Title
Optimal discrimination designs for semiparametric models

Permalink
https://escholarship.org/uc/item/51w4m5hr

Journal
Biometrika, 105(1)

Authors
Guchennko, Roman
Melas, V. B.
Wong, Weng Kee

Publication Date
2017-10-26

DOI
10.1093/biomet/asx058

Peer reviewed
Optimal discrimination designs for semiparametric models

BY H. DETTE
Fakultät für Mathematik, Ruhr-Universität Bochum, 44780 Bochum, Germany
holger.dette@rub.de

R. GUCHENKO, V. B. MELAS
Faculty of Mathematics and Mechanics, St. Petersburg State University,
198504 St. Petersburg, Russia
romanguchenko@ya.ru vbmelas@post.ru

AND W. K. WONG
Department of Biostatistics, University of California at Los Angeles,
650 Charles E. Young Dr. South, Los Angeles, California 90095, U.S.A.
wkwong@ucla.edu

SUMMARY

Much work on optimal discrimination designs assumes that the models of interest are fully specified, apart from unknown parameters. Recent work allows errors in the models to be nonnormally distributed but still requires the specification of the mean structures. Otsu (2008) proposed optimal discriminating designs for semiparametric models by generalizing the Kullback–Leibler optimality criterion proposed by López-Fidalgo et al. (2007). This paper develops a relatively simple strategy for finding an optimal discrimination design. We also formulate equivalence theorems to confirm optimality of a design and derive relations between optimal designs found here for discriminating semiparametric models and those commonly used in optimal discrimination design problems.

Some key words: Continuous design; Equivalence theorem; Kullback–Leibler divergence; T-optimality; Variational calculus.

1. Introduction

Optimal discrimination design problems have recently appeared in cognitive science (Covagnaro et al., 2010), psychology (Myung & Pitt, 2009) and chemical engineering (Alberton et al., 2011). A main motivation for such research is that in a scientific study, we often do not know the true underlying model that drives the responses but experts may have a number of candidate models that they believe should be adequate for studying the process. An informed and well-constructed design provides valuable information, so constructing an optimal design to find the most appropriate model among a few plausible models is important. In applications, the optimal discrimination design provides guidance on how data should be collected efficiently to infer the most plausible model before other inferential procedures are employed to attain
the study objectives using the identified model. Our work concerns the first part of such an approach, where the goal is to determine the most appropriate design to discriminate between the models.

The statistical theory for studying optimal discrimination designs dates back to the 1970s. An early reference is Atkinson & Fedorov (1975a,b), who proposed $T$-optimal designs to discriminate between models when errors are normally distributed. $T$-optimality assumes a known null model and we wish to test whether a rival parametric model with unknown parameters holds. When models are all parametric, the likelihood ratio test is typically used to discriminate between the models. The noncentrality parameter of the chi-squared distribution of the test statistic contains the unknown parameters from the alternative model and is proportional to the $T$-optimality criterion (Atkinson & Fedorov, 1975a; Wiens, 2009). Since a larger noncentrality parameter provides a more powerful test, $T$-optimal designs maximize the minimum value of the noncentrality parameter, where the minimum is taken over all possible values of the parameters in the alternative model. The $T$-optimality criterion is not differentiable and finding optimal discrimination designs under the maximin design criterion can be challenging even when relatively simple models are involved; see, for example, Dette et al. (2012, 2017b). Constructing efficient algorithms for finding $T$-optimal designs is likewise difficult in general, despite recent progress (Braess & Dette, 2013; Dette et al., 2015, 2017a; Aletti et al., 2016; Tommasi et al., 2016).

Recent advances in tackling discrimination design problems include the following. 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) and the class of models of interest now includes generalized linear models (Waterhouse et al., 2008). Methodologies are also available for finding a variety of optimal discriminating designs for multivariate dynamic models (Ucinski & Bogacka, 2005), Bayesian optimal designs for model discrimination (Felsenstein, 1992; Tommasi & López-Fidalgo, 2010; Dette et al., 2015), dual-objective optimal discrimination designs (Ng & Chick, 2004; Atkinson, 2008; Alberton et al., 2011; Abd El-Monsef & Seyam, 2011), optimal designs that discriminate between models with correlated errors (Campos-Barreiro & Lopez-Fidalgo, 2016) and adaptive designs for model discrimination (Myung & Pitt, 2009). References that describe alternative approaches and properties of optimal discrimination designs include López-Fidalgo et al. (2007), Dette & Titoff (2009) and Dette et al. (2015).

All references cited so far require a parametric conditional distribution of the response. This raises the question as to whether $T$-optimal discrimination designs are robust with respect to misspecification of this distribution. Some answers are provided by Wiens (2009), Ghosh & 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 but 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 & Lewis, 1991) and the resulting optimal designs are called semiparametric optimal discrimination designs.

