector of explanatory variables x defined on a given compact design space \mathcal{X} . Suppose the density of Y with respect to the Lebesgue measure is $f(y, x)$, and we have resources to take a fixed number of observations, say n , for the study. Following Kiefer (1974), we focus on approximate designs which are essentially probability measures defined on \mathcal{X} . If an approximate design has k support points, say x_1, \dots, x_k , and the corresponding weights are $\omega_1, \dots, \omega_k$, then approximately $n\omega_i$ observations are taken at $x_i, i = 1, \dots, k$. In practice, a rounding procedure is applied to every $n\omega_i$ so that they are positive integers n_i ($i = 1, \dots, k$) and $\sum_{i=1}^k n_i = n$. The experimenter then takes n_i independent observations with $Y_{i,1}, \dots, Y_{i,n_i}$ at x_i which are assumed to have a density $f(y, x_i)$ with respect to the Lebesgue measure, $i = 1, \dots, k$.

To construct efficient designs for discriminating between two competing models for $f(y, x)$,

López-Fidalgo et al. (2007) assumed parametric densities, say $f_j(y, x, \theta_j)$, where the parameter θ_j varies in a compact parameter space, say Θ_j , $j = 1, 2$. To fix ideas, we ignore nuisance parameters which may be present in the models. The Kullback-Leibler distance between two densities f_1 and f_2 is

$$I_{1,2}(x, f_1, f_2, \theta_1, \theta_2) = \int f_1(y, x, \theta_1) \log \frac{f_1(y, x, \theta_1)}{f_2(y, x, \theta_2)} dy \quad (2.1)$$

and it measures the discrepancy between the densities. López-Fidalgo et al. (2007) assumed that the model f_1 is the “true” model with a fixed parameter vector $\bar{\theta}_1$ and call a design local *KL-optimal discriminating design for the models f_1 and f_2* if it maximizes the criterion

$$\text{KL}_{1,2}(\xi, \bar{\theta}_1) = \inf_{\theta_2 \in \Theta_2} \int_{\mathcal{X}} I_{1,2}(x, f_1, f_2, \bar{\theta}_1, \theta_2) \xi(dx). \quad (2.2)$$

Otsu (2008) considered a novel setup and proposed a design criterion for discriminating a parametric model defined by its density and another nonparametric model. More precisely, this author considered the design problem for testing the hypothesis

$$H_0 : \eta(x) = \eta_1(x, \theta_1) \quad \text{versus} \quad H_1 : \eta(x) = \eta_2(x, \theta_2) \quad (2.3)$$

where

$$\eta_j(x, \theta_j) = \int y f_j(y, x, \theta_j) dy, \quad j = 1, 2,$$

is the conditional mean of the density $f_j(y, x, \theta_j)$ with support set

$$\mathcal{S}_{f_j, \theta_j, x} = \{y \mid f_j(y, x, \theta_j) > 0\}, \quad j = 1, 2. \quad (2.4)$$

The setup is therefore more general than that in López-Fidalgo et al. (2007), who assumed that f_1 and f_2 are known and one of the parametric models is fully specified. To be more specific, let $f_1(y, x, \bar{\theta}_1)$ be a parametric density with a fixed parameter $\bar{\theta}_1$ and define

$$\mathcal{F}_{2,x,\theta_2} = \left\{ f_2 : \int f_2(y, x, \theta_2) dy = 1, \int y f_2(y, x, \theta_2) dy = \eta_2(x, \theta_2), \mathcal{S}_{f_2, \theta_2, x} = \mathcal{S}_{f_1, \bar{\theta}_1, x} \right\}, \quad (2.5)$$

which is the class of all conditional densities (at the point x) with parameter θ_2 and conditional mean $\eta_2(x, \theta_2)$. Consider the set obtained from $\mathcal{F}_{2,x,\theta_2}$ by letting the ranges of x and θ_2 vary over all their possible values:

$$\mathcal{F}_2 = \{ \mathcal{F}_{2,x,\theta_2} \mid x \in \mathcal{X}; \theta_2 \in \Theta_2 \}$$

and call a design ξ^* *semi-parametric optimal design for discriminating between the model*

$f_1(y, x, \bar{\theta}_1)$ and models in the class \mathcal{F}_2 if it maximizes the optimality criterion

$$K_{(a)}(\xi, \bar{\theta}_1) = \inf_{\theta_2 \in \Theta_2} \int_X \inf_{f_2 \in \mathcal{F}_{2,x,\theta_2}} I_{1,2}(x, f_1, f_2, \bar{\theta}_1, \theta_2) \xi(dx) \quad (2.6)$$

among all approximate designs defined on the design space \mathcal{X} . Note that this criterion is a local optimality criterion in the sense of Chernoff (1953) as it depends on the parameter $\bar{\theta}_1$. The subscript (a) denotes the first design criterion with the subscript (b) for the second design criterion below.

Similarly, we may fix the conditional density $f_2(y, x, \theta_2)$ and define

$$\mathcal{F}_{1,x,\bar{\theta}_1} = \left\{ f_1 : \int f_1(y, x, \bar{\theta}_1) dy = 1, \int y f_1(y, x, \bar{\theta}_1) dy = \eta_1(x, \bar{\theta}_1), \mathcal{S}_{f_1, \bar{\theta}_1, x} = \mathcal{S}_{f_2, \theta_2, x} \right\}, \quad (2.7)$$

which is the class of all conditional densities with parameter $\bar{\theta}_1$ and conditional mean $\eta_1(x, \bar{\theta}_1)$. For fixed $\bar{\theta}_1$, let

$$\mathcal{F}_1 = \{ \mathcal{F}_{1,x,\bar{\theta}_1} \mid x \in \mathcal{X} \}$$

and we call a design ξ locally *semi-parametric optimal design for discriminating between the model $f_2(y, x, \theta_2)$ and the class \mathcal{F}_1* if it maximizes the optimality criterion

$$K_{(b)}(\xi, \bar{\theta}_1) = \inf_{\theta_2 \in \Theta_2} \int_X \inf_{f_1 \in \mathcal{F}_{1,x,\bar{\theta}_1}} I_{1,2}(x, f_1, f_2, \bar{\theta}_1, \theta_2) \xi(dx), \quad (2.8)$$

among all approximate designs defined on the design space \mathcal{X} .

Remark 2.1 We note that the above approach can be similarly applied when the response variable Y is discrete and it has one of two possible discrete probability measures that depends on x and $\bar{\theta}_1$. Let these two competing measures with the same support be

$$P(x, \bar{\theta}_1) = \begin{bmatrix} y_1(x, \bar{\theta}_1) & \dots & y_s(x, \bar{\theta}_1) \\ p_1(x, \bar{\theta}_1) & \dots & p_s(x, \bar{\theta}_1) \end{bmatrix} \quad \text{and} \quad Q(x, \theta_2) = \begin{bmatrix} y_1(x, \bar{\theta}_1) & \dots & y_s(x, \bar{\theta}_1) \\ q_1(x, \theta_2) & \dots & q_s(x, \theta_2) \end{bmatrix}.$$

If we let $y_i = y_i(x, \bar{\theta}_1)$, let $p_i = p_i(x, \bar{\theta}_1)$ and let $q_i = q_i(x, \theta_2)$ for simplicity, we have $\sum_{i=1}^s p_i = 1 = \sum_{i=1}^s q_i$ and $\sum_{i=1}^s y_i p_i = \eta_1(x, \bar{\theta}_1)$. Define two sets

$$Q_{x,\theta_2} = \left\{ q : \sum_{i=1}^s q_i = 1, \sum_{i=1}^s y_i q_i = \eta_2, 0 < q_i \leq 1 \right\} \quad \text{and} \quad Q = \left\{ Q_{x,\theta_2} \mid x \in \mathcal{X}; \theta_2 \in \Theta_2 \right\}.$$

To find a semi-parametric optimal design ξ^* for discriminating between the model $P(x, \bar{\theta}_1)$ and the class Q we may use the discrete version of the criterion (2.6), which is

$$\tilde{K}_{(a)}(\xi, \bar{\theta}_1) = \inf_{\theta_2 \in \Theta_2} \int_X \inf_{q \in Q_{x,\theta_2}} \sum_{i=1}^s p_i(x, \bar{\theta}_1) \log \frac{p_i(x, \bar{\theta}_1)}{q_i(x, \theta_2)} \xi(dx).$$

We then write down the discrete version of (2.8) in a similar manner and proceed, omitting details for space consideration.

