© Grace Scientific Publishing

Journal of
Statistical Theory and Practice
Volume 3, No. 2, June 2009

Downloaded by [University of Alberta], [Doug Wiens] at 09:54 30 November 2011

Asymptotic Properties of a Neyman-Pearson Test for Model Discrimination, with an Application to Experimental Design

Douglas P. Wiens, *Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alberta, Canada.*
Email: doug.wiens@ualberta.ca

Received: January 15, 2008 *Revised:* April 28, 2008

Abstract

We present and derive the asymptotic properties of a certain Neyman-Pearson test, relevant for problems concerning discrimination between two competing models. We then study an application in which this test is used to assess the efficacy of designs, the purpose of which is to aid in discriminating between two nonlinear regression models.

AMS Subject Classification: Primary: 62F03; 62F05. Secondary: 62K05; 62J02.

Key-words: Asymptotics; Design; Discrimination; DT-optimality; Kullback-Leibler divergence; Michaelis-Menten; T-optimality; KL-optimality.

1. Introduction

The purpose of this article is to present and derive the asymptotic properties of a certain Neyman-Pearson test, relevant for problems concerning discrimination between two competing models. We then study an application in which this test is used to assess the efficacy of designs, the purpose of which is to aid in such discrimination problems.

The problem of discrimination can be viewed as one of choosing between two hypotheses. We suppose that, under the null hypothesis H_0, the data at hand have arisen from a density $f_0(y|\mathbf{x}, \mu_0, \tau_0)$. Under the alternate hypothesis H_1 the density is $f_1(y|\mathbf{x}, \mu_1, \tau_1)$. The data are values of a random variable Y, together with covariates $\mathbf{x}$, possibly chosen by design. Under Model j – specified by H_j – the mean conditional response is $\mu_j(\mathbf{x})$, which depends upon the covariates through a, linear or nonlinear, regression response. The stochastic component

* 1559-8608/09-2/$5 + $1pp – see inside front cover
© Grace Scientific Publishing, LLC

Downloaded by [University of Alberta], [Doug Wiens] at 09:54 30 November 2011

is specified by the density f_j. The remaining terms τ_j represent possibly vector-valued nuisance parameters. Where there is no possibility of confusion, τ_j will not be explicitly mentioned.

Given a finite design space $\mathcal{S} = \{\mathbf{x}_i\}_{i=1}^N$ with $n_i \geq 0$ observations $\{y_{ij}\}_{j=1}^{n_i}$ made at $\mathbf{x}_i$, and if the parameters are completely specified under each hypothesis, the Neyman - Pearson test of H_0 vs. H_1 rejects for large values of $R = \sum_{i,j} R_{ij}$, where

$$R_{ij} = 2\log\left\{\frac{f_1(y_{ij}|\mathbf{x}_i,\mu_1)}{f_0(y_{ij}|\mathbf{x}_i,\mu_0)}\right\}.$$

Under the large-sample approximation to the distribution of R derived in §3 of this article and presented as Theorem 1.1, the power of the test is maximized by the design maximizing

$$E_{H_1}[R] = 2n\int_{\mathcal{S}} \mathcal{I}(\mu_0(\mathbf{x}),\mu_1(\mathbf{x}))\xi(d\mathbf{x}), \tag{1.1}$$

where $n = \sum_{i=1}^N n_i$, ξ is the design measure placing mass n_i/n at $\mathbf{x}_i$, and

$$\mathcal{I}(\mu_0(\mathbf{x}),\mu_1(\mathbf{x})) = \int_{-\infty}^{\infty} f_1(y|\mathbf{x},\mu_1)\log\left\{\frac{f_1(y|\mathbf{x},\mu_1)}{f_0(y|\mathbf{x},\mu_0)}\right\}dy$$

is the Kullback-Leibler divergence, measuring the information which is lost when f_0 is used to approximate f_1. This is the basis for the design problem considered in §2.

In the proof of Theorem 1.1 we assume the usual regularity conditions for likelihood estimation - these are stated in §3 - and we consider contiguous alternatives.

