Applied Stochastic Models in Business and Industry

Special Issue Paper

Received 1 February 2011, Revised 27 June 2011, Accepted 27 June 2011 Published online 1 September 2011 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/asmb.917

A robust treatment of a dose–response study

Douglas P. Wiens*† and Pengfei Li

We give a description of the steps leading to a robustification of a dose–response study. The original experiment was designed and described by Rosenberger and Grill (1997) and has since been discussed by several other groups of researchers. Our robustification consists of redesigning the experiment so as to build in flexibility over a range of possible link functions, including the logistic link assumed by the original experimenters. We consider as well an asymptotic Neyman–Pearson test of the validity of the assumed link function.

Keywords: Anderson–Darling distance; discrimination; logistic; Neyman–Pearson test; probit; quadratic programming; simulated annealing

1. Introduction

At the recent Joint Research Conference on Statistics in Quality, Industry and Technology (May 2010, National Institute of Standards and Technology, Gaithersburg, Maryland), the following experimental scenario was discussed. Rosenberger and Grill [1] designed an experiment, the goal of which was to elicit information about the relationship between stimulus level (x) and response, by estimating quantiles of the stimulus response curve. Subjects sequentially received 'marking stimuli' (auditory marking clicks) at various levels and at random times near to those of a certain event and were then asked whether the event occurred before or after the stimulus. The response Y was binary, with $Y = 1$ being recorded if the subject reported that the event occurred before the stimulus. The principal goal was to estimate the median of the stimulus response curve. A secondary goal was to design so as to allow for the estimation of other quantiles such as the lower and the upper quartiles. The investigators assumed a logistic link relating $P(Y = 1)$ to a linear function of the stimulus level and, in their discussion, stated:

> The effects of using logistic regression to analyse data that more closely follow a probit, Weibull, or other distribution ('violation of link assumption') is not particularly well-known among practitioners, even in the independent case. However, in estimating the median, it is likely that logistic and probit analysis would yield similar results.

This question of how the designs might change in response to uncertainty about the appropriate link was the subject of our presentation, and our observations are detailed here. The theoretical development and mathematical details may be found in Li and Wiens [2]. A basic feature of our approach is that we entertain a class of possible link functions, forming a neighbourhood of that used by the experimenter to analyse the data. We introduce certain loss functions, obtain the maximum loss as the links vary over the neighbourhood and, finally, choose a design to minimise the maximum loss.

In §2, we return to the experimental scenario described previously and compare our methods with those of Biedermann *et al.* [3] and Zhu and Wong [4] who have also constructed designs for this experiment. We note that there is another, rather different, approach to dose estimation—one might observe a *continuous* response Y to a dose x and then model $E[Y|x]$ via (typically nonlinear) regression. A referee has kindly pointed us to work by Bornkamp *et al.* [5], Bretz *et al.* [6], Dette *et al.* [7,8], Dragalin *et al.* [9,10] and Zhou *et al.* [11]; in these papers, the aforementioned regression approach is applied, and so there is only a passing relevance to the problems discussed here.

Despite the robustness arising from our minimax approach, questions of model discrimination remain—the experimenter may wish to test that other nearby links are more appropriate for his or her data. Thus, in §3, we investigate the properties of the Neyman–Pearson test to compare two link functions and illustrate the results in the context of the Rosenberger and Grill [1] experiment.

Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alberta T6G 2G1, Canada
**Correspondence to: Douglas P. Wiens, Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alberta T6G 2G1, Canada.*
†E-mail: doug.wiens@ualberta.ca

1.1. Model and robustness requirements

The experimental situation described previously is common to 'dose–response' studies, in which a binary response is linked to a 'dose' x via a cumulative distribution function evaluated at a linear function of the dose. The experimenter assumes that the mean conditional response is given by $E[Y|x] = F_0(\alpha_0 + \beta_0 x)$, for a specified distribution function (link) F_0. The common links are $F(t) = \Phi(t)$ (the 'probit' link) and $F(t) = L(t) = 1/\left(1+e^{-t}\right)$ (the 'logistic' link). The fitted link $F_0(t)$ might, however, not be the 'true' one, and we entertain a family of alternative response functions:

$$\Pr_n(Y = 1|x) = F_n(\alpha_n + \beta_n x) \overset{def}{=} H_n(x).$$

Here, $n = \sum_{i=1}^{I} n_i$ is the sample size, with n_i responses $\{y_{ij}\}_{j=1}^{n_i}$ being recorded at dose level x_i. The levels are chosen from a finite set $\{x_i\}_{i=1}^{N}$. Both the link F_n and the parameters $\boldsymbol{\theta}_n = (\alpha_n, \beta_n)^T$ may depend on n; this is necessary for an asymptotic treatment of the problem, which requires the true and the fitted links to be contiguous in order that errors due to link misspecification, and those due to variation, remain of the same order.