In what is to follow, we assume for simplicity that $f_1(y, x, \theta_1)$, $f_2(y, x, \theta_2)$, $\eta_1(x, \theta_1)$, $\eta_2(x, \theta_2)$ are differentiable with respect to y , x , θ_1 and θ_2 , even though these assumptions can be weakened a bit for what we plan to do here. Otsu (2008) derived an explicit form for the two criteria. For the criterion (2.6), he obtained

$$K_{(a)}(\xi, \bar{\theta}_1) = \inf_{\theta_2 \in \Theta_2} \int_{\mathcal{X}} \left\{ \mu + 1 + \int \log \{-\mu - \lambda(y - \eta_2(x, \theta_2))\} f_1(y, x, \bar{\theta}_1) dy \right\} \xi(dx),$$

where the constants λ and μ depend on x , $\bar{\theta}_1$ and θ_2 and are roots of the system of equations

$$- \int \frac{f_1(y, x, \bar{\theta}_1)}{\mu + \lambda(y - \eta_2(x, \theta_2))} dy = 1, \quad \int \frac{(y - \eta_2(x, \theta_2)) f_1(y, x, \bar{\theta}_1)}{\mu + \lambda(y - \eta_2(x, \theta_2))} dy = 0$$

that satisfy the constraint

$$\mu + \lambda(y - \eta_2(x, \theta_2)) < 0, \quad \text{for all } y \in \mathcal{S}_{f_1, \bar{\theta}_1, x}.$$

A similar result can be obtained for the criterion (2.8) using somewhat similar arguments and they are omitted for space consideration. In either case, the desired values for μ and λ have to be found numerically. Otsu (2008) did not provide an effective numerical procedure for finding solutions in the two dimensional space that solve the system of equations and satisfy the constraint. In what is to follow, we show that the inner optimization problems in (2.6) and (2.8) can be reduced to solving a single equation, which is much faster and more reliable. For this purpose we derive simpler expressions for the criteria (2.6) and (2.8) that facilitate the computation of the semi-parametric optimal discriminating designs. One of our key results is the following.

Theorem 2.1

(a) Assume that for each $x \in \mathcal{X}$ the support of the conditional density $f_1(y, x, \bar{\theta}_1)$ is an interval, i.e. $\mathcal{S}_{f_1, \bar{\theta}_1, x} = [y_{x, \min}, y_{x, \max}]$, such that $y_{x, \min} < \eta_2(x, \theta_2) < y_{x, \max}$ for all $\theta_2 \in \Theta_2$. Assume further that for all $x \in \mathcal{X}$ and for all $\theta_2 \in \Theta_2$ there exists a unique non-zero solution $\bar{\lambda}(x, \bar{\theta}_1, \theta_2)$ of the equation

$$\int \frac{f_1(y, x, \bar{\theta}_1)}{1 + \lambda(y - \eta_2(x, \theta_2))} dy = 1, \quad (2.9)$$

that satisfies

$$-\frac{1}{y_{x, \max} - \eta_2(x, \theta_2)} < \bar{\lambda}(x, \bar{\theta}_1, \theta_2) < -\frac{1}{y_{x, \min} - \eta_2(x, \theta_2)}. \quad (2.10)$$

The criterion (2.6) then takes the form

$$K_{(a)}(\xi, \bar{\theta}_1) = \inf_{\theta_2 \in \Theta_2} \int_{\mathcal{X}} \int \log \{1 + \bar{\lambda}(x, \bar{\theta}_1, \theta_2)(y - \eta_2(x, \theta_2))\} f_1(y, x, \bar{\theta}_1) dy \xi(dx) \quad (2.11)$$

and the “optimal” density f_2^* in (2.6) is given by

$$f_2^*(y, x, \theta_2) = \frac{f_1(y, x, \bar{\theta}_1)}{1 + \bar{\lambda}(x, \bar{\theta}_1, \theta_2)(y - \eta_2(x, \theta_2))}. \quad (2.12)$$

(b) Assume that the integrals

$$\int f_2(y, x, \theta_2) \exp(-\lambda y) dy \quad \text{and} \quad \int y f_2(y, x, \theta_2) \exp(-\lambda y) dy$$

exist for all $x \in \mathcal{X}$ and for all λ . The criterion (2.8) takes the form

$$K_{(b)}(\xi) = \inf_{\theta_2 \in \Theta_2} \int_{\mathcal{X}} \int [\log \{\mu'(\bar{\lambda}_x)\} - \bar{\lambda}_x y] f_2(y, x, \theta_2) \exp(-\bar{\lambda}_x y) \mu'(\bar{\lambda}_x) dy \xi(dx) \quad (2.13)$$

where

$$\mu'(\lambda) = \mu'(\lambda, x, \theta_2) = \frac{1}{\int f_2(y, x, \theta_2) \exp(-\lambda y) dy} \quad (2.14)$$

and $\bar{\lambda}_x = \bar{\lambda}(x, \bar{\theta}_1, \theta_2)$ is the nonzero root of the equation

$$\frac{\int y f_2(y, x, \theta_2) \exp(-\lambda y) dy}{\int f_2(y, x, \theta_2) \exp(-\lambda y) dy} = \eta_1(x, \bar{\theta}_1). \quad (2.15)$$

Moreover, the “optimal” density f_1^* in (2.8) is given by

$$f_1^*(y, x, \bar{\theta}_1) = \frac{f_2(y, x, \theta_2) \exp(-\bar{\lambda}(x, \bar{\theta}_1, \theta_2)y)}{\int f_2(y, x, \theta_2) \exp(-\bar{\lambda}(x, \bar{\theta}_1, \theta_2)y) dy}. \quad (2.16)$$

The main implication of Theorem 2.1 is that we first solve equations (2.9) and (2.15) numerically for λ . For the solution of (2.9), it is natural to presume that $\lambda < 0$ if $\eta_1(x, \bar{\theta}_1) < \eta_2(x, \theta_2)$. This is because, in this case, the function $1/[1 + \lambda(y - \eta_2(x, \theta_2))]$ is increasing whenever $y \in \mathcal{S}_{f_1, \bar{\theta}_1, x}$, allowing us to shift the average of the function $f_1(y, x, \bar{\theta}_1)/[1 + \lambda(y - \eta_2(x, \theta_2))]$ to the right. Similarly, if $\eta_1(x, \bar{\theta}_1) > \eta_2(x, \theta_2)$, we search for $\lambda > 0$. We state this intuitive consideration formally in the following lemma, whose proof is deferred to the appendix.

Lemma 2.1 *Assume that $v_2^2(x, \theta_2) = \int (y - \eta_2(x, \theta_2))^2 f_2(y, x, \theta_2) dy$ exists and is positive. If $\bar{\lambda}$ is a solution of the equation (2.9) and satisfies (2.10), then $\bar{\lambda}$ has the same sign as the difference $\eta_1(x, \theta_1) - \eta_2(x, \theta_2)$.*

Example 2.1 Let $f_1(y, x, \bar{\theta}_1)$ be the truncated normal density $\mathcal{N}(\eta(x, \bar{\theta}_1), 1)$ on the interval $[-3 + \eta_1(x, \bar{\theta}_1), 3 + \eta_1(x, \bar{\theta}_1)]$. This density is a function of $\eta_1(x, \bar{\theta}_1)$ and it follows from (2.12) that the optimal density $f_2^*(y, x, \theta_2)$ is a function of $\eta_1(x, \bar{\theta}_1)$ and $\eta_2(x, \theta_2)$ in this case. Figure 2 shows plots of the function f_2^* for $\eta_1(x, \bar{\theta}_1) \equiv 0$ and different values of $\eta_2(x, \theta_2)$ on the interval $[-3, 3]$. On the top of each figure the value $\bar{\lambda}$ shows the solution of equation (2.9).

