

EXPERIMENTAL DESIGN FOR BODY IMAGE TESTING

Craig J. Johns

Russel L. Boice

Univerisity of Colorado, Denver

June 26, 2003

Abstract

Body image experiments require a subject to give a binary response to a large number of presented images. We use simulation to compare several data collection methods based on experimental design theory in order to build an efficient sequential design for a specific type of body image experiment.

1 Introduction

A distorted view of one's body is regarded as a defining feature of anorexia nervosa (American Psychiatric Association, 1980). Many psychophysical experiments for estimating body-image distortion have been proposed, (see for example, Lacey and Hart, 1985, Fries, 1977, Garfinkle et al. 1979). Watt and Andrews (1985) proposed an adaptive probit estimation (APE) technique based on a local design with four points. Fonagy et al. (1990) showed that APE is relatively free of systematic biases, due to psychological factors, that hampered other objective techniques.

Gardner and Boice (2003) developed a portable, freely available software that runs a variety of body image experiments based on the APE method. However, their implementation of APE is *ad hoc* and does not incorporate statistical design theory. A battery of psychophysical tests are available in the software described in Gardner and Boice (2003). Here we focus on the specific test geared toward estimating self-perception of body image. The essence of the experiment is to present the subject with a sequence of digital photos that have been distorted in the horizontal axis, and ask the subject for a binary response: "Too thin" or "Too wide." The collection of responses and distortion levels provide the basis for estimating the subject-specific response curve via probit-regression. Of primary interest is the "point of subjective equality" (PSE), and the "difference limen" (DL). The PSE is frequently denoted by P50 in the psychology literature and $L_{.5}$ in the statistical literature and is the point at which the subject responds "Too thin" and "Too wide" with equal probability. The DL is also referred to as the sensitivity, and is related to the slope of the probit curve. The DL is commonly estimated by $L_{.75} - L_{.25}$.

In this paper we use Monte-Carlo simulation to compare several designs for probit estimation in order to make specific recommendations to improve the methodologies implemented in the software of Gardner and Boice (2003) that will also be useful for other psychophysical experiments.

2 Statistical Theory

Experimental design for logistic/probit regression and sensitivity analysis is a topic with deep statistical heritage. The up-and-down method of Dixon and Mood (1948) and Stochastic Approximation method of Robbins and Monro (1951) are well known and asymptotic results for these and other sequential designs are discussed at length in Wu (1985). The reader is referred to Young and Easterling (1994) for a broad review of statistical methodologies for quantile estimation.

Khan and Yazdi (1988), show that a 2-point design is D-optimal for both probit and logistic regression. These results were confirmed in a Bayesian setting by Chaloner and Larntz (1989). Both of these articles suggest that locally D-optimal designs be used when there is a lack of information about the true parameters. Sun and Tsutakawa (1997) give an example of designs for logistic regression based on changing the usual criterion from the typical expected posterior variance to one which penalizes for large prediction variances as well. However, optimal designs in logistic and probit regression cases, as well as other generalized linear models, depend on the parameters of the interest. Hence, the theory can only act as a guide for the practitioner.

To motivate the particular designs used below, we give a brief review of the theory. Let $\Phi(z)$ denote the cumulative distribution function for the standard normal distribution let $\phi(z)$ be the corresponding pdf. The typical probit model likelihood is of the form $\mathcal{L}(\mathbf{Y}, \mu, \sigma) = \prod_{i=1}^t \Phi(\frac{x_i - \mu}{\sigma})^{y_i} (1 - \Phi(\frac{x_i - \mu}{\sigma}))^{(1 - y_i)}$, and the expected information in a single observation y_t is:

$$\mathcal{I}(x_i, \mu, \sigma) = w(z_i) \begin{bmatrix} 1 & z_i \\ z_i & z_i^2 \end{bmatrix} \quad (1)$$

where $w(z) = \frac{\phi(z)^2}{\sigma^2 \Phi(z)(1 - \Phi(z))}$ and $z_i = (x_i - \mu)/\sigma$. The D-optimal design point for the $t + 1$ observation is based on the minimizing the determinant of $(Q_t + \mathcal{I}_{t+1})^{-1}$, where $Q_t = \sum_{i=1}^t \mathcal{I}_i$. Minimizing $\det(Q_{t+1}^{-1}) = \det(Q_t) + w(z_{t+1})(1, z_{t+1})Q_t^{-1}(1, z_{t+1})^T$ is equivalent to maximizing the objective function

$$D(z) = w(z) (q_{11}z^2 - 2q_{12}z + q_{22}) \quad (2)$$

where q_{ij} is the (i, j) element of Q_t and $x_{t+1} = \mu + z\sigma$.