### 2. Semiparametric Discrimination Designs

Following Kiefer (1974), we focus on approximate designs, which are probability measures defined on user-selected design space $X$. If an approximate design has $k$ support points at $x_1, \ldots, x_k$ with corresponding weights $\omega_1, \ldots, \omega_k$ and the total number of observations allowed for the study is $n$, then approximately $n\omega_i$ observations are taken at $x_1, \ldots, x_k$. In practice, each $n\omega_i$ is rounded to an integer $n_i$ so that $n_i$ observations are taken at $x_1, \ldots, x_k$ subject to $\sum_{i=1}^{k} n_i = n$. 
Let \( Y \) be the continuous response variable and let \( x \) denote a vector of explanatory variables 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 want to construct efficient designs for discriminating between two competing models. López-Fidalgo et al. (2007) assumed that there are two parametric densities, say \( f_j(y; x, \theta_j) \), where the parameter \( \theta_j \) varies in a compact parameter space \( \Theta_j \) (\( j = 1, 2 \)). To fix ideas, we ignore nuisance parameters which may be present in the models. The Kullback–Leibler divergence measures the discrepancy between the two densities and is given by

\[
I_{1,2}(x, f_1, f_2, \theta_1, \theta_2) = \int f_1(y; x, \theta_1) \log \frac{f_1(y; x, \theta_1)}{f_2(y; x, \theta_2)} \, dy. \tag{1}
\]

López-Fidalgo et al. (2007) assumed that the model \( f_1 \) is the true model with a fixed parameter vector \( \hat{\theta}_1 \) and call a design a local Kullback–Leibler-optimal discriminating design for the models \( f_1 \) and \( f_2 \) if it maximizes the criterion

\[
\text{KL}_{1,2}(\xi, \hat{\theta}_1) = \inf_{\theta_2 \in \Theta_2} \int_{\mathcal{X}} I_{1,2}(x, f_1, f_2, \theta_1, \theta_2) \, \xi(dx) \tag{2}
\]

over all designs on the design space \( \mathcal{X} \). Such a Kullback–Leibler-optimal design maximizes the power of the likelihood ratio test for the hypothesis

\[
H_0 : f(y; x) = f_2(y; x, \theta_2) \quad \text{versus} \quad H_1 : f(y; x) = f_1(y; x, \hat{\theta}_1)
\]

in the worst-case scenario when \( \theta_2 \in \Theta_2 \) (López-Fidalgo et al., 2007, p. 233). Otsu (2008) proposed a design criterion for discriminating between a parametric model defined by its density and another semiparametric model. The set-up is 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.

Specifically, suppose that the conditional mean of the density \( f_j(y; x, \theta_j) \) is

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

and its support set is

\[
S_{f_j, \theta_j, x} = \{ y : f_j(y; x, \theta_j) > 0 \} \quad (j = 1, 2). \tag{3}
\]

Further, let \( f_1(y; x, \hat{\theta}_1) \) be a parametric density with a fixed parameter \( \hat{\theta}_1 \). Define

\[
\mathcal{F}_{2,x, \theta_2} = \left\{ f_2 : \int f_2(y; x, \theta_2) \, dy = 1, \int y f_2(y; x, \theta_2) \, dy = \eta_2(x, \theta_2), \ S_{f_2, \theta_2, x} = S_{f_1, \eta_1, x} \right\},
\]

which is the class of all conditional densities at the point \( x \) with parameter \( \theta_2 \) 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 \( \theta_2 \) vary over all their possible values, i.e.,

\[
\mathcal{F}_2 = \bigcup_{x \in \mathcal{X}} \bigcup_{\theta_2 \in \Theta_2} \mathcal{F}_{2,x, \theta_2},
\]
and call a design $\xi^*$ semiparametric optimal for discriminating between the model $f_1(y;x, \tilde{\theta}_1)$ and models in the class $\mathcal{F}_2$ if it maximizes

$$K_1(\xi, \tilde{\theta}_1) = \inf_{\theta_2 \in \Theta_2} \int_{\mathcal{X}} \inf_{f_2 \in \mathcal{F}_2 \times \theta_2} I_{1,2}(x, f_1, f_2, \tilde{\theta}_1, \theta_2) \xi(dx)$$

(4)

among all approximate designs on $\mathcal{X}$. This is a local optimality criterion in the sense of Chernoff (1953), as it depends on the parameter $\tilde{\theta}_1$.

Another possibility is to fix the family of conditional densities for $f_2(y;x, \theta_2)$, where the form of $f_2$ is known apart from the values of $\theta_2$. Define

$$\mathcal{F}_{1,x,\tilde{\theta}_1} = \left\{ f_1 : \int f_1(y;x, \tilde{\theta}_1) dy = 1, \int yf_1(y;x, \tilde{\theta}_1) dy = \eta_1(x, \tilde{\theta}_1), S_{f_1, \tilde{\theta}_1,x} = S_{f_2, \theta_2,x} \right\},$$

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

$$\mathcal{F}_1 = \bigcup_{x \in \mathcal{X}} \mathcal{F}_{1,x,\tilde{\theta}_1}$$

and call a design $\xi^*$ locally semiparametric optimal for discriminating between the family of models $f_2(y;x, \theta_2)$ and the class $\mathcal{F}_1$ if it maximizes

$$K_2(\xi, \tilde{\theta}_1) = \inf_{\theta_2 \in \Theta_2} \int_{\mathcal{X}} \inf_{f_1 \in \mathcal{F}_{1,x,\tilde{\theta}_1}} I_{1,2}(x, f_1, f_2, \tilde{\theta}_1, \theta_2) \xi(dx)$$

(5)

among all approximate designs on $\mathcal{X}$. In the following discussion we refer to designs that maximize the criteria $K_1$ and $K_2$ as semiparametric Kullback–Leibler-optimal discriminating designs of type 1 and type 2, respectively.