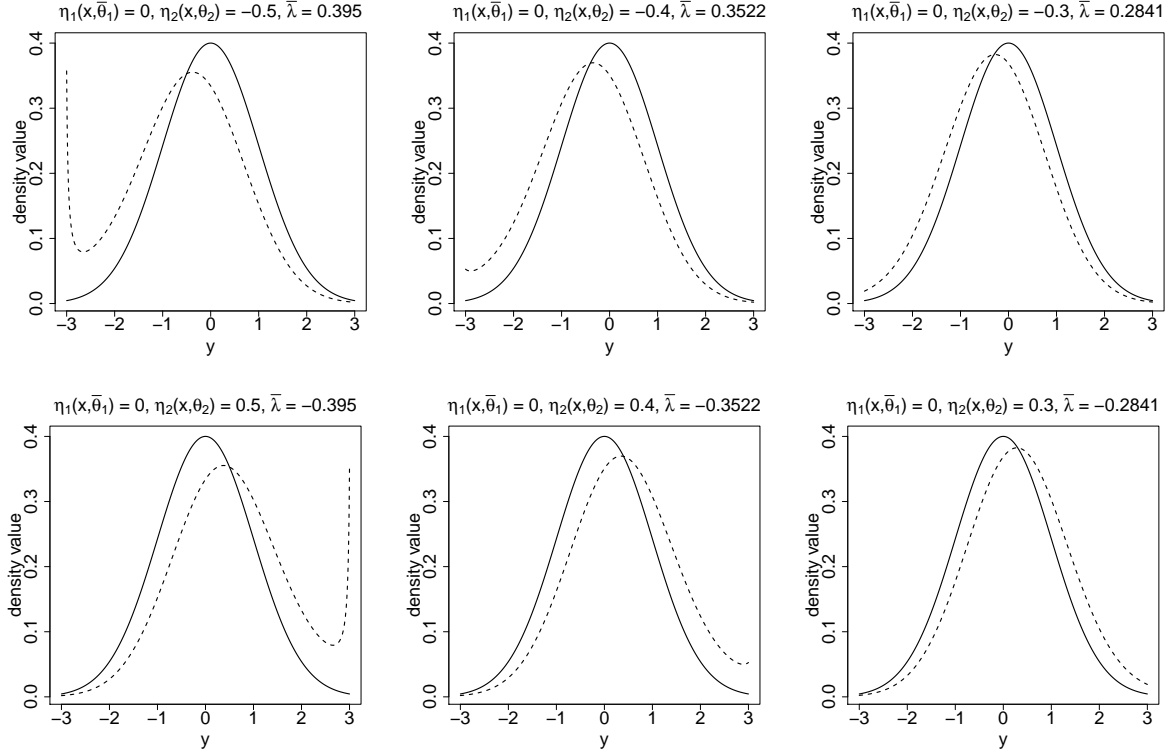


Figure 1: The density f_1 (solid line) of a truncated standard normal distribution on the interval $[-3, 3]$ and the solution f_2^* in (2.12) for various choices of the mean function, that is $\eta_2(x, \theta_2) = -0.5, -0.4, -0.3, +0.5, +0.4, +0.3$.

3 Equivalence theorems

In our setup, the variables we wish to optimize in the design problem are the number of support points, the locations of the support points and the weights at each of the support point. Because the number of support points in the optimal design is not known in advance, the dimension of the optimization is unknown at the onset. A common strategy is to confine the search for the optimal design to a smaller class of designs, say, with a fixed number point designs. This may simplify the search and result in a closed-form description of the optimal design; however, such an approach may not produce a design that is optimal among all designs on \mathcal{X} .

When the design problem is formulated as a convex or concave optimization problem, it is possible to derive an equivalence theorem to verify if a given design is optimal among all designs on the given design space. Each convex or concave functional has a unique equivalence

theorem but they all have same form with an inequality bounded above by 0 with equality at the support points of the optimal design. Frequently, the function on the left hand side of the inequality in the equivalence theorem is called the sensitivity function and if the dimension of the design space is small, optimality of a design can be directly verified by plotting the sensitivity function over the design space and examining the resulting plot. Design monographs, such as Pázman (1986); Silvey (1980); Pukelsheim (2006) present equivalence theorems for various convex criteria. When the design criterion is concave (or convex) and not differentiable, the equivalence theorems involve subgradients and consequently, the optimal designs are more difficult to find and confirm via an equivalence theorem. However, they still have the same form as just discussed, see Wong (1992); Dette (1997) and others, for example. For T -optimality and its other variations, equivalence theorems are available in López-Fidalgo et al. (2007); Dette and Titoff (2009) among others.

Our proposed optimality criterion is a concave functional on the space of all approximate designs and so it is possible to derive an equivalence theorem for confirming the optimality of a design as a semi-parametric optimal discrimination design. The next theorem presents the equivalence theorems for checking the optimality of a design under the two new criteria. A major difference here compared to the “traditional” equivalence theorems is that now ours’ does not involve the Fisher information matrix.

Theorem 3.1 *Suppose that the conditions of Theorem 2.1 hold and the infimum in (2.6) and (2.8) is attained at a unique point $\theta_2^* \in \Theta_2$ for the optimal design ξ^* .*

(a) A design ξ^ is a semi-parametric optimal design for discriminating between the model $f_1(y, x, \bar{\theta}_1)$ and the class \mathcal{F}_2 if and only if the following inequality holds for all $x \in \mathcal{X}$:*

$$I_{1,2}(x, f_1, f_2^*, \bar{\theta}_1, \theta_2^*) - \int_{\mathcal{X}} I_{1,2}(x, f_1, f_2^*, \bar{\theta}_1, \theta_2^*) \xi^*(dx) \leq 0, \quad (3.1)$$

with equality at the support points of ξ^ . Here $I_{1,2}(x, f_1, f_2, \bar{\theta}_1, \theta_2)$ is defined in (2.1),*

$$\theta_2^* = \arg \inf_{\theta_2 \in \Theta_2} \int_{\mathcal{X}} I_{1,2}(x, f_1, f_2^*, \bar{\theta}_1, \theta_2) \xi^*(dx),$$

$$f_2^*(y, x, \theta_2) = \frac{f_1(y, x, \bar{\theta}_1)}{1 + \lambda(y - \eta_2(x, \theta_2))}$$

and λ is found from (2.9). Moreover, there is equality in (3.1) for all support point of ξ^ .*

(b) A design ξ^ is a semi-parametric optimal design for discriminating between the model $f_2(y, x, \theta_2)$ and the class \mathcal{F}_1 if and only if the following inequality holds for all $x \in \mathcal{X}$:*

$$I_{1,2}(x, f_1^*, f_2, \bar{\theta}_1, \theta_2^*) - \int_{\mathcal{X}} I_{1,2}(x, f_1^*, f_2, \bar{\theta}_1, \theta_2^*) \xi^*(dx) \leq 0, \quad (3.2)$$

with equality at the support points of ξ^* . Here

$$\theta_2^* = \arg \inf_{\theta_2 \in \Theta_2} \int_{\mathcal{X}} I_{1,2}(x, f_1^*, f_2, \bar{\theta}_1, \theta_2^*) \xi^*(dx),$$

$$f_1^*(y, x, \bar{\theta}_1) = \frac{f_2(y, x, \theta_2) \exp(-\lambda y)}{\int f_2(y, x, \theta_2) \exp(-\lambda y) dy}$$

and λ is found from (2.15). Moreover, there is equality in (3.2) for all support point of ξ^* .

Example 3.1 This example illustrates the application of the equivalence theorem using an example from Otsu (2008). The goal was to construct semi-parametric optimal designs for discriminating between $f_1(y, x, \bar{\theta}_1)$, the truncated normal density on the interval $[\eta_1(x, \bar{\theta}_1) - 3, \eta_1(x, \bar{\theta}_1) + 3]$ with mean $\eta_1(x, \bar{\theta}_1)$ and variance 1, and densities in the class \mathcal{F}_2 . The two competing models defined on the same design space is $\mathcal{X} = [-1, 1]$ are

$$\eta_1(x, \theta_1) = \theta_{1,1} + \theta_{1,2}e^x + \theta_{1,3}e^{-x}, \quad (3.3)$$

$$\eta_2(x, \theta_2) = \theta_{2,1} + \theta_{2,2}x + \theta_{2,3}x^2, \quad (3.4)$$

where $\bar{\theta}_1 = (4.5, -1.5, -2)$. Otsu (2008) determined the semi-parametric optimal design for discriminating between the model $f_1(y, x, \bar{\theta}_1)$ and the class \mathcal{F}_2 as

$$\tilde{\xi} = \begin{bmatrix} -1.000 & -0.266 & 0.721 & 1.000 \\ 0.377 & 0.198 & 0.244 & 0.181 \end{bmatrix}. \quad (3.5)$$

The left part of Figure 3.1 shows the plots of the sensitivity function on the left-hand side of (3.1) for the design found by Otsu (2008) (left). We conclude that this design is not optimal for discriminating between the model $f_1(y, x, \bar{\theta}_1)$ and the class \mathcal{F}_2 , because the function is positive throughout the design space. The subfigure on the right shows the sensitivity function of the design

$$\xi^* = \begin{bmatrix} -1.000 & -0.670 & 0.142 & 0.959 \\ 0.253 & 0.428 & 0.247 & 0.072 \end{bmatrix}, \quad (3.6)$$

which has been found maximizing the criterion provided by Theorem 2.1 and confirms its optimality.