The purpose is to estimate the dose x_p^n required to attain $H_n(x) = p$ for one or more specified values of p:

$$x_p^n \overset{def}{=} x_p(\boldsymbol{\theta}_n; F_n) = \frac{F_n^{-1}(p) - \alpha_n}{\beta_n}.$$

The parameters are to be estimated by maximum likelihood based on the fitted link F_0, and so satisfy the likelihood equations

$$\sum_{i,j} \psi_0\left(y_{ij}; \mathbf{z}_i^T \hat{\boldsymbol{\theta}}_n\right) \begin{pmatrix} 1 \\ x_i \end{pmatrix} = \mathbf{0},$$

where $\psi_0(y;t) = (y - F_0(t))\, w_0(t)$ for $w_0(t) = \frac{d}{dt}\log\left(\frac{F_0(t)}{1-F_0(t)}\right) = \frac{f_0(t)}{F_0(t)(1-F_0(t))}$. The experimenter then estimates the appropriate dose by

$$\hat{x}_p \overset{def}{=} x_p\left(\hat{\boldsymbol{\theta}}_n; F_0\right) = \frac{F_0^{-1}(p) - \hat{\alpha}_n}{\hat{\beta}_n}.$$

He or she desires a design that will afford some robustness against errors arising from link misspecification. By 'design' we mean the design measure ξ_n, placing mass n_i/n at x_i. To construct a robust design, we first obtain the asymptotic normal distribution of $\sqrt{n}\left(\hat{x}_p - x_p^n\right)$. From this, we obtain the mean squared error; this will be maximised over F_n and then minimised by the choice of design.

We define

$$P_n(t) = \sqrt{n}\left(F_n(t) - F_0(t)\right) \ (t \in \mathbb{R}),$$
$$Q_n(p) = \sqrt{n}\left(F_n^{-1}(p) - F_0^{-1}(p)\right) \ (p \in (0,1)),$$

and allow F_n to vary over

$$\mathcal{D}_n(F_0) = \{F_n \mid |P_n(t)| \leqslant \tau \text{ for all } t\},$$

for a fixed τ. This then defines a sequence of shrinking Kolmogorov neighbourhoods (Figure 1).

We must impose a minor restriction on $\mathcal{D}_n(F_0)$ to prevent possible loss of identifiability. Note that if more than one member of a location/scale family belongs to $\mathcal{D}_n(F_0)$, then varying the linear parameters can yield the same response function, viz.

$$F_n(\alpha_n + \beta_n x) \equiv F_n'\left(\alpha_n' + \beta_n' x\right),$$

for appropriate α_n', β_n', if location/scale families

$$F_n'(t) = F_n\left(\frac{t-\mu}{\sigma}\right)$$

belong to $\mathcal{D}_n(F_0)$. This is precluded by the requirement that

$$Q_n(p_1) = Q_n(p_2) = 0,$$

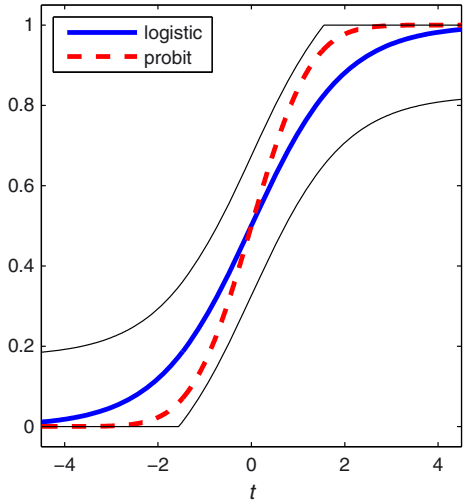

Figure 1. Kolmogorov neighbourhood $\mathcal{D}_n(F_0)$ with $\tau = 1$, $n = 33$ and $F_0(t) = L(t)$, the logistic link. The neighbourhood is large enough to include as well the probit link $F(t) = \Phi(t)$.

for two specified values p_1, p_2; this restriction ensures that only one of F_n, F_n' will belong. Choosing p_1 near 0 and p_2 near 1 ensures that this is a very mild restriction on the possible links.

As is typically the case in design problems for nonlinear models, the loss depends on the unknown parameters. This can be addressed in a number of ways; our treatment of the experiment, which is the subject of this article, is to assume a 'working parameter' $\boldsymbol{\theta}_0 = (\alpha_0, \beta_0)^T$ and construct locally optimal designs.

Results of Le Cam [12] on estimation in contiguous models apply, and yield that the asymptotic mean squared error of the estimate $\hat{x}_p$ of x_p^n is $\left(\sqrt{n}\beta_0\right)^{-2}$ times

$$\mathcal{L}(\xi_n, F_n; p) \overset{def}{=} \left[Q_n(p) + \left(1, x_p^0\right)\mathbf{R}_{\xi_n}^{-1}\boldsymbol{\delta}_{\xi_n,F_n}\right]^2 + \left(1, x_p^0\right)\mathbf{R}_{\xi_n}^{-1}\begin{pmatrix}1\\ x_p^0\end{pmatrix},$$