Theorem 1.1. *Suppose that the densities f_0, f_1 are the same, i.e. $f_j(y|\mathbf{x},\mu_j,\tau_j) = f(y|\mathbf{x},\mu_j,\tau)$ for a density f, and that $\mu_1(\mathbf{x}_i) = \mu_0(\mathbf{x}_i) + n^{-1/2}\Delta_i$, $i = 1,\ldots,N$. Define*

$$D = \int_{\mathcal{S}} \mathcal{I}(\mu_0(\mathbf{x}),\mu_1(\mathbf{x}))\xi(d\mathbf{x}) = \sum_{i=1}^N \mathcal{I}(\mu_0(\mathbf{x}_i),\mu_1(\mathbf{x}_i))\xi_i. \tag{1.2}$$

Then under the regularity conditions given in §3 we have the following.

(i) *Under this sequence of contiguous alternatives D is $O(n^{-1})$ and R is asymptotically normally distributed under each hypothesis:*

$$\text{under } H_0,\ \frac{R+2nD}{\sqrt{8nD}} \xrightarrow{L} N(0,1); \tag{1.3}$$

$$\text{under } H_1,\ \frac{R-2nD}{\sqrt{8nD}} \xrightarrow{L} N(0,1). \tag{1.4}$$

Thus a test with asymptotic size α test rejects for $R > z_\alpha\sqrt{8nD} - 2nD$ and has asymptotic power

$$\beta = \Phi(\sqrt{2nD} - z_\alpha), \tag{1.5}$$

where Φ is the $N(0,1)$ distribution function and $z_\alpha = \Phi^{-1}(1-\alpha)$.

(ii) *If f is the Normal density with variance σ^2 then $2nD = \sum_{i=1}^N \Delta_i^2\xi_i/\sigma^2$ and these asymptotic distributions are exact, for all n.*

2. Examples

Suppose that one is to design an experiment, the purpose of which is to determine which of two regression models is appropriate for the data. Under the null model the mean response is of the Michaelis-Menten form

$$\mu_0(x) = \eta_0(x|\boldsymbol{\theta}_0) = \frac{V_0 x}{K_0 + x}, \text{ with } \boldsymbol{\theta}_0 = (V_0, K_0)^T;$$

under the alternate it is exponential:

$$\mu_1(x) = \eta_1(x|\boldsymbol{\theta}_1) = V_1(1 - e^{-K_1 x}), \text{ with } \boldsymbol{\theta}_1 = (V_1, K_1)^T.$$

An optimal design will maximize (1.2).

If the data are normally distributed under each model, and equally varied, then

$$2\sigma^2 D = \sum_{i=1}^{N} \{\eta_1(\mathbf{x}_i|\boldsymbol{\theta}_1) - \eta_0(\mathbf{x}_i|\boldsymbol{\theta}_0)\}^2 \xi_i. \tag{2.1}$$

For this case Hunter and Reiner (1965) proposed a sequential method to construct the design: after n observations have been made and estimates $\hat{\boldsymbol{\theta}}_j$ computed, the next design point $\mathbf{x}_{n+1}$ should maximize $\{\eta_1(\mathbf{x}|\hat{\boldsymbol{\theta}}_1) - \eta_0(\mathbf{x}|\hat{\boldsymbol{\theta}}_0)\}^2$. Fedorov and Pazman (1968) extended this approach to heteroscedastic models. If instead the data follow a log-normal distribution, with a common coefficient of variation cv, then

$$2\log(1 + cv^2) D = \sum_{i=1}^{N} \{\log \eta_1(\mathbf{x}_i|\boldsymbol{\theta}_1) - \log \eta_0(\mathbf{x}_i|\boldsymbol{\theta}_0)\}^2 \xi_i; \tag{2.2}$$

some details are in López-Fidalgo, Tommasi and Trandafir (2007). Again a sequential approach is plausible.