We assume, for the sake of 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 $y, x, \theta_1$ and $\theta_2$, though these assumptions could be relaxed if necessary. In Theorem 3.1 of his paper, Otsu (2008) derived explicit forms for the two criteria. For criterion (4), he obtained

$$K_1(\xi, \tilde{\theta}_1) = \inf_{\theta_2 \in \Theta_2} \int_{\mathcal{X}} \left( \mu + 1 + \int \log \left[ -\mu - \lambda(y - \eta_2(x, \theta_2)) \right] f_1(y;x, \tilde{\theta}_1) dy \right) \xi(dx)$$

(6)

where the constants $\lambda$ and $\mu$ depend on $x, \tilde{\theta}_1$ and $\theta_2$ are roots of the system of equations

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

(7)

that satisfy the constraint $\mu + \lambda(y - \eta_2(x, \theta_2)) < 0$ for all $y \in S_{f_1, \tilde{\theta}_1,x}$. A similar result can be obtained for criterion (5) (Otsu, 2008, Theorem 3.2). Below we simplify Otsu’s approach, show that the inner optimization problems in (4) and (5) can be reduced to solving a single equation, and derive simpler expressions for criteria (4) and (5) that facilitate the computation of the semiparametric optimal discriminating designs.
THEOREM 1.

(i) Assume that for each \( x \in \mathcal{X} \) the support of the conditional density \( f_1(y; x, \tilde{\theta}_1) \) is an interval, i.e., \( S_{f_1,\tilde{\theta}_1,x} = [y_{x,\min}, y_{x,\max}] \), 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 \( \theta_2 \in \Theta_2 \), the equation

\[
\int \frac{f_1(y; x, \tilde{\theta}_1)}{1 + \lambda \{y - \eta_2(x, \theta_2)\}} \, dy = 1
\]

(8)

has a unique nonzero root \( \lambda(x, \tilde{\theta}_1, \theta_2) \) that satisfies

\[
-\frac{1}{y_{x,\max} - \eta_2(x, \theta_2)} < \lambda(x, \tilde{\theta}_1, \theta_2) < -\frac{1}{y_{x,\min} - \eta_2(x, \theta_2)}.
\]

(9)

Criterion (4) then takes the form

\[
K_1(\xi, \tilde{\theta}_1) = \inf_{\theta_2 \in \Theta_2} \int \int f_1(y; x, \tilde{\theta}_1) \log \frac{f_1(y; x, \tilde{\theta}_1)}{f_2^*(y; x, \theta_2)} \, dy \, \xi(dx)
\]

(10)

and the optimal density \( f_2^* \) in (4) is

\[
f_2^*(y; x, \theta_2) = \frac{f_1(y; x, \tilde{\theta}_1)}{1 + \lambda(x, \tilde{\theta}_1, \theta_2) \{y - \eta_2(x, \theta_2)\}}.
\]

(11)

(ii) Assume that the integrals

\[
\int f_2(y; x, \theta_2) \exp(-\lambda y) \, dy, \quad \int y f_2(y; x, \theta_2) \exp(-\lambda y) \, dy
\]

exist for all \( x \in \mathcal{X} \) and for all \( \lambda \). Criterion (5) takes the form

\[
K_2(\xi, \tilde{\theta}_1) = \inf_{\theta_2 \in \Theta_2} \int \int f_1^*(y; x, \tilde{\theta}_1) \log \frac{f_1^*(y; x, \tilde{\theta}_1)}{f_2(y; x, \theta_2)} \, dy \, \xi(dx)
\]

(12)

and the optimal density \( f_1^* \) in (5) is given by

\[
f_1^*(y; \tilde{\theta}_1) = \frac{f_2(y; x, \theta_2) \exp\{-\lambda(x, \tilde{\theta}_1, \theta_2) y\}}{\int f_2(y; x, \theta_2) \exp\{-\lambda(x, \tilde{\theta}_1, \theta_2) y\} \, dy},
\]

(13)

where \( \lambda_x = \lambda(x, \tilde{\theta}_1, \theta_2) \) is the nonzero root of the equation

\[
\frac{\int y f_2(y; x, \theta_2) \exp(-\lambda y) \, dy}{\int f_2(y; x, \theta_2) \exp(-\lambda y) \, dy} = \eta_1(x, \tilde{\theta}_1).
\]

(14)

The main implication of Theorem 1 is that we first solve equations (8) and (14) numerically for \( \lambda \). As this has to be done for several values of \( \theta_2 \) it is quite demanding, though not so computationally expensive as finding the solution of the two equations in (7) for Otsu’s approach.

For solving (8), it is natural to assume that \( \lambda < 0 \) if \( \eta_1(x, \tilde{\theta}_1) < \eta_2(x, \theta_2) \), because if \( y \in S_{f_1,\tilde{\theta}_1,x} \), the function \( 1/[1 + \lambda \{y - \eta_2(x, \theta_2)\}] \) is increasing and so allows us to shift the average of the
function \( f_1(y; x, \tilde{\theta}_1) /[1 + \lambda \{ y - \eta_2(x, \theta_2) \}] \) to the right. Similarly, if \( \eta_1(x, \tilde{\theta}_1) > \eta_2(x, \theta_2) \), we search for \( \lambda > 0 \). The following lemma formalizes this consideration and its proof, and all other proofs are deferred to the final section.

**Lemma 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 \( \tilde{\lambda} \) solves (8) and satisfies (9), \( \tilde{\lambda} \) has the same sign as the difference \( \eta_1(x, \tilde{\theta}_1) - \eta_2(x, \theta_2) \).