In the special case where no initial design is available, we make the following heuristic argument for a two-point design and refer the reader to Khan and Yazdi (1988) for more powerful results. Restricted to sampling at only two points, the determinant of the information matrix reduces to

$$D(z_1, z_2) = w(z_1)w(z_2)(z_1 - z_2)^2. \quad (3)$$

By Jensen's inequality, $w(z_1)w(z_2) \leq (w(z_1) + w(z_2))^2/4$ with equality if and only if $w(z_1) = w(z_2)$. Furthermore, $(z_1 - z_2)^2 \leq (2z_1)^2$ for any z_2 with equality if and only if $z_2 = -z_1$. Thus, the objective function is maximized when $z_1 = -z_2$ since w is symmetric about zero. When $z_1 = -z_2$, the objective function becomes $(2w(z)z)^2$ which is maximized at $z \approx \pm 1.138$ and the two-point D-optimal design uses $x = \mu \pm 1.138\sigma$.

The sequential A-optimal point design minimizes the trace of $(Q_t + \mathcal{I}_{t+1})^{-1}$, so the objective function to minimize is:

$$A(z) = \frac{q_{11} + q_{22} + w(z)(1 + z^2)}{(q_{11}q_{22} - q_{12}^2) + w(z)(q_{11}z^2 - 2q_{12}z + q_{22})}. \quad (4)$$

Thus, sequential A-optimal design points will minimize $A(z)$. In the special case of a two-point design with no other information, the objective function simplifies to

$$A(z_1, z_2) = \frac{w(z_1)(z_1^2 + 1) + w(z_2)(z_2^2 + 1)}{w(z_1)w(z_2)(z_1 - z_2)^2} \quad (5)$$

Again by Jensen's inequality, $2\sqrt{w(z_1)(z_1^2 + 1)w(z_2)(z_2^2 + 1)} \leq w(z_1)(z_1^2 + 1) + w(z_2)(z_2^2 + 1)$ with equality if and only if $w(z_1)z_1^2 = w(z_2)z_2^2$, which by the symmetry of w is equivalent to $z_1^2 = z_2^2$. Arguing as in the D-optimal case for the denominator, we see that $A(z_1, z_2)$ is minimized if and only if $z_1 = -z_2$ in which case the objective function becomes $(z^2 + 1)/(2z^2w(z))$. Thus, the two-point A-optimal design uses $x = \mu \pm 1.090\sigma$.

A Bayesian approach to the design is to find x_{t+1} that minimizes the posterior variance of $E(\theta|Y_{t+1}, \mathbf{Y}_t)$ where $\mathbf{Y}_t = (Y_1, \dots, Y_t)$ and $\theta = (\mu, \sigma)$. Note that

$$V[E(\theta|Y_{t+1}, \mathbf{Y}_t)] = V(Y_{t+1}) [2E(\theta\Phi(x_{t+1}, \mu, \sigma)|\mathbf{Y}_t) - E(\theta|\mathbf{Y}_t)]^2 \quad (6)$$

and so the objective is to find x_{t+1} that is a root of

$$\int \int \mu\sigma \left(2\Phi\left(\frac{x_{t+1} - \mu}{\sigma}\right) - 1 \right) p(\mu, \sigma|\mathbf{Y}_t) d\mu d\sigma \quad (7)$$

where $p(\mu, \sigma|\mathbf{Y}_t)$ is the posterior distribution of (μ, σ) given data up to t . Note that if $x_{t+1} \ll \tilde{\mu}_t$, then $(2\Phi - 1)$ is positive for $\mu \approx \tilde{\mu}_t$ where the posterior has high density. Likewise, if $x_{t+1} \gg \tilde{\mu}_t$, then $(2\Phi - 1)$ is negative where the posterior has high density. Thus, it is reasonable to assume that a root exists. We found that the posterior mean of $p(\mu, \sigma|\mathbf{Y}_t)$ is a reasonable estimate of the root.

The adaptive Robbins-Monro (Robbins-Monro, 1951) method uses $x_{t+1} = x_t - c(y_t - p)/t$ to estimate L_p where c is a constant. Robbins-Monro was designed for estimating the quantile of an unknown curve, and in this case we have roughly equal interest in the three quartiles. Chung (1954) showed that $c^{-1} = F'(L_p)$ minimizes the variance of $\sqrt{n}(x_t - L_p)$. An asymptotically equivalent modification is to use $c^{-1} = \max[\min(\hat{\beta}_t, d), \delta]$ where $\hat{\beta}_t$ is the usual least squares regression of the observations y_t on the locations x_t and the truncation is used to reduce the erratic behavior of $\hat{\beta}_t$.

3 Simulation Study

In order to determine a reasonable strategy to program into the next version of the software, we simulated data using 8 different scenarios. We assumed a normal distribution with mean zero and variance 2 for μ and a gamma distribution with mean 1 and variance 1/4 as the prior on σ . These priors allow for fast and accurate estimates of integrals using Gauss-Hermite and Gauss-Laguerre approximations. We determined efficacy of a method by comparing MSE for the posterior mode at $n = 80, 120, \text{ and } 200$.