It is our purpose here to construct static, i.e. nonsequential, designs. For this, one can clearly not rely on estimates of the parameters; in this case Fedorov (1975) suggested the *maximin* procedure of maximizing (1.1) after first minimizing over $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$. Atkinson and Fedorov (1975a) - see also Atkinson and Fedorov (1975b) - assumed that Model 1 was known to be the correct one, that $\boldsymbol{\theta}_1$ was known, and constructed designs, termed *T-optimal* designs, maximizing $\inf_{\boldsymbol{\theta}_0} \int \{\eta_1(\mathbf{x}|\boldsymbol{\theta}_1) - \eta_0(\mathbf{x}|\boldsymbol{\theta}_0)\}^2 \xi(d\mathbf{x})$. López-Fidalgo, Tommasi and Trandafir (2007) studied extensions of these notions to non-normal models, leading to the maximization of $\inf_{\boldsymbol{\theta}_0} \int \mathcal{I}(\eta_0(\mathbf{x}|\boldsymbol{\theta}_0), \eta_1(\mathbf{x}|\boldsymbol{\theta}_1))\xi(d\mathbf{x})$; this criterion is termed *KL-optimality*.

We instead take the following approach. We first introduce a "working response" $E[Y|\mathbf{x}]$; this can be constructed in several ways. In Example 2.1 we take $E[Y|\mathbf{x}] = \eta_1(\mathbf{x}|\boldsymbol{\theta}_1)$ for $\boldsymbol{\theta}_1 = (1,1)^T$. In Example 2.2 we take $E[Y|x] = .35 + .12x$, which is a linear response approximating $\eta_0(x|\boldsymbol{\theta}_0)$ when $\boldsymbol{\theta}_0 = (1,1)^T$. Another possibility, as mentioned by Atkinson and Fedorov (1975a), is to take $E[Y|x] = \int \eta_1(x|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}$ for a prior $p(\boldsymbol{\theta})$.

Once a working response is chosen, we *define*

$$\boldsymbol{\theta}_j = \arg\min_{\boldsymbol{\theta}} \sum_{\mathcal{S}} \{E[Y|\mathbf{x}_i] - \eta_j(\mathbf{x}_i|\boldsymbol{\theta})\}^2, \tag{2.3}$$

i.e. $\boldsymbol{\theta}_j$ provides the closest agreement, in this L^2 sense, between the true regression response and its approximation in Model j. We then seek a design $\boldsymbol{\xi}^*$maximizing (1.2).

Downloaded by [University of Alberta], [Doug Wiens] at 09:54 30 November 2011

Downloaded by [University of Alberta], [Doug Wiens] at 09:54 30 November 2011

Example 2.1. Here we consider normal data, and take $E[Y|\mathbf{x}] = \eta_1(\mathbf{x}|\boldsymbol{\theta}_1)$ for $\boldsymbol{\theta}_1 = (1,1)^T$. We take $x \in \mathcal{S} = .1(.1)5$; thus $N = 50$. Then from (2.3), $\boldsymbol{\theta}_0 = (1.22, .91)^T$. See Figure 2.1. If one-point designs are allowed, then the optimal choice $\boldsymbol{\xi}^*$ maximizing (2.1) places all mass at

$$\arg\max_{\mathcal{S}} |\eta_1(\mathbf{x}|\boldsymbol{\theta}_1) - \eta_0(\mathbf{x}|\boldsymbol{\theta}_0)| = .3.$$

The resulting powers (1.5) of level $\alpha = .1$ tests of $\eta_0(\mathbf{x}|\boldsymbol{\theta}_0)$ against $\eta_1(\mathbf{x}|\boldsymbol{\theta}_1)$, evaluated with $n = 20$ at a range of values of σ^2, are

$$\begin{pmatrix} \sigma^2: & 1 & .5 & .1 & .01 \\ \beta: & .14 & .15 & .25 & .73 \end{pmatrix}.$$

In this case even the power attained by this best possible design is disappointingly low, unless σ^2 is quite small. Most other values of $\boldsymbol{\theta}_1$ which we applied resulted in substantially larger powers.