**Example 1.** Let \( f_1(y; x, \tilde{\theta}_1) \) be the truncated normal density \( N[\eta(x, \tilde{\theta}_1), 1] \) on the interval \([-3 + \eta_1(x, \tilde{\theta}_1), 3 + \eta_1(x, \tilde{\theta}_1)]\). This density is a function of \( \eta_1(x, \tilde{\theta}_1) \) and it follows from (11) that the optimal density \( f_2^\ast(y; x, \theta_2) \) is a function of \( \eta_1(x, \tilde{\theta}_1) \) and \( \eta_2(x, \theta_2) \). Figure 1 displays the function \( f_2^\ast \) for \( \eta_1(x, \tilde{\theta}_1) \equiv 0 \) and different values of \( \eta_2(x, \theta_2) \) on the interval \([-3, 3]\).

The main difference between our approach and that of Otsu (2008) is that we provide an easier and quicker way to compute the quantity

\[
\inf_{f_2 \in \Phi_{2, x, \theta_2}} I_{1, 2}(x, f_1, f_2, \tilde{\theta}_1, \theta_2).
\]  

(15)

This difference has very important implications for the numerical calculation of the semiparametric discrimination designs. To be precise, the result in Otsu (2008) requires us to solve the two nonlinear equations in (7) numerically for all design points \( x \) involved in the determination of the optimal design maximizing criterion (5) and all parameter values \( \theta_2 \in \Theta_2 \) involved in the minimization of the simplified version (6) derived by Otsu (2008). From a numerical viewpoint, it is very challenging to tackle this unstable problem because the solution depends sensitively on the specification of an initial point for the iterative procedure to solve (7). In contrast, Theorem 1 reduces the problem to the solution of one nonlinear equation, which can be found, for example, by a bisection search or a golden ratio search.

The numerical instability becomes apparent also in the numerical study in § 5, where we tried to compare the two methods in three examples. There we implemented Newton’s method to find the solution of the system of two equations in (7) required by Otsu’s method. We observed that for many values of the explanatory variable \( x \), the function in (15) could not be computed because the Newton method did not converge to the solution of system (7) that satisfies the condition \( \mu + \lambda \{ y - \eta_2(x, \theta_2) \} < 0 \). Such a problem was even observed in cases where we used a starting point in the iteration which is very close to the solution determined by the new method proposed in this paper. As a consequence, in many examples the semiparametric optimal discrimination
design could not be determined by the algorithm of Otsu (2008). Moreover, we observe that in the cases where Otsu’s method was able to determine the solution of the two nonlinear equations in (15), our method is still, on average, about two times faster; see Example 4.

3. Equivalence theorems

Equivalence theorems are useful because they confirm optimality of a design among all designs on the given design space $X$. These tools exist if the criterion is a convex or concave function over the set of all approximate designs on $X$, and their derivations are discussed in design monographs (Silvey, 1980; Pukelsheim, 2006). The next theorem states the equivalence results for the semiparametric Kullback–Leibler-optimal discriminating designs.

**Theorem 2.** Suppose that the conditions of Theorem 1 hold and the infimum in (4) and (5) is attained at a unique point $\theta_2^* \in \Theta_2$ for the optimal design $\xi^*$.

(a) A design $\xi^*$ is a semiparametric Kullback–Leibler-optimal discriminating design of type 1 if and only if

$$I_{1,2}(x,f_1,f_2^*,\bar{\theta}_1,\theta_2^*) - \int_X I_{1,2}(x,f_1,f_2^*,\bar{\theta}_1,\theta_2^*) \xi^*(dx) \leq 0, \quad x \in X, \quad (16)$$

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

$$\theta_2^* = \arg \inf_{\theta_2 \in \Theta} \int_X I_{1,2}(x,f_1,f_2^*,\bar{\theta}_1,\theta_2) \xi^*(dx), \quad f_2^*(y;x,\theta_2) = \frac{f_1(y;x,\bar{\theta}_1)}{1 + e^{\lambda} \{ y - \eta_2(x,\theta_2) \}},$$

and $\lambda$ is found from (8). Moreover, there is equality in (16) for all support points of $\xi^*$.

(b) A design $\xi^*$ is a semiparametric Kullback–Leibler-optimal discriminating design of type 2 if and only if

$$I_{1,2}(x,f_1^*,f_2,\bar{\theta}_1,\theta_2^*) - \int_X I_{1,2}(x,f_1^*,f_2,\bar{\theta}_1,\theta_2^*) \xi^*(dx) \leq 0, \quad x \in X, \quad (17)$$

with equality at the support points of $\xi^*$. Here

$$\theta_2^* = \arg \inf_{\theta_2 \in \Theta} \int_X I_{1,2}(x,f_1^*,f_2,\bar{\theta}_1,\theta_2) \xi^*(dx), \quad f_1^*(y;x,\bar{\theta}_1) = \frac{f_2(y;x,\theta_2) \exp(-\lambda y)}{\int f_2(y;x,\theta_2) \exp(-\lambda y) dy},$$

and $\lambda$ is found from (14). Moreover, there is equality in (17) for all support points of $\xi^*$.