where $x_p^0 = \left(F_0^{-1}(p) - \alpha_0\right)/\beta_0$,

$$\mathbf{R}_{\xi_n} = \sum_{i=1}^{N}\frac{n_i}{n}\left(f_0(t_i)w_0(t_i)\right)_{|t_i=\alpha_0+\beta_0 x_i}\begin{pmatrix}1 & x_i\\ x_i & x_i^2\end{pmatrix}$$

is the information matrix (under F_0) and

$$\boldsymbol{\delta}_{\xi_n,F_n} = \sum_{i=1}^{N}\frac{n_i}{n}\cdot d_{i,n}\cdot w_0(\alpha_0+\beta_0 x_i)\cdot\begin{pmatrix}1\\ x_i\end{pmatrix},$$

a linear function of the discrepancies $d_{i,n} = \sqrt{n}\left(F_n(\alpha_0+\beta_0 x_i) - F_0(\alpha_0+\beta_0 x_i)\right)$. In Li and Wiens [2], we also establish a representation of the limit of $\sqrt{n}\left(\hat{x}_p - x_p^n\right)$ as a Gaussian process indexed by p; this allows for the treatment of the case in which an interval of values of p is of interest to the experimenter. In this case, the loss function is the integral of $\mathcal{L}(\xi_n, F_n; p)$ over the interval of interest.

1.2. Maximising the mean squared error

Note that $\mathcal{L}(\xi_n, F_n; p)$ is a quadratic function of the $d_{i,n}$ and also of $Q_n(p)$. Its maximisation over the appearance of F_n in $d_{i,n}$ can be handled by quadratic programming. Membership of F_n in $\mathcal{D}_n(F_0)$ imposes certain (linear) constraints on F_n^{-1}, and the maximisation over the appearance of F_n^{-1} in $Q_n(p)$ can then also be handled by quadratic programming. Of course, these are very different quadratic programming problems. Our approach is to solve the pair of problems iteratively, until convergence is attained. This typically happens very quickly.

1.3. Minimising the maximum mean squared error

The maximised loss

$$\mathcal{R}_1(\xi) = \max_{F_n} \mathcal{L}(\xi, F_n; p), \tag{1}$$

or

$$\mathcal{R}_2(\xi) = \max_{F_n} \int_{p_{01}}^{p_{02}} \mathcal{L}(\xi, F_n; p)\, dp \tag{2}$$

if an interval $[p_{01}, p_{02}]$ is of interest, is now to be minimised over designs ξ, represented as vectors $(\xi_1, \ldots, \xi_N)$ of relative frequencies $\xi_i = n_i/n$. This is an integer optimisation problem that evidently has no particular structure that might allow it to be handled by standard optimisers. We thus use simulated annealing. Briefly, this involves iteratively making small random changes in ξ and accepting an improved design with probability 1 and a worse design with positive but small—and decreasing—probability (see Fang and Wiens [13], Bohachevsky *et al.* [14] and Haines [15] for background material on simulated annealing in design theory). The method is computationally very intensive, because each design to be tested requires the solution of the quadratic programming problems described in Section 1.2.

2. Design construction and comparisons

The Rosenberger and Grill [1] experiment described previously, and the question of how the designs might change in response to varying link functions, has been discussed in the literature by other investigators. Biedermann *et al.* [3] (henceforth referred to as 'BDP') addressed this question by constructing designs intended to be simultaneously efficient with respect to various choices of link functions and parameter regions. Their work followed upon that of Zhu and Wong [4] (henceforth referred to as 'ZW'; see also [16]) who constructed Bayesian optimal designs minimising a certain linear combination of loss functions. We note that these studies were concerned with efficiency, that is, with minimising the variance of the estimates, either at the nominal link function or over a small set of possible links. None sought to control the increase in the bias, hence in the mean squared error, of the estimates arising from link misspecification.

Following Rosenberger and Grill [1], we took $n = 71$ possible levels of the stimuli, coded $1, \ldots, 71$ as in the studies to which we make our comparisons. We also took $n = 33$. Following BDP and ZW, we took $\boldsymbol{\theta}_0 = (-10.89, 0.33)^T$. Recall (1) and (2). The loss $\mathcal{R}_1(\xi)$, with $p = 0.5$, is appropriate when the experimenter is solely concerned with estimating the median of the stimulus response curve. If the concern is also on estimation of the quartiles, we minimise instead $\mathcal{R}_2(\xi)$, with $p_{01} = 0.25$, $p_{02} = 0.75$.

The designs of BDP and ZW were required to be symmetric around $x = 33$, and so for purposes of comparison, we will, in Section 2.2, restrict to symmetric robust designs. We first exhibit designs without this restriction and consider both the case $\tau = 0$—appropriate only if one is entirely confident in the appropriateness of the logistic link—and $\tau = 1$, a value large enough that $\mathcal{D}_n(F_0)$ includes the probit link (Figure 1).

2.1. Designs constructed without symmetry restrictions

Using the methods described in Sections 1.1 and 1.2, we constructed designs

LW1: Optimal for loss $\mathcal{R}_1$ when $\tau = 0$;
LW2: Optimal for loss $\mathcal{R}_2$ when $\tau = 0$;
LW3: Optimal for loss $\mathcal{R}_1$ when $\tau = 1$;
LW4: Optimal for loss $\mathcal{R}_2$ when $\tau = 1$.