One-point designs are of course poor if parameter estimation is also a goal. In such cases we dictate a minimum number q of support points and use a simulated annealing algorithm (see, e.g., Fang and Wiens, 2000) to obtain the designs. For instance if $q = 4$ - allowing for estimation of $\boldsymbol{\theta}$ and σ^2 - and $\mathcal{S} = .1(.175)5$, then when $n = 20$ we find that

$$\boldsymbol{\xi}^* = \begin{pmatrix} .275 & .45 & 4.825 & 5.0 \\ .85 & .05 & .05 & .05 \end{pmatrix}. \tag{2.4}$$

There is a negligible deterioration in the powers, which in this case are

$$\begin{pmatrix} \sigma^2: & 1 & .5 & .1 & .01 \\ \beta: & .14 & .15 & .24 & .71 \end{pmatrix}.$$

Although the powers were variable - and typically higher - we found little change in these designs upon experimenting with different inputs $\boldsymbol{\theta}_1$. Under normality our design criterion is quite similar to that of T-optimality, which requires the maximization of (2.1). A difference is in our treatment of $\boldsymbol{\theta}_0$, which we define through (2.3) with (in this example) $E[Y|x] = \eta_1(x|\boldsymbol{\theta}_1)$ but which, in the classical approach, is *design-dependent* and defined as the minimizer of $\sum_{i=1}^n \{\eta_1(x_i|\boldsymbol{\theta}_1) - \eta_0(x_i|\boldsymbol{\theta})\}^2 \xi(x_i)$. We were motivated to define the parameters as we did partially by robustness considerations. The assumptions that $\mu_j(x) = \eta_j(x|\boldsymbol{\theta}_j)$ might hold only approximately, in which case the parameters can be unidentifiable. To avoid this, the definition (2.3) is particularly convenient. In Wiens (2009) the means $\mu_j(x)$ are taken to be unknown members of certain neighbourhoods of the $\eta_j(x|\boldsymbol{\theta}_j)$ and the robustness of the discrimination designs is investigated.

Our emphasis on exact designs, in finite design spaces, has led to the use of annealing algorithms as in Example 2.1. There is a further consequence. Some authors - Atkinson and Fedorov (1975a), López-Fidalgo, Tommasi and Trandafir (2007) - have exploited convex design theory to prove equivalence theorems which can then be used to establish the optimality of designs. In our case the various restrictions imposed on the designs render the class of such designs non-convex, so that analogous results cannot be expected.

Example 2.2. Here we consider lognormal data, so that (2.2) is to be maximized. We take $n = 20$ and $E[Y|x] = .35 + .12x$, and find that when $\mathcal{S} = .1(.1)5$ the parameters defined

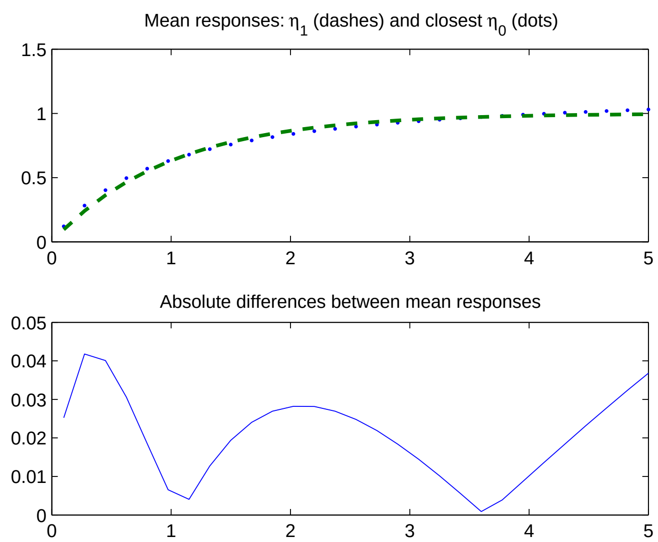

Figure 2.1. Mean reponses for Example 2.1.

by (2.3) are

$$\boldsymbol{\theta}_0 = (1.02, 1.13)^T, \ \boldsymbol{\theta}_1 = (.85, .73)^T.$$

See Figure 2.2. With $q = 1$ the design $\boldsymbol{\xi}^*$ concentrates all mass at

$$\arg\max_{\mathcal{S}} |\log \eta_1(\mathbf{x}|\boldsymbol{\theta}_1) - \log \eta_0(\mathbf{x}|\boldsymbol{\theta}_0)| = .1,$$

resulting in powers

$$\begin{pmatrix} cv^2: & 1 & .5 & .1 & .01 \\ \beta: & .69 & .85 & 1.00 & 1.00 \end{pmatrix}.$$