Theorem 2 is a direct consequence of the equivalence theorem for Kullback–Leibler-optimal designs from López-Fidalgo et al. (2007). Part (a) states that $K_1(\xi,\bar{\theta}_1)$ is the Kullback–Leibler criterion for discrimination between $f_1(y;x,\bar{\theta}_1)$ and $f_2^*(y;x,\theta_2)$ defined in (11). Part (b) states that $K_2(\xi,\bar{\theta}_1)$ is the Kullback–Leibler criterion for discrimination between $f_1^*(y;x,\bar{\theta}_1)$ defined in (13) and $f_2(y;x,\theta_2)$. Following convention in the case where all models are parametric, we call the function on the left-hand side of (16) or (17) the sensitivity function of the design under investigation. Clearly, different design criteria lead to different sensitivity functions for the same design. The usefulness of the equivalence theorem is that if the sensitivity function of a design does not satisfy the conditions required in the equivalence theorem, then the design is not optimal under the given criterion. Figure 2 illustrates these sensitivity plots.
4. CONNECTIONS WITH THE $T$-OPTIMALITY CRITERION

We now show that under homoscedastic symmetrically distributed errors, the semiparametric optimal design for discriminating between the model $f_1(y; x, \bar{\theta}_1)$ and the class $\mathcal{F}_2$ coincides with the $T$-optimal design proposed by Atkinson & Fedorov (1975a). We first recall the classical set-up for finding an optimal design to discriminate between two models, where we assume that the mean functions in the models are known and the parameters in the null model are fixed at, say, $\bar{\theta}_1$. When errors in both models are normally distributed, a $T$-optimal design proposed by Atkinson & Fedorov (1975a). Throughout this section, we assume that the assumptions of Theorem 1(i) hold and $f_1(y; x, \bar{\theta}_1)$ satisfies

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

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 criterion (18) is a semiparametric Kullback–Leibler-optimal discriminating design of type 1.

A similar result is available for the semiparametric Kullback–Leibler-optimal discriminating designs of type 2. Suppose that $f_2(y; x, \theta_2)$ and $f_1(y; x, \bar{\theta}_1)$ are normal distributions $\mathcal{N}([\eta_2(x, \theta_2), v_2^2(x, \theta_2)])$ and $\mathcal{N}([\eta_1(x, \bar{\theta}_1), v_2^2(x, \theta_2)])$, respectively. It can be shown that the power of the likelihood ratio test for hypotheses (19) is an increasing function of

$$\int_{\mathcal{X}} \{\eta_1(x, \bar{\theta}_1) - \eta_2(x, \theta_2)\}^2 \xi(dx)$$

where $KL_{1,2}(\xi, \bar{\theta}_1)$ is the Kullback–Leibler criterion defined in (2). The next result shows that this design is also a semiparametric Kullback–Leibler-optimal discriminating design of type 2.

**Theorem 4.** Suppose that $f_2(y; x, \theta_2)$ is a normal density with mean $\eta_2(x, \theta_2)$ and variance $v_2^2(x, \theta_2).$ The best approximation $f_1^*(y; x, \bar{\theta}_1)$ is a normal density with mean $\eta_1(x, \bar{\theta}_1)$ and variance $v_2^2(x, \theta_2)$, and the optimal design maximizing (20) is a semiparametric Kullback–Leibler-optimal discriminating design of type 2 and vice versa.

5. NUMERICAL RESULTS

We now illustrate the new techniques for finding semiparametric optimal designs using three examples. From § 2, the first step is to solve equations (8) and (14) efficiently. In the second step,
any numerical method that determines Kullback–Leibler-optimal discrimination designs can be adapted to solve the minimax problems obtained from Theorem 1 because the representations (10) and (12) have the same structure as the Kullback–Leibler optimality criteria considered in López-Fidalgo et al. (2007). The second step defines a very challenging problem, and some recent results and algorithms for Kullback–Leibler optimality criteria can be found in Stegmaier et al. (2013), Braess & Dette (2013), Dette et al. (2015) and Dette et al. (2017a). Below we focus on the first step because our aim is to find new semiparametric designs. In the second step, we use an adaptation of the first-order algorithm of Atkinson & Fedorov (1975a), which is not the most efficient algorithm but is very easy to implement.

Let δ be a user-selected positive constant. By Lemma 1 and inequality (9), we solve equation (8) in the following regions: if \( \eta_1(x, \hat{\theta}_1) = \eta_2(x, \theta_2) \), set \( \lambda = 0 \); if \( \eta_1(x, \hat{\theta}_1) < \eta_2(x, \theta_2) \), choose a solution in the interval \( \Lambda^- = \left[ -1/(y_{x,\max} - \eta_2(x, \theta_2)), -\delta \right] \); if \( \eta_1(x, \hat{\theta}_1) > \eta_2(x, \theta_2) \), choose a solution in the interval \( \Lambda^+ = [\delta, -1/(y_{x,\min} - \eta_2(x, \theta_2))]. \)

Similarly, the solution of (14) can be obtained as follows. We search for \( \lambda > 0 \) if \( \eta_1(x, \hat{\theta}_1) < \eta_2(x, \theta_2) \) so that \( \lambda \) shifts the predefined density \( f_2(y; x, \theta_2) \) to the left, and search for \( \lambda < 0 \) if \( \eta_1(x, \hat{\theta}_1) > \eta_2(x, \theta_2) \). If \( \delta \) is chosen to be a small enough positive constant and \( \beta \) is a user-selected large positive constant, we can assume that the solution of (14) is in \([\beta, +\beta]\). We suggest searching for the numerical solution of equation (14) in the following regions: if \( \eta_1(x, \hat{\theta}_1) = \eta_2(x, \theta_2) \), set \( \lambda = 0 \); if \( \eta_1(x, \hat{\theta}_1) < \eta_2(x, \theta_2) \), choose a solution in the interval \( \Lambda^+ = [\delta, +\beta] \); if \( \eta_1(x, \hat{\theta}_1) > \eta_2(x, \theta_2) \), choose a solution in the interval \( \Lambda^- = [-\beta, -\delta] \).