The designs LW1 and LW3 turn out to be identical and symmetric; thus, they are also minimax under the restriction of symmetry and are discussed in Section 2.2. Designs LW2 and LW4 have dose levels and design weights

$$\text{LW2} = \begin{pmatrix} 29 & 30 & 36 \\ 0.15 & 0.33 & 0.52 \end{pmatrix},$$

$$\text{LW4} = \begin{pmatrix} 30 & 31 & 32 & 33 & 35 & 36 & 40 & 41 \\ 0.03 & 0.39 & 0.03 & 0.09 & 0.12 & 0.15 & 0.09 & 0.09 \end{pmatrix},$$

respectively (Figure 2). As they must, LW2 and LW4 minimise $\mathcal{R}_2$ among all designs considered (Table I) although the best of the symmetric designs are close competitors.

2.2. *Symmetric designs*

The construction of symmetric designs was carried out by slightly modifying our MATLAB code (available from the authors) at the annealing stage—each trial design was symmetrised by averaging it with its reflection across $x = 33$ before being assessed. This resulted in the designs LW1 and LW3, optimal for $\mathcal{R}_1$, as in Section 2.1. We constructed as well designs

LW5: Optimal among symmetric designs for loss $\mathcal{R}_2$ when $\tau = 0$;
LW6: Optimal among symmetric designs for loss $\mathcal{R}_2$ when $\tau = 1$.

The designs are illustrated in Figure 3 and have support points and weights

$$\text{LW1} = \text{LW3} = \begin{pmatrix} 32 & 33 & 34 \\ 0.03 & 0.94 & 0.03 \end{pmatrix},$$

$$\text{LW5} = \begin{pmatrix} 29 & 30 & 36 & 37 \\ 0.07 & 0.43 & 0.43 & 0.07 \end{pmatrix},$$

$$\text{LW6} = \begin{pmatrix} 25 & 26 & 30 & 31 & 32 & 34 & 35 & 36 & 40 & 41 \\ 0.03 & 0.03 & 0.06 & 0.29 & 0.08 & 0.08 & 0.29 & 0.06 & 0.03 & 0.03 \end{pmatrix}.$$

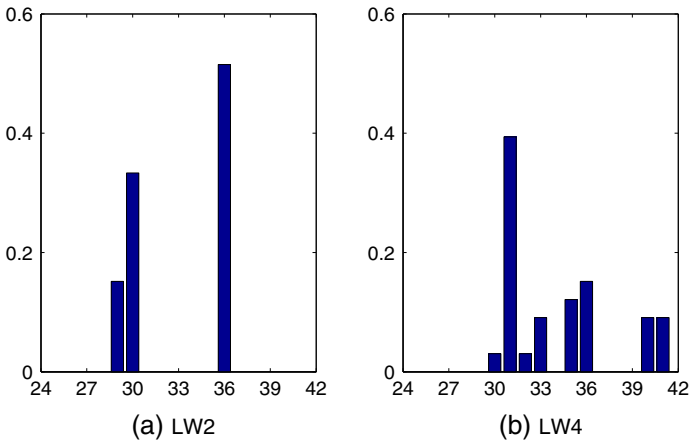

Figure 2. Designs LW2 and LW4, minimising $\mathcal{R}_2$ when $\tau = 0$ and $\tau = 1$, respectively, without symmetry restrictions.

Table I. Comparative losses of the designs.

Design	$\mathcal{R}_1(\xi)$		$\mathcal{R}_2(\xi)$	
	$\tau = 0$[1]	$\tau = 1$	$\tau = 0$[1]	$\tau = 1$
ZW1	6.51	46.32	[3.86]	[28.33]
ZW2	[9.01]	[57.32]	5.08	49.91
BDP1	6.15	41.81	[3.74]	[26.04]
BDP2	[8.83]	[18.99]	5.12	27.85
LW1	**4.01**	4.03	[117.87]	[129.06]
LW2	[5.19]	[92.85]	**3.49**	24.83
LW3	4.01	**4.03**	[117.87]	[129.06]
LW4	[5.16]	[5.80]	3.71	**4.06**
LW5	[5.18]	[92.74]	**3.49**	25.06
LW6	[4.89]	[5.75]	3.75	**4.37**

[1]When $\tau = 0$, the loss is due solely to variation.

Square brackets denote estimation situations in which the indicated design is not intended to be appropriate. Figures in bold correspond to situations in which the LW designs are intended to be optimal (among symmetric designs, for LW5 and LW6).