When $q = 4$ and $\mathcal{S} = .1(.175)5$ we find that

$$\boldsymbol{\xi}^* = \begin{pmatrix} .1 & .275 & .45 & .625 \\ .85 & .05 & .05 & .05 \end{pmatrix}, \qquad (2.5)$$

with powers

$$\begin{pmatrix} cv^2: & 1 & .5 & .1 & .01 \\ \beta: & .64 & .81 & 1.00 & 1.00 \end{pmatrix}.$$

Even with the imposition of the requirement $q > 1$, the designs presented here are poor for estimation or prediction from the chosen model. This seems intuitively clear, and can be quantified by the D-efficiencies

$$\text{eff} = \left(\frac{|\mathbf{M}(\boldsymbol{\xi}^*)|}{\max_{\xi \in \Xi_{n,q}} |\mathbf{M}(\boldsymbol{\xi})|} \right)^{\frac{1}{p}},$$

Downloaded by [University of Alberta], [Doug Wiens] at 09:54 30 November 2011

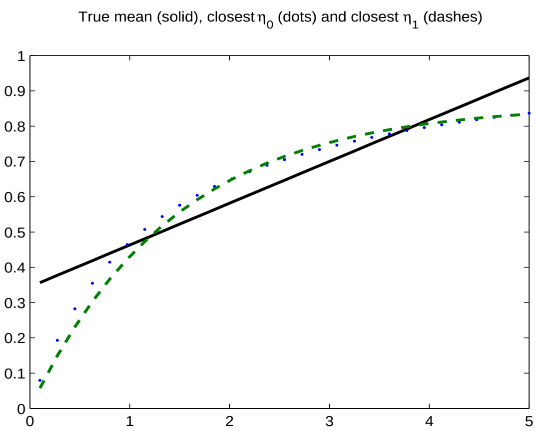

Figure 2.2. Hypothesized responses $\eta_j\left(x,\boldsymbol{\theta}_j\right)$ for Example 2.2, with $E\left[Y|x\right]$ linear.

where $|\mathbf{M}(\boldsymbol{\xi})|$ denotes the determinant of the information matrix of the design $\boldsymbol{\xi}$, $\Xi_{n,q}$ is the class of q-point exact designs on $\mathcal{S}$ using the specified sample size, and p is the number of regression parameters in the model being considered. For the design (2.4) we compute a D-efficiency of .483 for Model 0 and .355 for Model 1. The locally D-optimal designs in these cases were again found by simulated annealing and are

$$\begin{pmatrix} .625 & .80 & 4.825 & 5.0 \\ .45 & .05 & .05 & .45 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} .975 & 1.15 & 4.825 & 5.0 \\ .45 & .05 & .05 & .45 \end{pmatrix},$$

respectively. For (2.5) these D-efficiencies are .205 and .106 for Model 0 and Model 1 respectively; for both models the locally D-optimal design is

$$\begin{pmatrix} .1 & 4.625 & 4.825 & 5.0 \\ .50 & .05 & .05 & .40 \end{pmatrix}.$$

This inefficiency, for estimation or prediction, of discrimination designs was noted by Hill, Hunter and Wichern (1968), who proposed a sequential method to maximize a convex combination of the discriminatory power and a measure of estimation efficiency such as a weighted average of the determinants of the information matrices in the two models. Atkinson (2008) has proposed the maximization of a convex combination of the logarithm of $\mathcal{D}(\boldsymbol{\xi}|\mathbf{0},\mathbf{0})$, as at (2.1), and the logarithm of the determinant of the information matrix under Model 0. This criterion, combining as it does classical D- and T-optimality criteria, is termed DT-optimality. Wiens (2009) has proposed, in a robustness context, a sequential method which shifts emphasis from maximization of (1.2) towards minimization of mean squared error of the predictions, as evidence accrues in favour of one of the two models.

3. Proof of Theorem 1.1

In proving Theorem 1.1 we will make use of the following assumptions.