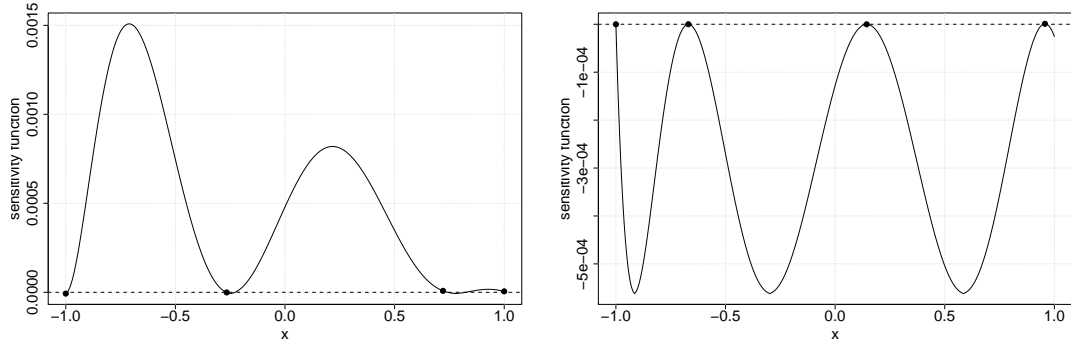


Figure 2: *The sensitivity functions of two designs for discriminating between models (3.3) and (3.4). The left figure is for the design (3.5) calculated by Otsu (2008) and the right figure is for the T -optimal design (3.6).*

The values of the criterion (2.6) for different designs are given by $K_{(a)}(\tilde{\xi}) = 3.27779 * 10^{-6}$ and $K_{(a)}(\xi^*) = 0.0005580455$ demonstrating that ξ^* performs substantially better than $\tilde{\xi}$; in fact, the efficiency of $\tilde{\xi}$ of the design determined by Otsu (2008) relative to ξ^* is approximately 0.006. It turns out that the design in (3.6) is also the T -optimal design. In the next section, we prove in Theorem 4.1 that for models (3.3) and (3.4), the semi-parametric optimal design for discriminating between the model $f_1(y, x, \bar{\theta}_1)$ and the class \mathcal{F}_2 actually coincides with the T -optimal design proposed by Atkinson and Fedorov (1975a).

4 Analytical results

In this section we investigate relationships between the semi-parametric optimality criteria considered in this paper and the optimality criteria for model discrimination with normal errors [see Atkinson and Fedorov (1975a) and Ucinski and Bogacka (2005)]. A "classical" T -optimal design, say ξ_T^* , for discriminating between two models maximizes the criterion

$$\inf_{\theta_2 \in \Theta_2} \int_{\mathcal{X}} [\eta_1(x, \bar{\theta}_1) - \eta_2(x, \theta_2)]^2 \xi(dx) \quad (4.1)$$

among all designs on the design space [see Atkinson and Fedorov (1975a)], where $\bar{\theta}_1$ is fixed. We will presume throughout this section that the infimum in (4.1) is attained at a unique point θ_2^* when $\xi = \xi_T^*$. It follows by similar arguments as given in Wiens (2009) that under the assumption of a normal distribution, the power of the likelihood ratio test for the hypotheses

$$H_0 : \eta(x) = \eta_1(x, \bar{\theta}_1) \text{ versus } H_1 : \eta(x) = \eta_2(x, \theta_2) \text{ for some } \theta_2 \in \Theta_2 \quad (4.2)$$

is an increasing function of the criterion (4.1). The following results give a sufficient condition for the T -optimal discriminating design to be a semi-parametric optimal design in the sense of Section 2.

Theorem 4.1 *Let the assumptions of the Theorem 2.1 (a) be fulfilled and assume further that the density $f_1(y, x, \bar{\theta}_1)$ can be represented in the form*

$$f_1(y, x, \bar{\theta}_1) = g(y - \eta_1(x, \bar{\theta}_1)) \quad (4.3)$$

where g is a symmetric density function supported in the interval $[-a, a]$, i.e. f_1 has support $[-a + \eta_1(x, \bar{\theta}_1), a + \eta_1(x, \bar{\theta}_1)]$. The T -optimal discriminating design maximizing the criterion (4.1) is a semi-parametric optimal design for discriminating between the model $f_1(y, x, \bar{\theta}_1)$ and the class \mathcal{F}_2 .

A similar result is available for the semi-parametric optimal designs for discriminating between the model $f_2(y, x, \theta_2)$ and the class \mathcal{F}_1 . For this purpose we consider the situation, where $f_2(y, x, \theta_2)$ is the density of the normal distributions $\mathcal{N}(\eta_2(x, \theta_2), v_2^2(x, \theta_2))$. If $f_1(y, x, \bar{\theta}_1)$ is also the density of a $\mathcal{N}(\eta_1(x, \bar{\theta}_1), v_2^2(x, \theta_2))$ it can be shown that the power of the likelihood ratio test for the hypotheses (4.2) is an increasing function of the quantity

$$\inf_{\theta_2 \in \Theta_2} \int_{\mathcal{X}} \frac{[\eta_1(x, \bar{\theta}_1) - \eta_2(x, \theta_2)]^2}{v_2^2(x, \theta_2)} \xi(dx). \quad (4.4)$$

The maximization of this expression corresponds to the KL-optimal design problem for discriminating between two normal distributions with the same variance as considered by López-Fidalgo et al. (2007). Our following result shows that this design is also optimal for discriminating between the model $f_2(y, x, \theta_2)$ and the class \mathcal{F}_1 .

Theorem 4.2 *Assume that $f_2(y, x, \theta_2)$ is the density of a normal distribution with mean $\eta_2(x, \theta_2)$ and variance $v_2^2(x, \theta_2)$. The optimal design maximizing (4.4) is a semi-parametric optimal design for discriminating between the model $f_2(y, x, \theta_2)$ and the class \mathcal{F}_1 and vice versa. Moreover, the best approximation $f_1^*(y, x, \bar{\theta}_1)$ is a normal density with mean $\eta_1(x, \bar{\theta}_1)$ and variance $v_2^2(x, \theta_2)$.*

5 Numerical results

The numerical construction of semi-parametric optimal discrimination designs is a very challenging problem. In this section we describe techniques for finding semi-parametric optimal designs and illustrate our approach in two examples. It follows from the results of Section 2 that the first step in the determination of the optimal designs consists in an efficient solution of the equations (2.9) and (2.15). In a second step any numerical method for the determination of KL-optimal discrimination designs can be adapted to the minimax problems obtained from Theorem 2.1 as the representations (2.11) and (2.13) have a similar structure as the KL-optimality criteria considered in López-Fidalgo et al. (2007). Even the second step defines a very challenging problem and some recent results and algorithms for KL-optimality criteria can be found in (Stegmaier et al., 2013), Braess and Dette (2013); Dette et al. (2015) and Dette et al. (2016a). As the focus in this paper is on the new semi-parametric design criteria

we concentrate on the first step in the following discussion. For the second step we used an adaptation of the first-order algorithm of Atkinson and Fedorov (1975a), because it can easily be implemented.

Let δ be a user-selected positive constant. For finding the numerical solution of the equation (2.9) we propose the following algorithm:

- if $\eta_1(x, \bar{\theta}_1) = \eta_2(x, \theta_2)$, set $\lambda = 0$;
- if $\eta_1(x, \bar{\theta}_1) < \eta_2(x, \theta_2)$, choose a solution in the interval $\Lambda^- = [-1/(y_{x,\max} - \eta_2(x, \theta_2)), -\delta]$;
- if $\eta_1(x, \bar{\theta}_1) > \eta_2(x, \theta_2)$, choose a solution in the interval $\Lambda^+ = [\delta, -1/(y_{x,\min} - \eta_2(x, \theta_2))]$.

Similarly, the solution of (2.15) can be obtained as follows. We search for $\lambda > 0$ if $\eta_1(x, \bar{\theta}_1) < \eta_2(x, \theta_2)$ so that λ shifts the predefined density $f_2(y, x, \theta_2)$ to the left and, search for $\lambda < 0$ if $\eta_1(x, \bar{\theta}_1) > \eta_2(x, \theta_2)$. Let β be a large user-selected positive constant, let δ be a small positive constant and assume that the solution of (2.15) is in $[-\beta, +\beta]$. We note that

$$\mu'(\lambda)f_2(y, x, \theta_2) \exp(-\lambda y) > 0, \quad \forall \lambda \in [-\beta, +\beta],$$

for all $y \in \mathcal{S}_{f_2, \theta_2, x}$, where $\mu'(\lambda)$ is defined in (2.14). We suggest the following algorithm for finding a numeral solution of (2.15):

- if $\eta_1(x, \bar{\theta}_1) = \eta_2(x, \theta_2)$, set $\lambda = 0$;
- if $\eta_1(x, \bar{\theta}_1) < \eta_2(x, \theta_2)$, choose a solution in the interval $\Lambda^+ = [+ \delta, +\beta]$;
- if $\eta_1(x, \bar{\theta}_1) > \eta_2(x, \theta_2)$, choose a solution in the interval $\Lambda^- = [-\beta, -\delta]$.