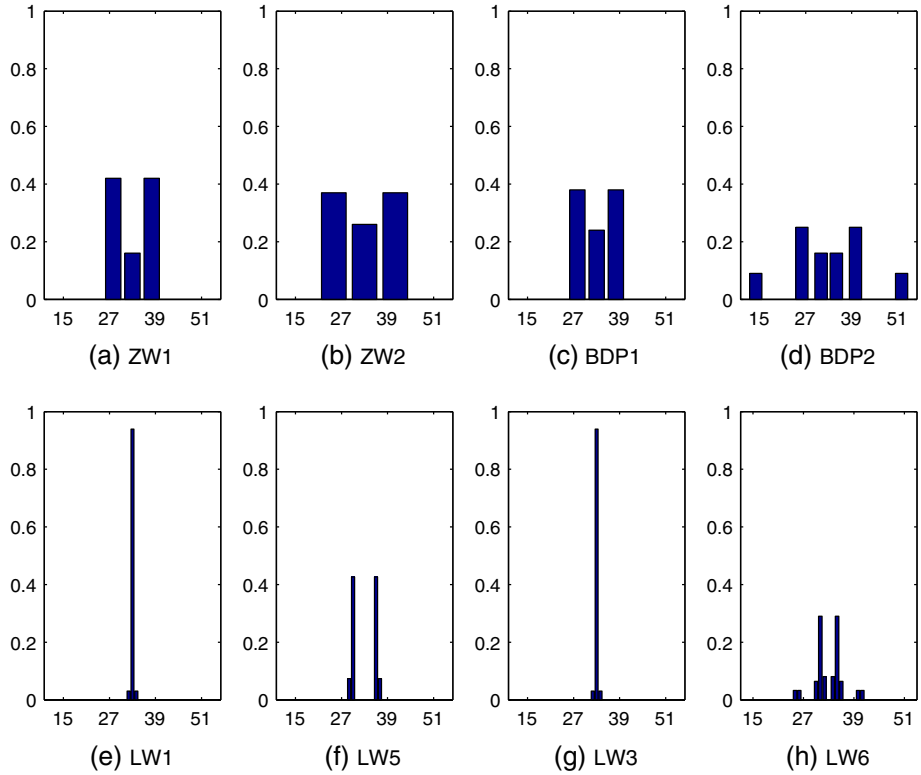

Figure 3. Symmetric designs compared in Section 2.2.

We compare these with the symmetric designs of ZW and BDP:

$$\text{ZW1} = \begin{pmatrix} 28 & 33 & 38 \\ 0.42 & 0.16 & 0.42 \end{pmatrix},$$
$$\text{ZW2} = \begin{pmatrix} 25 & 33 & 41 \\ 0.37 & 0.26 & 0.37 \end{pmatrix},$$
$$\text{BDP1} = \begin{pmatrix} 28 & 33 & 38 \\ 0.38 & 0.24 & 0.38 \end{pmatrix},$$
$$\text{BDP2} = \begin{pmatrix} 14 & 26 & 31 & 35 & 40 & 52 \\ 0.09 & 0.25 & 0.16 & 0.16 & 0.25 & 0.09 \end{pmatrix}.$$

Designs ZW1 and ZW2 are the Bayesian designs of ZW, optimal for estimating the median (ZW1) or the quartiles (ZW2) of the stimulus response curve. Designs BDP1 and BDP2 are the correspondingly optimal 'maximin efficient' designs of BDP. We note that ZW and BDP constructed continuous designs, and so some rounding of the levels has been carried out for the purpose of comparison with our designs. From the performance measures in Table I, we see that the symmetric designs LW1 and LW5 constructed for optimality against variation alone ($\tau = 0$) are indeed the most efficient of those studied here. The symmetric designs LW3 and LW6, constructed for robustness as well as efficiency, perform better than the corresponding ZW and BDP designs and at least almost as well as their non-symmetric competitors.

One might ask if the robustness is achieved at too great a loss in efficiency. Of course both $\mathcal{R}_1$ and $\mathcal{R}_2$ measure efficiency when $\tau = 0$; from Table I, we see that, with the exception of LW1 (LW3) when used—inappropriately—for quartile estimation, the LW designs are quite efficient. To further answer this question, we computed the root of the total variance, that is, $\left[\det\left(\mathbf{R}_\xi\right)\right]^{-1/2}$, for all designs under consideration. With respect to this measure, the design LW1 = LW3 fares poorly. Designs LW2, LW4, LW5 and LW6 for $\mathcal{R}_2$ are at least almost as efficient as, and in some cases more efficient than, the competing designs of ZW and BDP constructed for efficiency alone:

Design	ZW1	ZW2	BDP1	BDP2	LW1	LW2	LW3	LW4	LW5	LW6
$\lvert\mathbf{R}_\xi\rvert^{-1/2}$	1.51	1.75	1.55	1.91	16.48	1.65	16.48	1.84	1.65	1.93.