We now present two examples, where the \( T \)-optimal and semiparametric Kullback–Leibler-optimal designs are determined numerically and are different.

**Example 2.** Consider the optimal design problem from López-Fidalgo et al. (2007), where they wanted to discriminate between the two models

\[
\eta_1(x, \theta_1) = \theta_{1,1} x + \frac{\theta_{1,2} x}{x + \theta_{1,3}}, \quad \eta_2(x, \theta_2) = \frac{\theta_{2,1} x}{x + \theta_{2,2}}.
\]

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 four different types of optimal discrimination design for this problem: a \( T \)-optimal design; a Kullback–Leibler-optimal design for lognormal errors, with fixed variances \( v_1^2(x, \hat{\theta}_1) = v_2^2(x, \theta_2) = 0.1 \); a semiparametric Kullback–Leibler-optimal discriminating design of type 1 for a mildly truncated lognormal density \( f_1(y; x, \hat{\theta}_1) \) with location \( \mu_1(x, \hat{\theta}_1) \) and scale \( \sigma_1^2(x, \theta_1) \); and a semiparametric Kullback–Leibler-optimal discriminating design of type 2 for a mildly truncated lognormal density \( f_2(y; x, \theta_2) \) with location \( \mu_2(x, \theta_2) \) and scale \( \sigma_2^2(x, \theta_2) \), where

\[
\mu_i(x, \theta) = \log \eta_i(x, \theta) - \frac{1}{2} \sigma_i^2(x, \theta) \quad \text{and} \quad \sigma_i^2(x, \theta) = \log \left( 1 + v_i^2(x, \theta)/\eta_i^2(x, \theta) \right) \quad (i = 1, 2).
\]

The ranges for those densities are the intervals from \( Q_1(0.0001, x, \hat{\theta}_1) \) to \( Q_1(0.9999, x, \hat{\theta}_1) \) and from \( Q_2(0.0001, x, \theta_2) \) to \( Q_2(0.9999, x, \theta_2) \) respectively, where \( Q_i(p, x, \theta) \) is the quantile function of the ordinary lognormal density with mean \( \eta_i(x, \theta) \) and variance \( v_i^2(x, \theta) = 0.1 \). We note that because of the mild truncation, \( \eta_1(x, \hat{\theta}_1) \) and \( \eta_2(x, \theta_2) \) are not exactly the means of the densities \( f_1(y; x, \hat{\theta}_1) \) and \( f_2(y; x, \theta_2) \), respectively, but are very close to them.

Table 1 displays the optimal discrimination designs under the four different criteria, along with the optimal parameter \( \theta_2^* \) of the second model corresponding to the minimal value with respect to the parameter \( \theta_2 \). All four types of optimal discrimination designs are different, with the smallest
Table 1. Optimal discrimination designs for the two models in (21)

| Design type          | $\xi^*$ | $\theta^*$ |
|----------------------|---------|------------|
| $T$-optimal          | 0.508   | 2.992      | 5.000     | (22, 564, 14, 637) |
| $KL$-optimal         | 0.580   | 0.298      | 0.122     |                      |
| $SKL_1$-optimal      | 0.218   | 2.859      | 5.000     | (21, 112, 13, 436)  |
| $SKL_2$-optimal      | 0.629   | 0.260      | 0.111     |                      |
|                      | 0.454   | 2.961      | 5.000     | (22, 045, 14, 197)  |
|                      | 0.531   | 0.344      | 0.125     |                      |
|                      | 0.509   | 2.994      | 5.000     | (22, 824, 14, 857)  |
|                      | 0.611   | 0.273      | 0.116     |                      |

$KL$, Kullback–Leibler; $SKL_i$, semiparametric Kullback–Leibler of type $i$.

Fig. 2. Plots of the sensitivity functions of the following discrimination designs: (a) $T$-optimal, (b) Kullback–Leibler-optimal, (c) semiparametric Kullback–Leibler-optimal of type 1 and (d) semiparametric Kullback–Leibler-optimal of type 2, from Table 1.

Table 2. Efficiencies of optimal discrimination designs for the two models in (21) under various optimality criteria. For example, the value 0.321 in the first row is the efficiency of the Kullback–Leibler-optimal design with respect to the $T$-optimality criterion

|               | $T$-optimal | $KL$-optimal | $SKL_1$-optimal | $SKL_2$-optimal |
|---------------|-------------|--------------|-----------------|-----------------|
| $T$-criterion | 1.000       | 0.321        | 0.741           | 0.830           |
| $KL$-criterion| 0.739       | 1.000        | 0.796           | 0.650           |
| $SKL_1$-criterion | 0.552   | 0.544        | 1.000           | 0.454           |
| $SKL_2$-criterion | 0.876   | 0.254        | 0.633           | 1.000           |

$KL$, Kullback–Leibler; $SKL_i$, semiparametric Kullback–Leibler of type $i$.

support point of the Kullback–Leibler-optimal design being noticeably different from those of the other three designs. The semiparametric Kullback–Leibler-optimal discriminating design of type 2 has nearly the same support as the $T$-optimal design. Figure 2 shows the sensitivity functions of the four optimal designs and confirms their optimality.