We now present two examples, where the T -optimal and semi-parametric KL -optimal designs are determined numerically and shown to be different. To be precise, consider the optimal design problem from López-Fidalgo et al. (2007), where they were interested to discriminate between the two models:

$$\eta_1(x, \theta_1) = \theta_{1,1}x + \frac{\theta_{1,2}x}{x + \theta_{1,3}}, \quad (5.1)$$

$$\eta_2(x, \theta_2) = \frac{\theta_{2,1}x}{x + \theta_{2,2}}. \quad (5.2)$$

The design space for both models is the interval $[0.1, 5]$ and we assume that the first model has fixed parameters $\bar{\theta}_1 = (1, 1, 1)$. We construct three different types of optimal discrimination designs for this problem: a T -optimal design, a KL -optimal design for lognormal errors (with fixed variances $v_1^2(x, \bar{\theta}_1) = v_2^2(x, \theta_2) = 0.1$) and a semi-parametric KL -optimal design (case a)) for a mildly truncated lognormal density $f_1(y, x, \bar{\theta}_1)$ with location and scale parameters, respectively, given by

$$\begin{aligned} \mu_1(x, \bar{\theta}_1) &= \log [\eta_1(x, \bar{\theta}_1)] - \frac{1}{2} \log [1 + v_1^2(x, \bar{\theta}_1)/\eta_1^2(x, \bar{\theta}_1)], \\ \sigma_1^2(x, \bar{\theta}_1) &= \log [1 + v_1^2(x, \bar{\theta}_1)/\eta_1^2(x, \bar{\theta}_1)]. \end{aligned}$$

| Design type | ξ^* | | | θ_2^* |
|----------------|---------|-------|-------|------------------|
| T -optimal | 0.508 | 2.992 | 5.000 | (22.564, 14.637) |
| KL -optimal | 0.580 | 0.298 | 0.122 | |
| | 0.206 | 2.826 | 5.000 | (20.552, 12.962) |
| | 0.574 | 0.308 | 0.118 | |
| SKL -optimal | 0.454 | 2.961 | 5.000 | (22.045, 14.197) |
| | 0.531 | 0.344 | 0.125 | |

Table 1: T , KL - and semi-parametric KL -optimal discrimination (SKL) designs for discriminating between models (5.1) and (5.2).

| $K(\xi) \setminus \xi$ | T | KL | SKL |
|------------------------|-------|-------|-------|
| T | 1 | 0.247 | 0.741 |
| KL | 0.653 | 1 | 0.787 |
| SKL | 0.55 | 0.397 | 1 |

Table 2: Efficiencies of the T , KL - and semi-parametric KL -optimal discrimination (SKL) designs for the models (5.1) and (5.2) under different optimality criteria.

The range for this density is the interval from $Q(0.0001, x, \bar{\theta}_1)$ to $Q(0.9999, x, \bar{\theta}_1)$, where $Q(p, x, \bar{\theta}_1)$ is quantile function of ordinary lognormal density with mean $\eta_1(x, \bar{\theta}_1)$ and variance $v_1^2(x, \bar{\theta}_1) = 0.1$. We note that because of mild truncation $\eta_1(x, \bar{\theta}_1)$ does not correspond exactly to the mean of $f_1(y, x, \bar{\theta}_1)$ but is very close to it. The optimal discrimination designs under the three different criteria are displayed in Table 1, where we also show the ‘‘optimal’’ parameter θ_2^* of the second model corresponding to the minimal value with respect to the parameter θ_2 . We observe substantial differences between the T , KL - and semi-parametric KL -optimal discrimination (SKL) designs. Figure 3 displays the sensitivity functions of the three designs and confirms the optimality of each of the designs.

Table 2 displays the efficiencies of T , KL - and semi-parametric KL -optimal discrimination (SKL) designs with respect to the different criteria. For example, the value 0.247 in the first row is the efficiency of the KL -optimal design with respect to the T -optimality criterion. We observe that the T - and KL - optimal discrimination design are not very robust under a variation of the criteria, where the KL -optimal discrimination design has slight advantages. On the other hand, the semi-parametric KL -optimal discrimination design yields moderate efficiencies (about 75%) with respect to the T - and KL -optimality criterion. Figure 4 shows the plots of the functions $f_1(y, x_i^*, \bar{\theta}_1)$ and $f_2^*(y, x_i^*, \theta_2^*)$ for the support points x_i^* , $i = 1, 2, 3$ of SKL -optimal design. Above each figure the values $\eta_1(x_i, \bar{\theta}_1)$, $\eta_2(x_i, \theta_2^*)$ and $\bar{\lambda}(x_i^*, \bar{\theta}_1, \theta_2^*)$, the solution of (2.9), are presented. We note that the densities f_1 and f_2^* depend on the parameters $\bar{\theta}_1$ and θ_2^* only through $\eta_1(x, \bar{\theta}_1)$ and $\eta_2(x, \theta_2^*)$.

As a second example we now consider a similar problem with a different function $\eta_1(x, \theta_1)$ (Wiens, 2009). The two models of interest are

$$\eta_1(x, \theta_1) = \theta_{1,1} \{1 - \exp(-\theta_{1,2}x)\}, \quad (5.3)$$

$$\eta_2(x, \theta_2) = \frac{\theta_{2,1}x}{\theta_{2,2} + x}, \quad (5.4)$$

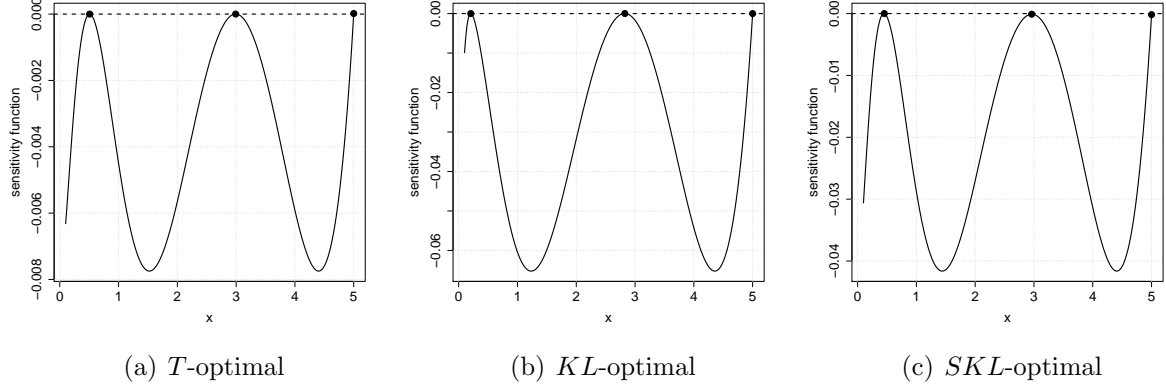


Figure 3: The sensitivity functions of the *T*-, *KL*- and semi-parametric *KL*-optimal discrimination (*SKL*) designs in Table 1.

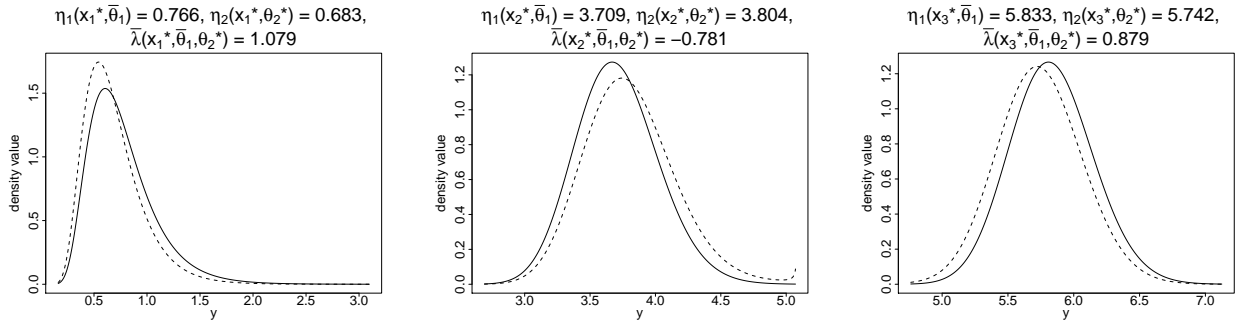


Figure 4: The density $f_1(y, x_i^*, \bar{\theta}_1)$ of a truncated lognormal distribution (solid line) on the interval $[Q(0.0001, x_i^*, \bar{\theta}_1), Q(0.9999, x_i^*, \bar{\theta}_1)]$ and the corresponding “optimal” density $f_2^*(y, x_i^*, \theta_2^*)$, defined by (2.12), for the support points x_i^* , $i = 1, 2, 3$ of *SKL*-optimal design and parameter values θ_2^* from Table 1.

| Design type | ξ^* | | | θ_2^* |
|----------------|---------|-------|-------|----------------|
| T -optimal | 0.308 | 2.044 | 5.000 | (1.223, 0.948) |
| | 0.316 | 0.428 | 0.256 | |
| KL -optimal | 0.140 | 1.916 | 5.000 | (1.242, 1.006) |
| | 0.333 | 0.402 | 0.465 | |
| SKL -optimal | 0.395 | 2.090 | 5.000 | (1.216, 0.920) |
| | 0.396 | 0.355 | 0.249 | |