The design LW1 performs very well with respect to $\mathcal{R}_1(\xi)$, yet poorly with respect to total variance. This illustrates the differing emphasis of these loss functions—when $\tau = 0$ and $p = 0.5$, $\mathcal{R}_1(\xi) = (1, 33)\,\mathbf{R}_{\xi_n}^{-1}\binom{1}{33}$ is the variance of the linear estimate $\hat{\alpha} + 33\hat{\beta}$ at the particular dose level $x = 33$ and can be expected to be minimised by a design such as LW1 with most mass at this level. In contrast, the total variance measures the quality of the parameter estimates and typically calls for a design with greater spread. We also note that Yang and Stufken [17] have shown the optimality of three-point (possibly two-point) designs, when the loss depends on variance alone and the design space is an asymmetric (but continuous) interval. Our designs LW1 and LW2 for $\tau = 0$ are in line with these results, whereas LW5 appears to be an approximation, in our discrete design space, of a two-point design.

3. Post-design link discrimination

Having constructed a design and gathered the data, the experimenter is able to test if his or her assumed link F_0 is appropriate. In the context of the experiment described in this article, this is best accomplished by the Neyman–Pearson test: one tests

$$\mathrm{H}_0 : F = F_0,\ \mathrm{H}_1 : F = F_1, \tag{3}$$

with each link evaluated at $\mathbf{z}^T\boldsymbol{\theta}$ for $\boldsymbol{\theta}$ fixed, for example, at $(-10.89, 0.33)^T$ in our example. The pair F_0, F_1 might be plausible competitors with $F_1 \in \mathcal{D}_n(F_0)$, for example, F_0 logistic and F_1 probit. Or—and we shall also investigate this case here—F_1 could be on the boundary of $\mathcal{D}_n(F_0) : F_1 = F_{\mathrm{bdry}}$ with

$$F_{\mathrm{bdry}}(t) = \begin{cases} \min\left(F_0(t) + \tau/\sqrt{n}, 1\right), & t \geqslant 0, \\ \max\left(F_0(t) - \tau/\sqrt{n}, 0\right), & t < 0. \end{cases}$$

In this case, a failure to reject H_0 asserts that F_0 does 'well enough' relative to the closest competitor, which is (almost) not in $\mathcal{D}_n(F_0)$—a kind of goodness-of-fit criterion.

Given data $\{x_i, y_{ij}\}$, where $y_{ij} \in \{0, 1\}$ denotes the response in the jth replicate at dose x_i, the likelihood function under H_k $(k = 0, 1)$ is $\prod_{i,j} P_{H_k}\left(Y = y_{ij}|\mathbf{z}_i\right)$, with

$$P_{H_k}(Y = y|\mathbf{z}) = \left[F_k\left(\mathbf{z}^T\boldsymbol{\theta}\right)\right]^y \left[\bar{F}_k\left(\mathbf{z}^T\boldsymbol{\theta}\right)\right]^{1-y}.$$

We use the notation $\bar{F} = 1 - F$. The Neyman–Pearson test of (3) rejects for large values of

$$R = 2\sum_{i,j} \log \frac{P_{H_1}\left(Y = y_{ij}|\mathbf{z}_i\right)}{P_{H_0}\left(Y = y_{ij}|\mathbf{z}_i\right)} = 2\sum_{i,j}\left[Y_{ij}\kappa_i + \left(1 - Y_{ij}\right)\bar{\kappa}_i\right],$$

where

$$\kappa_i = \log \frac{F_1\left(\mathbf{z}_i^T\boldsymbol{\theta}\right)}{F_0\left(\mathbf{z}_i^T\boldsymbol{\theta}\right)}, \bar{\kappa}_i = \log \frac{\bar{F}_1\left(\mathbf{z}_i^T\boldsymbol{\theta}\right)}{\bar{F}_0\left(\mathbf{z}_i^T\boldsymbol{\theta}\right)}.$$

In Appendix A, we prove Theorem 1, which gives the asymptotic properties of this test.

Theorem 1
For $F_1 \in \mathcal{D}_n(F_0)$, define

$$\Delta = \int \frac{\left\{F_1\left(\mathbf{z}^T\boldsymbol{\theta}\right) - F_0\left(\mathbf{z}^T\boldsymbol{\theta}\right)\right\}^2}{2F_0\left(\mathbf{z}^T\boldsymbol{\theta}\right)\bar{F}_0\left(\mathbf{z}^T\boldsymbol{\theta}\right)} d\xi_n.$$

Then Δ is $O\left(n^{-1}\right)$ as $n \to \infty$. Under H_0, the mean of R is $\mu_0 = -2n\Delta + o(1)$; under H_1, it is $\mu_1 = 2n\Delta + o(1)$. Under either hypothesis, the variance is $\sigma_R^2 = 8n\Delta + o(1)$. The test statistic is asymptotically normal under each hypothesis: under H_0,

$$\frac{R + 2n\Delta}{\sqrt{8n\Delta}} \xrightarrow{L} N(0, 1),$$

whereas under H_1,

$$\frac{R - 2n\Delta}{\sqrt{8n\Delta}} \xrightarrow{L} N(0, 1).$$

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

$$\beta = \Phi(\sqrt{2n\Delta} - z_\alpha).$$

We note that Δ is similar to the Anderson–Darling statistic, which weights by F_0 rather than by the design measure ξ_n and measures the discrepancy of F_1 from F_0.