Table 2 displays the four different types of efficiencies of the $T$-, Kullback–Leibler-, and semiparametric Kullback–Leibler-optimal discriminating designs. Small changes in the design can have large effects, and the $T$- and Kullback–Leibler-optimal discrimination designs are not very robust under a variation of the criteria, where the Kullback–Leibler-optimal discrimination design has slight advantages. On the other hand, the semiparametric Kullback–Leibler-optimal discriminating design of type 1 yields moderate efficiencies, about 75%, with respect to the $T$- and Kullback–Leibler optimality criteria.
**Table 3. Optimal discrimination designs for the two models in (22)**

| Design type  | $x$  | $w$  |
|--------------|-----|-----|
| $T$-optimal  | 0.308 | 2.044 | 5.000 | (1.223, 0.948) |
| $KL$-optimal | 0.316 | 0.428 | 0.256 | (1.244, 1.020) |
| $SKL_1$-optimal | 0.136 | 1.902 | 5.000 | (1.216, 0.920) |
| $SKL_2$-optimal | 0.297 | 0.457 | 0.252 | (1.225, 0.956) |

$KL$, Kullback–Leibler; $SKL_i$, semiparametric Kullback–Leibler of type $i$.

**Example 3.** Consider a similar problem with a function $\eta_1(x, \theta_1)$ taken from Wiens (2009). The two models of interest are

$$\eta_1(x, \theta_1) = \theta_{1,1} \{1 - \exp(-\theta_{1,2} x)\}, \quad \eta_2(x, \theta_2) = \frac{\theta_{2,1} x}{\theta_{2,2} + x}, \quad (22)$$

where the design space is $\mathcal{X} = [0, 1, 5]$. Here we fix the parameters of the first model in (22) to $\tilde{\theta}_1 = (1, 1)$ and determine the $T$-optimal, Kullback–Leibler-optimal for lognormal errors, and semiparametric Kullback–Leibler-optimal discriminating designs of type 1 and type 2 for mildly truncated lognormal errors. The error variances for the Kullback–Leibler-optimal discrimination design are $v_1^2(x, \tilde{\theta}_1) = v_2^2(x, \theta_2) = 0.02$; for the semiparametric Kullback–Leibler-optimal discriminating design of type 1 the variance is $v_1^2(x, \tilde{\theta}_1) = 0.02$, and for the semiparametric Kullback–Leibler-optimal discriminating design of type 2 the variance is $v_2^2(x, \theta_2) = 0.02$.

Table 3 displays the various optimal designs, along with the minimal values of the parameters $\theta_2^*$ and $\theta_2^*$ in the second model sought in the criterion. The optimality of the numerically determined $T$-optimal, Kullback–Leibler-optimal and semiparametric Kullback–Leibler-optimal discriminating designs of type 1 and type 2 can be verified by plotting the corresponding sensitivity functions. We again observe substantial differences between the optimal discrimination designs with respect to the different criteria. 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. In particular, we note that for our two examples, the other optimal discrimination designs are especially sensitive to the Kullback–Leibler optimality criteria; in the first example their Kullback–Leibler efficiencies are at best 54%, and in the second example their Kullback–Leibler efficiencies are not higher than 40%. One reason may be that the smallest support point of the Kullback–Leibler-optimal design is noticeably smaller than the minimum support point of each of the other three optimal designs.

**Example 4.** It is difficult to compare our approach with that of Otsu (2008) because of the computational difficulties described at the end of §2. In particular, the latter algorithm is often not able to determine the semiparametric Kullback–Leibler-optimal discrimination design. For instance, in the situations considered in Examples 2 and 3 we were unable to obtain convergence. For a comparison of the speed of the methods we therefore have to choose a relatively simple example for which a comparison of both methods is possible. For this purpose we again considered models (22) and constructed the semiparametric Kullback–Leibler-optimal discriminating designs of type 1. For the density $f_1(y; x, \tilde{\theta}_1)$ we used the density of the random variable.
6. Conclusions

Much of the present work on optimal design for discriminating between models assumes that the models are fully parametric. Our work allows the alternative models to be nonparametric, where only their mean functions have to be specified, apart from the parameter values. Our approach is simpler and more reliable than other approaches in the literature for tackling such challenging and more realistic optimal discrimination design problems.

We expect potential applications of our work to systems biology, where frequently the underlying model generating the responses is unknown and very complex. In practice, the mean response is approximated in a few ways and these approximations become the conditional means of nonparametric models that need to be efficiently discriminated to arrive at a plausible model. The optimal design method presented here will save costs by helping biological researchers to efficiently determine an adequate mean model among several postulated. There are also rich opportunities for further methodological research. For example, an important problem is to relax the assumption that the set $S_{\eta_1, \theta_1}$ defined in (3) is fixed for each $x$, so that the method can be applied to a broader class of conditional densities.

Acknowledgement

We are grateful to the reviewers for their constructive comments on the first version of our paper. Dette and Guchenko were supported by the Deutsche Forschungsgemeinschaft. Dette and Wong were partially supported by the National Institute of General Medical Sciences of the U.S. National Institutes of Health. Melas and Guchenko were partially supported by St. Petersburg State University and the Russian Foundation for Basic Research.

Supplementary Material

Supplementary material available at Biometrika online contains all proofs.
REFERENCES

ABD EL-MONSEF, M. M. E. & SEYAM, M. M. (2011). CDT-optimum designs for model discrimination, parameter estimation and estimation of a parametric function. *J. Statist. Plan. Infer.* 141, 639–43.

ALBERTON, A. L., SWAOBA, M., LABAO, M. W. N. & PINTO, J. C. (2011). Experimental design for the joint model discrimination and precise parameter estimation through information measures. *Chem. Eng. Sci.* 66, 1940–52.