(A1) The density $f(y|\mathbf{x},\mu)$ is three times differentiable with respect to μ, and the derivative of $\int f(y|\mathbf{x},\mu)\,dy$ can be obtained by differentiating under the integral sign.

Downloaded by [University of Alberta], [Doug Wiens] at 09:54 30 November 2011

Downloaded by [University of Alberta], [Doug Wiens] at 09:54 30 November 2011

(A2) With

$$\psi(y|\mathbf{x}_i,\mu_0(\mathbf{x}_i)) \stackrel{def}{=} \frac{\partial}{\partial\mu}\log f(y|\mathbf{x}_i,\mu)_{|\mu=\mu_0(\mathbf{x}_i)},$$

(and with dots denoting differentiation with respect to μ), there exists a function $M(y|\mu_0(\mathbf{x}))$ and a number $c>0$ such that

$$\sup_{|s|\le c}|\ddot{\psi}(y|\mathbf{x},\mu_0(\mathbf{x})+s)| \quad \le M(y|\mu_0(\mathbf{x})),$$

with

$$\int_{-\infty}^{\infty} M(y|\mathbf{x},\mu_0(\mathbf{x}))\,f(y|\mathbf{x},\mu_0(\mathbf{x}))\,dy \quad <\infty \text{ for } \mathbf{x}\in\mathcal{S}.$$

(A3) The random variables $\dot{\psi}(Y|\mathbf{x},\mu_0(\mathbf{x}))$ have second moments which, under either hypothesis, are uniformly bounded for $\mathbf{x}\in\mathcal{S}$.

(A4) The random variables $\psi(Y|\mathbf{x},\mu_0(\mathbf{x}))$ have third absolute moments which, under either hypothesis, are uniformly bounded for $\mathbf{x}\in\mathcal{S}$.

Remark 3.1. Since $\mathcal{S}$ is finite, (A3) and (A4) require only that the specified moments be finite for each $\mathbf{x}\in\mathcal{S}$. Assumptions (A1)-(A4) are easily seen to hold for the examples of Section 2.

Proof. (i) For some $t_i\in[0,1]$ we have the expansion

$$\begin{aligned}
R_{ij} &= 2\log\left\{\frac{f_1(y_{ij}|\mathbf{x}_i,\mu_1)}{f_0(y_{ij}|\mathbf{x}_i,\mu_0)}\right\} \\
&= 2\log\left\{\frac{f(y_{ij}|\mathbf{x}_i,\mu_0(\mathbf{x}_i)+\Delta_i/\sqrt{n})}{f(y_{ij}|\mathbf{x}_i,\mu_0(\mathbf{x}_i))}\right\} \\
&= 2\psi(y_{ij}|\mathbf{x}_i,\mu_0(\mathbf{x}_i))\frac{\Delta_i}{\sqrt{n}} + \dot{\psi}(y_{ij}|\mathbf{x}_i,\mu_0(\mathbf{x}_i))\frac{\Delta_i^2}{n} \\
&\quad + \ddot{\psi}\left(y_{ij}|\mathbf{x}_i,\mu_0(\mathbf{x}_i)+t_i\frac{\Delta_i}{\sqrt{n}}\right)\frac{\Delta_i^3}{3n^{3/2}}.
\end{aligned}$$

Thus, using (A2),

$$R=\sum_{i,j}R_{ij}=U_n+V_n+O_p(n^{-1/2}),$$

where

$$\begin{aligned}
U_n &= \frac{2}{\sqrt{n}}\sum_{i,j}\Delta_i\psi(y_{ij}|\mathbf{x}_i,\mu_0(\mathbf{x}_i)), \\
V_n &= \frac{1}{n}\sum_{i,j}\Delta_i^2\dot{\psi}(y_{ij}|\mathbf{x}_i,\mu_0(\mathbf{x}_i)).
\end{aligned}$$