We have obtained the asymptotic powers, of the designs presented in Section 2, for the two hypothesis testing situations discussed previously. For testing a logistic null against a probit alternative, these are given in Table II (for the 'boundary' alternative, see Table III). In each case, the most powerful tests result from designs ZW1, BDP1, LW2 and LW5, with the first two of these outperforming the second two. The robust designs LW3 and LW4 have lower power than the others—something that is perhaps to be expected, because they are constructed to accommodate competing links, rather than to discriminate between them. In each case, we computed as well the power that would result if the experimenter, dissatisfied with the current power, were to take two more observations at each of two points, chosen to maximise the integrand in the definition of Δ. For both alternatives (Figure 4), these are the points $x = 28$ and $x = 38$. Not surprisingly, the increase in the power is less for those designs that already place some mass at these points.

Table II. Comparative powers of the designs in a size $\alpha = 0.05$ test of H_0: $F = L$ versus H_1: $F = \Phi$.

Design	$n = 33$	Augmented[1]
ZW1	0.482	0.526
ZW2	0.342	0.395
BDP1	0.451	0.497
BDP2	0.358	0.410
LW1	0.066	0.153
LW2	0.426	0.474
LW3	0.066	0.153
LW4	0.304	0.360
LW5	0.425	0.473
LW6	0.298	0.353

[1]Two points added at each of $x = 28$ and $x = 38$.

Table III. Comparative powers of the designs in a size $\alpha = 0.05$ test of H_0: $F = L$ versus H_1: $F =$ bdry.

Design	$n = 33$	Augmented[1]
ZW1	0.787	0.829
ZW2	0.509	0.595
BDP1	0.776	0.819
BDP2	0.543	0.624
LW1	0.639	0.706
LW2	0.738	0.788
LW3	0.639	0.706
LW4	0.652	0.716
LW5	0.738	0.788
LW6	0.663	0.725

[1]Two points added at each of $x = 28$ and $x = 38$.

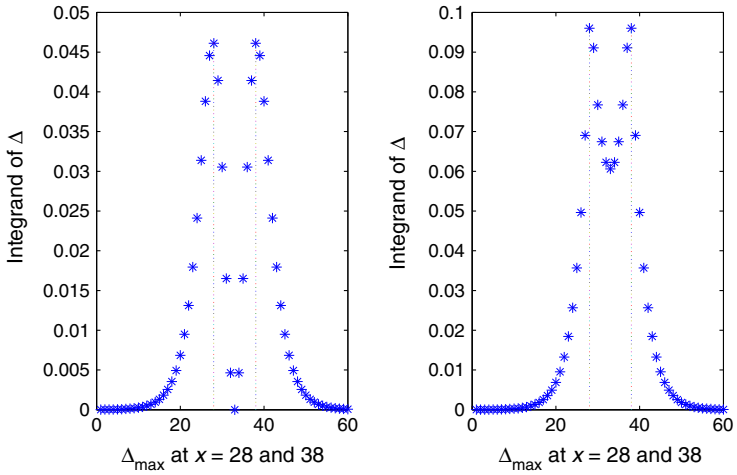

Figure 4. Integrand of Δ versus x; probit alternative (left) and boundary alternative (right).

APPENDIX A.

Proof of Theorem 1

Under H_k, Y_{ij} has mean $F_k\left(\mathbf{z}_i^T\boldsymbol{\theta}\right)$, and so the mean of R is

$$\mu_k = 2\sum_{i,j}\left[F_k\left(\mathbf{z}_i^T\boldsymbol{\theta}\right)\kappa_i + \bar{F}_k\left(\mathbf{z}_i^T\boldsymbol{\theta}\right)\bar{\kappa}_i\right] = 2nD_{k,n},$$

for

$$D_{k,n} = \int\left[F_k\left(\mathbf{z}^T\boldsymbol{\theta}\right)\log\frac{F_1\left(\mathbf{z}^T\boldsymbol{\theta}\right)}{F_0\left(\mathbf{z}^T\boldsymbol{\theta}\right)} + \bar{F}_k\left(\mathbf{z}^T\boldsymbol{\theta}\right)\log\frac{\bar{F}_1\left(\mathbf{z}^T\boldsymbol{\theta}\right)}{\bar{F}_0\left(\mathbf{z}^T\boldsymbol{\theta}\right)}\right]d\xi_n.$$

The variance under H_k is

$$\sigma_k^2 = 4\sum_{i,j}F_k\left(\mathbf{z}_i^T\boldsymbol{\theta}\right)\bar{F}_k\left(\mathbf{z}_i^T\boldsymbol{\theta}\right)(\kappa_i-\bar{\kappa}_i)^2 = 4nV_{k,n},$$

for

$$V_{k,n} = \int F_k\left(\mathbf{z}^T\boldsymbol{\theta}\right)\bar{F}_k\left(\mathbf{z}^T\boldsymbol{\theta}\right)\log^2\left(\frac{F_1}{F_0}\left(\mathbf{z}^T\boldsymbol{\theta}\right)\Big/\frac{\bar{F}_1}{\bar{F}_0}\left(\mathbf{z}^T\boldsymbol{\theta}\right)\right)d\xi_n.$$