Table 3: T , KL - and semi-parametric KL -optimal discrimination (SKL) designs for discriminating between the models (5.3) and (5.4).

| $K(\xi) \setminus \xi$ | T | KL | SKL |
|------------------------|-------|-------|-------|
| T | 1 | 0.360 | 0.663 |
| KL | 0.835 | 1 | 0.655 |
| SKL | 0.407 | 0.361 | 1 |

Table 4: Efficiencies of the T , KL - and semi-parametric KL -optimal discrimination (SKL) designs for the models (5.3) and (5.4) under various optimality criteria.

where the design space is given by the interval $\mathcal{X} = [0.1, 5]$. Here we fix the parameters of the model (5.3) as $\bar{\theta}_1 = (1, 1)$ and determine the T -optimal, KL -optimal (for lognormal errors) and a semi-parametric KL -optimal design (case a)) for mildly truncated lognormal errors. The error variances for the KL -optimal discrimination design are $v_1^2(x, \bar{\theta}_1) = v_2^2(x, \theta_2) = 0.02$ and for the semi-parametric KL -optimal the variance is $v_1^2(x, \bar{\theta}_1) = 0.02$. The optimal designs with respect to the different criteria are presented in Table 3, where we also show the corresponding parameter θ_2^* of the second model corresponding to the minimal value with respect to the parameter θ_2 . We observe again substantial differences between the optimal discrimination designs with respect to the different criteria, and a comparison of the efficiencies of the optimal designs with respect to the different criteria in Table 4 shows a similar picture as in the first example. Figure 5 displays the sensitivity functions of the T -, KL - and semi-parametric KL -optimal discrimination designs and the three subplots confirm optimality of the three three-point designs. Figure 6 shows the plots of the functions $f_1(y, x_i^*, \bar{\theta}_1)$ and $f_2^*(y, x_i^*, \theta_2^*)$ for the support points x_i^* , $i = 1, 2, 3$ of SKL -optimal design from Table 3, the legend above each plot has the same meaning as before.

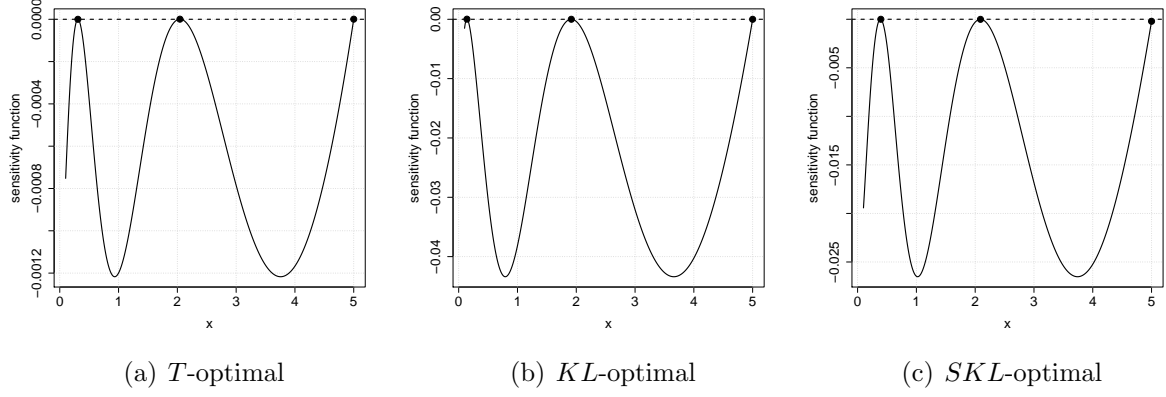


Figure 5: The sensitivity functions of the *T*-, *KL*- and semi-parametric *KL*-optimal discrimination (*SKL*) designs in Table 3.

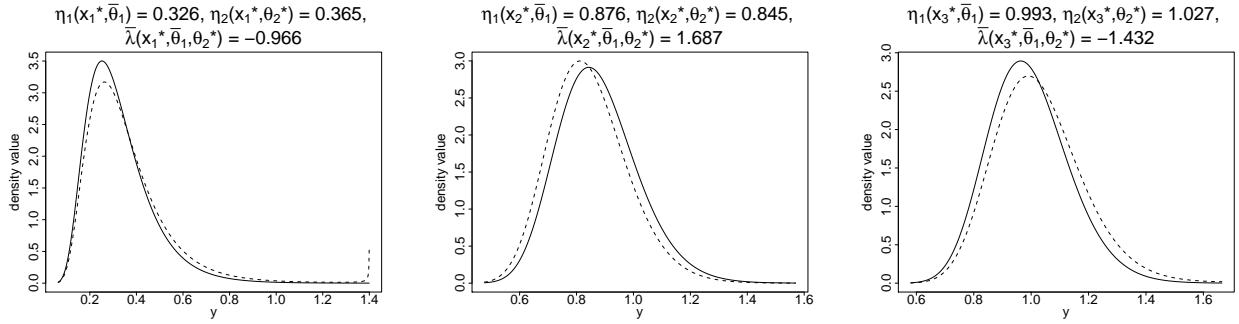


Figure 6: The density $f_1(y, x_i^*, \bar{\theta}_1)$ of a truncated lognormal distribution (solid line) and the “optimal” density $f_2^*(y, x_i^*, \theta_2^*)$ for the support points x_i^* , $i = 1, 2, 3$ of *SKL*-optimal design and corresponding “optimal” parameter values θ_2^* from Table 3.

6 Proofs

Proof of Theorem 2.1: We only prove the first part of the proposition. The second statement follows by similar arguments which are omitted for the sake of brevity. We first introduce the following temporary notation

$$f_1(y, x, \bar{\theta}_1) = f_1(y), \quad f_2(y, x, \theta_2) = f_2(y), \quad \eta_1(x, \bar{\theta}_1) = \eta_1, \quad \eta_2(x, \theta_2) = \eta_2, \quad (6.1)$$

and note that from (2.6), we wish to minimize

$$I_{1,2}(x, f_1, f_2, \theta_1, \theta_2) = \int \log \left\{ \frac{f_1(y)}{f_2(y)} \right\} f_1(y) dy$$

with constraints

$$\int f_2(y) dy = 1, \quad \text{and} \quad \int y f_2(y) dy = \eta_2. \quad (6.2)$$

Consequently, with the notation

$$H = \log \left\{ \frac{f_1(y)}{f_2(y)} \right\} f_1(y) + \mu f_2(y) + \lambda y f_2(y),$$

we obtain the Euler-Lagrange equation

$$\frac{\partial H}{\partial f_2} = -\frac{f_1(y)}{f_2(y)} + \mu + \lambda y = 0$$

and the relation

$$f_2(y) = \frac{f_1(y)}{\mu + \lambda y}. \quad (6.3)$$

We may assume without loss of generality that $\lambda \neq 0$; otherwise, we have $\mu = 1$ and $f_2(y) = f_1(y)$. From the condition (6.2) and (6.3) we obtain

$$\begin{aligned} \eta_2 &= \int y f_2(y) dy = \int \frac{y}{\mu + \lambda y} f_1(y) dy \\ &= \frac{1}{\lambda} \int \frac{\mu + \lambda y}{\mu + \lambda y} f_1(y) dy - \frac{\mu}{\lambda} \int \frac{1}{\mu + \lambda y} f_1(y) dy = \frac{1}{\lambda} - \frac{\mu}{\lambda}. \end{aligned}$$

Hence we have $\mu = 1 - \eta_2 \lambda$ and

$$f_2(y) = \frac{f_1(y)}{1 + \lambda(y - \eta_2)}.$$

Substituting the new expression for $f_2(y)$ into the remaining condition gives

$$\int \frac{f_1(y)}{1 + \lambda(y - \eta_2)} dy = 1. \quad (6.4)$$

By the assumption in the theorem, equation (6.4) has a unique solution in the interval (2.10) and the inequality

$$1 + \bar{\lambda}(y - \eta_2) > 0$$

holds for all $y \in \mathcal{S}_{f_1, \bar{\theta}_1, x}$, where $\bar{\lambda}$ solves (6.4). This implies that f_2 is a density. \square