ALETTI, G., MAY, C. & TOMMACI, C. (2016). KL-optimum designs: Theoretical properties and practical computation. *Statist. Comp.* 26, 107–17.

ATKINSON, A. C. (2008). DT-optimum designs for model discrimination and parameter estimation. *J. Statist. Plan. Infer.* 138, 56–64.

ATKINSON, A. C. & FEDOROV, V. V. (1975a). The designs of experiments for discriminating between two rival models. *Biometrika* 62, 57–70.

ATKINSON, A. C. & FEDOROV, V. V. (1975b). Optimal design: Experiments for discriminating between several models. *Biometrika* 62, 289–303.

BORWEIN, J. M. & LEWIS, A. S. (1991). Duality relationships for entropy-like minimization problems. *SIAM J. Contr. Optimiz.* 29, 325–38.

BRAESS, D. & DETTE, H. (2013). Optimal discriminating designs for several competing regression models. *Ann. Statist.* 41, 897–922.

CAMPOS-BARREIRO, S. & LOPEZ-FIDALGO, J. (2016). KL-optimal experimental design for discriminating between two growth models applied to a beef farm. *Math. Biosci. Eng.* 13, 67–82.

CHERNOFF, H. (1953). Locally optimal designs for estimating parameters. *Ann. Math. Statist.* 24, 586–602.

COVAGNARO, D. R., MYUNG, J. I., PIT, M. A. & KUJALA, J. V. (2010). Adaptive design optimization: A mutual information-based approach to model discrimination in cognitive science. *Neural Comp.* 22, 887–905.

DEHTE, H., GUCHENKO, R. & MELAS, V. B. (2017a). Efficient computation of Bayesian optimal discriminating designs. *J. Comp. Graph. Statist.* 26, 424–33.

DEHTE, H., MELAS, V. B. & GUCHENKO, R. (2015). Bayesian T-optimal discriminating designs. *Ann. Statist.* 43, 1959–85.

DEHTE, H., MELAS, V. B. & SHIPLEY, P. (2012). T-optimal designs for discrimination between two polynomial models. *Ann. Statist.* 40, 188–205.

DEHTE, H., MELAS, V. B. & SHIPLEY, P. (2013). Robust T-optimal discriminating designs. *Ann. Statist.* 41, 1693–715.

DEHTE, H., MELAS, V. B. & SHIPLEY, P. (2017b). T-optimal discriminating designs for Fourier regression models. *Comp. Statist. Data Anal.* 113, 196–206.

DEHTE, H. & TITOFF, S. (2009). Optimal discrimination designs. *Ann. Statist.* 37, 2056–82.

FESELENESTEIN, K. (1992). Optimal Bayesian design for discrimination among rival models. *Comp. Statist. Data Anal.* 14, 427–36.

GHOSH, S. & DUTTA, S. (2013). Robustness of designs for model discrimination. *J. Mult. Anal.* 115, 193–203.

JAMSIN, K. M., DUFFULL, S. B., TARNING, J., PRICE, R. N. & SIMPSON, J. (2013). A robust design for identification of the parasite clearance estimator. *Malaria J.* 12, 410–6.

KIEFER, J. (1974). General equivalence theory for optimum designs (approximate theory). *Ann. Statist.* 2, 849–79.

LOPEZ-FIDALGO, J., TOMMASI, C. & TRANDAFIR, P. C. (2007). An optimal experimental design criterion for discriminating between non-normal models. *J. R. Statist. Soc.* B 69, 231–42.

MYUNG, J. I. & PIT, M. A. (2009). Optimal experimental design for model discrimination. *Psychol. Rev.* 116, 499–518.

NG, S. H. & CHICK, S. E. (2004). Design of follow-up experiments for improving model discrimination and parameter estimation. *Naval Res. Logist.* 2, 1–11.

OTSU, T. (2008). Optimal experimental design criterion for discriminating semi-parametric models. *J. Statist. Plan. Infer.* 138, 4141–50.

PUKELSHEIM, F. (2006). *Optimal Design of Experiments*. Philadelphia: SIAM.

SILVEY, S. (1980). *Optimal Design*. London: Chapman & Hall.

STEIGMAIER, J., SKANDA, D. & LEBIEDZ, D. (2013). Robust optimal design of experiments for model discrimination using an interactive software tool. *PLoS ONE* 8, e55723, doi:10.1371/journal.pone.0055723.

TOMMASI, C. & LOPEZ-FIDALGO, J. (2010). Bayesian optimum designs for discriminating between models with any distribution. *Comp. Statist. Data Anal.* 54, 143–50.

TOMMASI, C., MARTIN-MARTIN, R. & LOPEZ-FIDALGO, J. (2016). Max-min optimal discriminating designs for several statistical models. *Statist. Comp.* 26, 1163–72.

UCNISKO, D. & BOGACKA, B. (2005). T-optimum designs for discrimination between two multiresponse dynamic models. *J. R. Statist. Soc.* B 67, 3–18.

WATERHOUSE, T. H., WOODS, D. C., ECCLESTON, J. A. & LEWIS, S. M. (2008). Design selection criteria for discrimination/estimation for nested models and a binomial response. *J. Statist. Plan. Infer.* 138, 132–44.

WIENS, D. P. (2009). Robust discrimination designs. *J. R. Statist. Soc.* B 71, 805–29.

[Received on 25 November 2016. Editorial decision on 12 August 2017]