We apply the expansions

$$t_n\log\left(\frac{t_n}{t_0}\right) + (1-t_n)\log\left(\frac{1-t_n}{1-t_0}\right) = \frac{1}{2}\frac{(t_n-t_0)^2}{t_0(1-t_0)} + o\{(t_n-t_0)^2\},$$
$$t_0\log\left(\frac{t_n}{t_0}\right) + (1-t_0)\log\left(\frac{1-t_n}{1-t_0}\right) = -\frac{1}{2}\frac{(t_n-t_0)^2}{t_0(1-t_0)} + o\{(t_n-t_0)^2\}$$

to $D_{1,n}$ and $D_{0,n}$, respectively, to obtain

$$D_{1,n} = \Delta + o(n^{-1}),$$
$$D_{0,n} = -\Delta + o(n^{-1}).$$

We then apply the expansions

$$t_n(1-t_n)\log^2\left(\frac{t_n}{t_0}\Big/\frac{1-t_n}{1-t_0}\right) = \frac{(t_n-t_0)^2}{t_0(1-t_0)} + o(n^{-1}),$$
$$t_0(1-t_0)\log^2\left(\frac{t_n}{t_0}\Big/\frac{1-t_n}{1-t_0}\right) = \frac{(t_n-t_0)^2}{t_0(1-t_0)} + o(n^{-1})$$

to $V_{1,n}$ and $V_{0,n}$, respectively, to obtain

$$V_{k,n} = 2\Delta + o\left(n^{-1}\right).$$

This gives the asymptotic means and variances stated in the theorem. The asymptotic normality is a consequence of Liapounov's central limit theorem. □

Acknowledgements

The research of both authors is supported by the Natural Sciences and Engineering Research Council of Canada. We are grateful to two anonymous referees for their helpful comments.

References

1. Rosenberger WF, Grill SE. A sequential design for psychophysical experiments: an application to estimating timing of sensory events. *Statistics in Medicine* 1997; **16**:2245–2260.
2. Li P, Wiens DP. Robustness of design in dose–response studies. *Journal of the Royal Statistical Society: Series B* 2011; **17**:215–238.
3. Biedermann S, Dette H, Pepelyshev A. Some robust design strategies for percentile estimation in binary response models. *The Canadian Journal of Statistics* 2006; **4**:603–622.
4. Zhu W, Wong WK. Bayesian optimal designs for estimating a set of symmetrical quantiles. *Statistics in Medicine* 2001; **20**:123–137.
5. Bornkamp B, Bretz F, Dette H, Pinheiro J. Response-adaptive dose-finding under model uncertainty. *Annals of Applied Statistics* 2011; **5**:1611–1631.
6. Bretz F, Dette H, Pinheiro J. Practical considerations for optimal designs in clinical dose finding studies. *Statistics in Medicine* 2010; **29**:731–742.
7. Dette H, Bretz F, Pepelyshev A, Pinheiro J. Optimal designs for dose finding studies. *Journal of the American Statistical Association* 2008; **103**(483):1225–1237.
8. Dette H, Kiss C, Bevanda M, Bretz F. Optimal designs for the emax, log-linear and exponential models. *Biometrika* 2010; **97**(2):513–518.
9. Dragalin V, Bornkamp B, Bretz F, Miller F, Padmanabhan SK, Patel N, Perevozskaya I, Pinheiro J, Smith JR. A simulation study to compare new adaptive dose-ranging designs. *Statistics in Biopharmaceutical Research* 2010; **2**(4):487–512.
10. Dragalin V, Hsuan F, Padmanabhan SK. Adaptive designs for dose-finding studies based on sigmoid emax model. *Journal of Biopharmaceutical Statistics* 2007; **17**:1051–1070.
11. Zhou X, Joseph L, Wolfson DB, Blisle P. A Bayesian A-optimal and model robust design criterion. *Biometrics* (2003); **59**:1082–1088.
12. Le Cam L. On some asymptotic properties of maximum likelihood estimates and related Bayes procedures. *University of California Publications in Statistics* 1953; **1**:277–330.
13. Fang Z, Wiens DP. Integer-valued, minimax robust designs for estimation and extrapolation in heteroscedastic, approximately linear models. *Journal of the American Statistical Association* 2000; **95**:807–818.
14. Bohachevsky IO, Johnson ME, Stein ML. Generalized simulated annealing for function optimization. *Technometrics* 1986; **28**:209–217.
15. Haines LM. The application of the annealing algorithm to the construction of exact optimal designs for linear-regression models. *Technometrics* 1987; **29**:439–447.
16. Zhu W, Wong WK. Multiple-objective designs in a dose–response experiment. *Journal of Biopharmaceutical Statistics* 2000; **10**:1–14.
17. Yang M, Stufken J. Support points of locally optimal designs for nonlinear models with two parameters. *Annals of Statistics* 2009; **37**:518–541.