Proof of Lemma 2.1: We again use the notation in (6.1). Let $\bar{\lambda}$ be the solution of equation (2.9) and f_2^* be the “optimal” density defined by (2.12), then

$$\begin{aligned} \eta_1 - \eta_2 &= \int y f_1(y) dy - \int y f_2^*(y) dy = \int y f_1(y) \left\{ 1 - \frac{1}{1 + \bar{\lambda}(y - \eta_2)} \right\} dy \\ &= \int y f_1(y) \frac{\bar{\lambda}(y - \eta_2)}{1 + \bar{\lambda}(y - \eta_2)} dy = \bar{\lambda} \int y(y - \eta_2) f_2(y) dy \\ &= \bar{\lambda} \int (y - \eta_2)^2 f_2(y) dy + \int (y\eta_2 - \eta_2^2) f_2(y) dy = \bar{\lambda} v_2^2, \end{aligned}$$

where $v_2^2 > 0$ is the variance of $f_2(y)$, and the last equality follows from the fact that

$$\int y f_2(y) dy = \eta_2.$$

\square

Proof of Theorem 3.1: Again we only prove part (a). Part (b) follows by similar arguments. Roughly speaking, the equivalence theorem is a consequence of the equivalence theorem for KL-optimal designs using the specific density for the second model

$$f_2^*(y, x, \theta_2) = \frac{f_1(y, x, \bar{\theta}_1)}{1 + \lambda(y - \eta_2(x, \theta_2))}$$

in the KL-criterion. More specifically, the criterion $K_{(a)}(\xi, \bar{\theta}_1)$ is concave and to obtain the equivalence theorem, we calculate the directional derivative as follows:

$$\frac{\partial K_{(a)}([1 - \alpha]\xi + \alpha\xi_{x^*}, \bar{\theta}_1)}{\partial \alpha} = \frac{\partial \inf_{\theta_2 \in \Theta_2} \int_{\mathcal{X}} I_{1,2}(x, f_1, f_2, \bar{\theta}_1, \theta_2)([1 - \alpha]\xi + \alpha\xi_{x^*})(dx)}{\partial \alpha},$$

where ξ_{x^*} is the Dirac measure at x^* . By theorem in Pshenichnyi (1971) on page 75, and the

assumptions in our theorem, we have

$$\begin{aligned} \frac{\partial K_{(a)}([1 - \alpha]\xi + \alpha\xi_{x^*}, \bar{\theta}_1)}{\partial \alpha} &= \int_{\mathcal{X}} I_{1,2}(x, f_1, f_2, \bar{\theta}_1, \widehat{\theta}_2) \xi_{x^*}(dx) - \inf_{\theta_2 \in \Theta_2} \int_{\mathcal{X}} I_{1,2}(x, f_1, f_2, \bar{\theta}_1, \theta_2) \xi(dx) \\ &= I_{1,2}(x^*, f_1, f_2, \bar{\theta}_1, \widehat{\theta}_2) - K_{(a)}(\xi, \bar{\theta}_1), \end{aligned}$$

where

$$\widehat{\theta}_2 = \arg \inf_{\theta_2 \in \Theta_2} \int_{\mathcal{X}} I_{1,2}(x, f_1, f_2, \bar{\theta}_1, \theta_2) \xi(dx).$$

The design ξ^* is optimal if and only if the inequality

$$\frac{\partial K_{(a)}([1 - \alpha]\xi^* + \alpha\xi_x, \bar{\theta}_1)}{\partial \alpha} \leq 0.$$

holds for every $x \in \mathcal{X}$, which means that we can not improve the concave functional $K_{(a)}(\xi^*, \bar{\theta}_1)$ by moving in any direction from ξ^* . \square

Proof of Theorem 4.1: With the notations (6.1) consider the function

$$\begin{aligned} F(\eta_1, \eta_2, \lambda) &= \int_{-a+\eta_1}^{a+\eta_1} \log \{1 + \lambda(y - \eta_2)\} f_1(y) dy \\ &= \int_{-a}^a \log \{1 + \lambda(t - \Delta_\eta)\} g(t) dt =: F(\Delta_\eta, \lambda), \end{aligned}$$

where we have used the representation (4.3) and the notation $\Delta_\eta = \eta_2 - \eta_1$. Similarly, we obtain for the left-hand side of (2.9) the representation

$$G(\eta_1, \eta_2, \lambda) = \int_{-a+\eta_1}^{a+\eta_1} \frac{f_1(y)}{1 + \lambda(y - \eta_2)} dy = \int_{-a}^a \frac{g(t)}{1 + \lambda(t - \Delta_\eta)} dt.$$

We now define the function

$$F(\Delta_\eta) = F(\Delta_\eta, \bar{\lambda}(\Delta_\eta)),$$

where $\bar{\lambda}(\Delta_\eta)$ is a non-zero solution of the equation $G(\eta_1, \eta_2, \lambda) = 1$, which exists for all $x \in \mathcal{X}$ and for all $\theta_2 \in \Theta_2$ by the assumption of the theorem. Note that the function

$$h(t, \Delta_\eta) = \frac{g(t)}{1 + \bar{\lambda}(\Delta_\eta)(t - \Delta_\eta)}$$

is a density function with mean Δ_η , because $g(t)$ is symmetric on $[-a, a]$ with mean 0. For

the derivative of $F(\Delta_\eta)$ we therefore obtain

$$\begin{aligned}\frac{\partial F}{\partial \Delta_\eta}(\Delta_\eta) &= \frac{\partial}{\partial \Delta_\eta} \int_{-a}^a \log \{1 + \bar{\lambda}(\Delta_\eta)(t - \Delta_\eta)\} g(t) dt \\ &= \int_{-a}^a \left[\frac{\partial \bar{\lambda}}{\partial \Delta_\eta}(\Delta_\eta)(t - \Delta_\eta) - \bar{\lambda}(\Delta_\eta) \right] \frac{g(t)}{1 + \bar{\lambda}(\Delta_\eta)(t - \Delta_\eta)} dt \\ &= \frac{\partial \bar{\lambda}}{\partial \Delta_\eta}(\Delta_\eta) \int_{-a}^a (t - \Delta_\eta) h(t, \Delta_\eta) dt - \bar{\lambda}(\Delta_\eta) \int_{-a}^a h(t, \Delta_\eta) dt = -\bar{\lambda}(\Delta_\eta).\end{aligned}$$

Here the first integral vanishes because h has mean Δ_η and the second integral is equal to 1 because h is a density. From Lemma 2.1, we have the following implications: If

$$\eta_1 > \eta_2, \text{ then } \Delta_\eta < 0 \Rightarrow \bar{\lambda}(\Delta_\eta) > 0 \Rightarrow F(\Delta_\eta) \text{ is decreasing,}$$

and if

$$\eta_1 < \eta_2, \text{ then } \Delta_\eta > 0 \Rightarrow \bar{\lambda}(\Delta_\eta) < 0 \Rightarrow F(\Delta_\eta) \text{ is increasing.}$$

We also note that the symmetry of g implies the symmetry of F , i.e. $F(\Delta_\eta) = F(-\Delta_\eta)$. Let $\xi_T^* = \{x_1^*, \dots, x_k^*; \omega_1^*, \dots, \omega_k^*\}$ be a T -optimal discriminating design maximizing (4.1) and define

$$\theta_2^* = \arg \inf_{\theta_2} \int [\eta_1(x, \bar{\theta}_1) - \eta_2(x, \theta_2)]^2 \xi_T^*(dx).$$

By the equivalence theorem for T -optimal designs (Dette and Titoff, 2009), it follows that for all $x \in \mathcal{X}$, we have

$$|\eta_1(x, \bar{\theta}_1) - \eta_2(x, \theta_2^*)| \leq \epsilon = |\eta_1(x_1^*, \bar{\theta}_1) - \eta_2(x_1^*, \theta_2^*)| = \dots = |\eta_1(x_n^*, \bar{\theta}_1) - \eta_2(x_n^*, \theta_2^*)|.$$

These arguments and the representation (2.12) for the ‘‘optimal’’ density f_2^* yield for all $x \in \mathcal{X}$

$$\begin{aligned}I_{1,2}(x, f_1, f_2^*, \theta_1, \theta_2^*) &= F(\eta_1(x, \bar{\theta}_1) - \eta_2(x, \theta_2^*)) \\ &= F(|\eta_1(x, \bar{\theta}_1) - \eta_2(x, \theta_2^*)|) \leq F(\epsilon) = \sum_{i=1}^k \omega_i^* F(\epsilon) \\ &= \sum_{i=1}^k \omega_i^* F(\eta_1(x_i^*, \bar{\theta}_1) - \eta_2(x_i^*, \theta_2^*)) = \int I_{1,2}(x, f_1, f_2^*, \theta_1, \theta_2^*) d\xi^*(x).\end{aligned}$$

Here the second equality follows from the symmetry of F and the inequality is a consequence of the monotonicity of F . By Theorem 3.1 this means that the design ξ^* is a semi-parametric optimal design for discriminating between the model $f_1(y, x, \bar{\theta}_1)$ and the class \mathcal{F}_2 . \square

Proof of Theorem 4.2: We will show that the criterion (2.13) is equivalent to the criterion (4.4) when $f_2(y, x, \theta_2)$ is a normal density. For simplicity, let $\eta_1 = \eta_1(x, \bar{\theta}_1)$, let