Downloaded by [University of Alberta], [Doug Wiens] at 09:54 30 November 2011

An expansion similar to that above, and using (A2)-(A4), gives

$$2nD = \sum_{i=1}^{N} \Delta_i^2 J\left(\mathbf{x}_i, \mu_0\left(\mathbf{x}_i\right)\right) \xi_i + O\left(n^{-1/2}\right),$$

where

$$\begin{aligned} J\left(\mathbf{x}_i, \mu_0\left(\mathbf{x}_i\right)\right) &= \int_{-\infty}^{\infty} \psi\left(y|\mathbf{x}_i, \mu_0\left(\mathbf{x}_i\right)\right) \dot{f}\left(y|\mathbf{x}_i, \mu_0\left(\mathbf{x}_i\right)\right) dy \\ &= \int_{-\infty}^{\infty} \psi^2\left(y|\mathbf{x}_i, \mu_0\left(\mathbf{x}_i\right)\right) f\left(y|\mathbf{x}_i, \mu_0\left(\mathbf{x}_i\right)\right) dy. \end{aligned}$$

We then calculate (using (A3)) and Chebyshev's Inequality) that, under either hypothesis,

$$V_n = -\sum_{i=1}^{N} \Delta_i^2 J\left(\mathbf{x}_i, \mu_0\left(\mathbf{x}_i\right)\right) \xi_i + o_p(1) = -2nD + o_p(1).$$

Thus

$$R = U_n - 2nD + o_p(1).$$

Under the null hypothesis, the term U_n has mean 0 (this follows from (A1)); under the alternate hypothesis

$$E_{H_1}[U_n] = 2\sum_{i=1}^{N} \Delta_i^2 J(\mathbf{x}_i, \mu_0(\mathbf{x}_i))\xi_i + O(n^{-1/2}) = 4nD + O(n^{-1/2}).$$

Under either hypothesis,

$$VAR[U_n] = 8nD + O(n^{-1/2}).$$

By (A4), Liapounov's Central Limit Theorem applies:

$$\frac{U_n - E\left[U_n\right]}{\sqrt{VAR\left[U_n\right]}} \xrightarrow{L} N(0,1);$$

(1.3) and (1.4) follow.

(ii) If f is the normal density, then

$$R = \frac{1}{\sigma^2}\left[2\sum_{i,j} Y_{ij}\left(\mu_1\left(\mathbf{x}_i\right) - \mu_0\left(\mathbf{x}_i\right)\right) - \sum_{i=1}^{N} n_i\left(\mu_1^2\left(\mathbf{x}_i\right) - \mu_0^2\left(\mathbf{x}_i\right)\right)\right],$$

which is normally distributed with variance $8nD$, where D is as given in the statement of the theorem. Under the null hypothesis the mean is $-2nD$; under the alternate it is $2nD$. □

Acknowledgements

This research was supported by the Natural Sciences and Engineering Research Council of Canada. The work has benefited from the insightful comments of two anonymous referees.

References

Atkinson, A.C., Fedorov, V.V., 1975a. The design 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.

Atkinson, A.C., 2008. DT-optimum designs for model discrimination and parameter estimation. *Journal of Statistical Planning and Inference*, 138, 56–64.

Fang, Z., Wiens, D.P., 2000. Integer-valued, minimax robust designs for estimation and extrapolation in heteroscedastic, approximately linear models. *Journal of the American Statistical Association*, 95, 807–818.

Fedorov, V.V., 1975. Optimal experimental designs for discriminating two rival regression models. *A Survey of Statistical Design and Linear Models*, Srivastava, J.N. (editor), North Holland, Amsterdam.

Fedorov, V.V., Pazman, A., 1968. Design of Physical Experiments. *Fortschritte der Physik.*, 16, 325–355.

Hill, W.J., Hunter, W.G., Wichern, D.W., 1968. A joint design criterion for the dual problem of model discrimination and parameter estimation. *Technometrics*, 10, 145–160.

Hunter, W.G., Reiner, A.M., 1965. Designs for discriminating between two rival models. *Technometrics*, 7, 307–323.

López-Fidalgo, J., Tommasi, C., Trandafir, P.C., 2007. An optimal experimental design criterion for discriminating between non-normal models. *Journal of the Royal Statistical Society* B, 69, 231–242.

Wiens, D.P., 2009. Robust discrimination designs. *Journal of Royal Statistical Society* (*Series B*), 71, 805–829.

Downloaded by [University of Alberta], [Doug Wiens] at 09:54 30 November 2011