$\eta_2 = \eta_2(x, \theta_2)$ and let $v_2 = v_2(x, \theta_2)$. It follows from part (b) of Theorem 2.1 that

$$\begin{aligned} f_1^*(y, x, \bar{\theta}_1) \propto f_2(y) \exp(-\lambda y) &= \frac{1}{\sqrt{2\pi v_2}} \exp \left[-\frac{\{y - \eta_2\}^2}{2v_2^2} - \lambda y \right] \\ &= \frac{1}{\sqrt{2\pi v_2}} \exp \left[-\frac{\{y - (\eta_2 - v_2^2 \lambda)\}^2}{2v_2^2} \right] \exp \left[-\eta_2 \lambda + \frac{v_2^2 \lambda^2}{2} \right]. \end{aligned}$$

The condition that $f_1^*(y, x, \bar{\theta}_1)$ is a density with mean η_1 yields

$$\lambda = \frac{\eta_2 - \eta_1}{v_2^2}, \quad \mu' = \exp \left[\eta_2 \lambda - \frac{v_2^2 \lambda^2}{2} \right],$$

which implies that $f_1^*(y, x, \bar{\theta}_1)$ is a normal density with mean η_1 and variance v_2^2 . Then the KL-divergence between $f_1^*(y, x, \bar{\theta}_1)$ and $f_2(y, x, \theta_2)$ is given by $[\eta_1 - \eta_2]^2/v_2^2$, which yields the criterion (4.4) and completes the proof. \square

Acknowledgements. Parts of this work were done during a visit of the second author at the Department of Mathematics, Ruhr-Universität Bochum, Germany. The work of H. Dette and R. Guchenko was supported by the Deutsche Forschungsgemeinschaft (SFB 823: Statistik nichtlinearer dynamischer Prozesse, Teilprojekt C2). The research of H. Dette and W.K. Wong reported in this publication was also partially supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number R01GM107639. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The work of V. Melas and R. Guchenko was also partially supported by St. Petersburg State University (project "Actual problems of design and analysis for regression models", 6.38.435.2015).

References

- Abd El-Monsef, M. M. E. and Seyam, M. M. (2011). A CDT-optimum designs for model discrimination, parameter estimation and estimation of a parametric function. *Journal of Statistical Planning and Inference*, 141:639–643.
- Alberston, A. L., Schwaab, M., Labao, M. W. N., and C., P. J. (2011). Experimental design for the joint model discrimination and precise parameter estimation through information measures. *Chemical Engineering Science*, 66.
- Atkinson, A. C. (2008). *DT*-optimum designs for model discrimination and parameter estimation. *Journal of Statistical Planning and Inference*, 138:56–64.
- Atkinson, A. C., Bogacka, B., and Bogacki, M. B. (1998). *D*- and *T*-optimum designs for the kinetics of a reversible chemical reaction. *Chemometrics and Intelligent Laboratory Systems*, 43:185–198.
- Atkinson, A. C. and Fedorov, V. V. (1975a). The designs of experiments for discriminating between two rival models. *Biometrika*, 62:57–70.

- Atkinson, A. C. and Fedorov, V. V. (1975b). Optimal design: Experiments for discriminating between several models. *Biometrika*, 62:289–303.
- Borwein, J. M. and Lewis, A. S. (1991). Duality relationships for entropy-like minimization problems. *SIAM J. Control Optim.*, 29:325–338.
- Braess, D. and Dette, H. (2013). Optimal discriminating designs for several competing regression models. *Annals of Statistics*, 41(2):897–922.
- Campos-Barreiro, S. and Lopez-Fidalgo, J. (2016). KL-optimal experimental design for discriminating between two growth models applied to a beef farm. *Mathematical Biosciences and Engineering*, 13:In press.
- Chernoff, H. (1953). Locally optimal designs for estimating parameters. *Annals of Mathematical Statistics*, 24:586–602.
- Covagnaro, D. R., Myung, J. I., Pitt, M. A., and Kujala, J. V. (2010). Adaptive design optimization: a mutual information-based approach to model discrimination in cognitive science. *Neural Computation*, 22.
- Dette, H. (1997). Designing experiments with respect to “standardized” optimality criteria. *Journal of the Royal Statistical Society, Ser. B*, 59:97–110.
- Dette, H., Melas, V. B., and Guchenko, R. (2015). Bayesian T -optimal discriminating designs. *Annals of Statistics*, to appear,.
- Dette, H., Melas, V. B., and Guchenko, R. (2016a). Efficient computation of bayesian optimal discriminating designs. *Journal of Computational and Graphical Statistics*, to appear,.
- Dette, H., Melas, V. B., and Shpilev, P. (2012). T -optimal designs for discrimination between two polynomial models. *Annals of Statistics*, 40(1):188–205.
- Dette, H., Melas, V. B., and Shpilev, P. (2013). Robust T -optimal discriminating designs. *Annals of Statistics*, 41(4):1693–1715.
- Dette, H., Melas, V. B., and Shpilev, P. (2016b). T -optimal discriminating designs for fourier regression models. *Computational Statistics and Data Analysis*, to appear,.
- Dette, H. and Titoff, S. (2009). Optimal discrimination designs. *Annals of Statistics*, 37(4):2056–2082.
- Donckels, B. M. R., De Pauw, D. J. W., De Baets, B., Maertens, J., and Vanrolleghem, P. A. (2009). An anticipatory approach to optimal experimental design for model discrimination. *Chemometrics and Intelligent Laboratory Systems*, 95:53–63.
- Felsenstein, K. (1992). Optimal Bayesian design for discrimination among rival models. *Computational Stat. & Data Analysis*, 14:427–436.
- Ghosh, S. and Dutta, S. (2013). Robustness of designs for model discrimination. *J. of Multivariate Analysis*, 115.
- Jansen, K. M., Duffull, S. B., Tarning, J., Price, R. N., and Simpson, J. (2013). A robust design for identification of the parasite clearance estimator. *Malaria Journal*, 12(1):410–416.
- Kiefer, J. (1974). General equivalence theory for optimum designs (approximate theory). *Annals of Statistics*, 2(5):849–879.
- López-Fidalgo, J., Tommasi, C., and Trandafir, P. C. (2007). An optimal experimental design criterion for discriminating between non-normal models. *Journal of the Royal Statistical Society, Series B*, 69:231–242.

- Myung, J. I. and Pitt, M. A. (2009). Optimal experimental design for model discrimination. *Psychol Rev.*, 116(3):499–518.
- Ng, S. H. and Chick, S. E. (2004). Design of follow-up experiments for improving model discrimination and parameter estimation. *Naval Research Logistics*, 2(22):1–11.
- Otsu, T. (2008). Optimal experimental design criterion for discriminating semi-parametric models. *Journal of Statistical Planning and Inference*, 138:4141–4150.
- Pázman, A. (1986). *Foundations of Optimum Experimental Design*. D. Reidel Publishing Company, Dordrecht.
- Pshenichnyi, B. N. (1971). *Necessary Conditions for an Extremum*. Dekker, New York.
- Pukelsheim, F. (2006). *Optimal Design of Experiments*. SIAM, Philadelphia.
- Silvey, S. (1980). *Optimal Design*. Chapman & Hall.
- Stegmaier, J., Skanda, D., and Lebedz, D. (2013). Robust optimal design of experiments for model discrimination using an interactive software tool. *PLOS ONE*, 8(2):e55723. doi:10.1371/journal.pone.0055723.
- Stigler, S. (1971). Optimal experimental design for polynomial regression. *Journal of the American Statistical Association*, 66:311–318.
- Tommasi, C. and López-Fidalgo, J. (2010). Bayesian optimum designs for discriminating between models with any distribution. *Computational Statistics & Data Analysis*, 54(1):143–150.
- Tommasi, C., Martin-Martin, R., and Jopez-Fidalgo, J. (2016). Max-min optimal discriminating designs for several statistical models. *Statistics and Computing*, 26(6):1163–1172.
- Ucinski, D. and Bogacka, B. (2005). T -optimum designs for discrimination between two multiresponse dynamic models. *Journal of the Royal Statistical Society, Ser. B*, 67:3–18.
- Waterhouse, T. H., Woods, D. C., Eccleston, J. A., and Lewis, S. M. (2008). Design selection criteria for discrimination/estimation for nested models and a binomial response. *Journal of Statistical Planning and Inference*, 138:132–144.
- Wiens, D. P. (2009). Robust discrimination designs. *Journal of the Royal Statistical Society, Ser. B*, 71:805–829.
- Wong, W. (1992). A unified approach to the construction of minimax designs. *Biometrika*, 79:611–620.
- Yanagisawa, Y. (1990). Designs for discrimination between bivariate binary response models. *Biometrical J.*, 1:25–34.