We tested 2 variations each of the D and A optimal methods, two variations of the adaptive Robbins-Monro method (Robbins and Monro 1951, and Wu, 1985), an approximate minimization of the conditional variance as well as an approximate implementation of the APE method in Gardner and Boice (2003) over a two-way complete factorial design. The "true" values for μ were one of the (.01, .05, .10, .15, .25, .5) quantiles of the prior distribution for μ and σ was set to the (.05, .25, .5, .75, .95) quantiles from the prior distribution of σ . These design points are shown in Figure 3 along with the median denoted by "x"

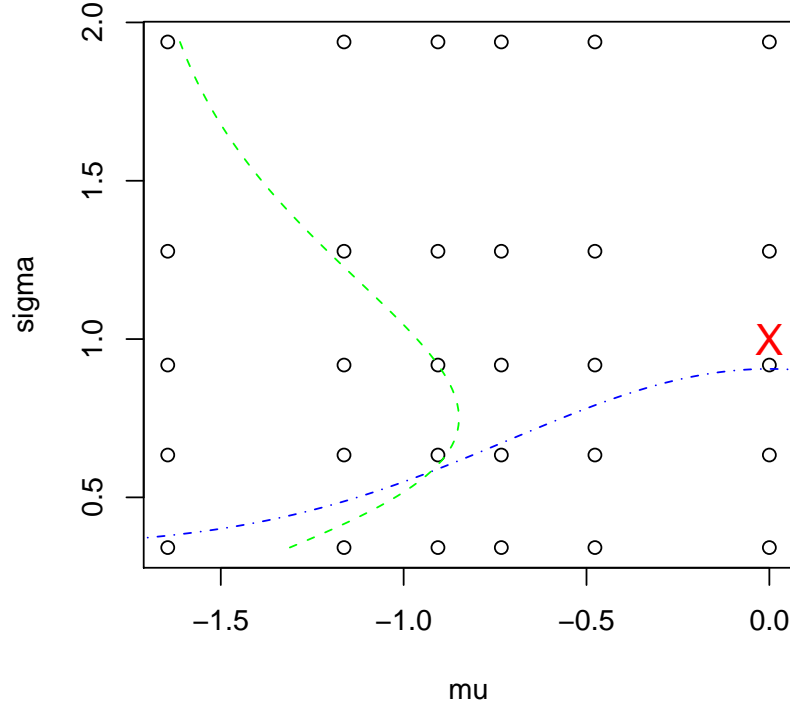


Figure 1: The design points for the simulation are shown with the “o” symbol, and the median of the prior distributions with an “x”. The dashed lines show the prior densities.

and dashed lines that show the density of the prior distributions of μ and σ . Four hundred replicates at each of these 30 combinations were calculated for each of the eight methods. The initial design is based on a random sample from the prior distribution for μ . To reduce variability due to initial design, 400 random designs based on 20 points were fixed for each of the $6 \times 5 \times 8$ combinations we tested. The eight different experimental designs are explicitly defined below.

- D1. Find x_{t+1} to minimize the determinant of the inverse-information matrix (2) evaluated at the posterior mode $(\tilde{\mu}, \tilde{\sigma})$.
- D2. Use the local D-optimal design with $x_{t+1} = \tilde{\mu}_t + 1.138\tilde{\sigma}$ and $x_{t+2} = \tilde{\mu}_t - 1.138\tilde{\sigma}$ as in (3).
- A1. Find x_{t+1} to minimize the trace of the inverse information matrix (4) using Q_t evaluated at the posterior mode.
- A2. Use the local A-optimal design with $x_{t+1} = \tilde{\mu}_t + 1.09\tilde{\sigma}$ and $x_{t+2} = \tilde{\mu}_t - 1.09\tilde{\sigma}$ as in (5).
- R1. Use the Robbins-Monro method with three design points where

$$c_t^{-1} = \phi((x - \tilde{\mu}_t)/\tilde{\sigma}_t) / \sigma$$

$$x_{t+1} = x_{t-2} - (y_{t-2} - 1/4)c_t/t$$

$$x_{t+2} = x_{t-1} - (y_{t-1} - 1/2)c_t/t$$

$$x_{t+3} = x_t - (y_t - 3/4)c_t/t$$
- R2. Use the Robbins-Monro method with three design points where

$$c_t^{-1} = \hat{\beta}_t$$

and the updates are as in R1.

- VE. Choose x_{t+1} to minimize the variance of the conditional expectation where the integrals of (6) are estimated by Gauss-Hermite and Gauss-Laguerre approximations;

$$\frac{\sum_{i=1}^m \sum_{j=1}^n \left[2 \Phi \left(\frac{x_{t+1} - \mu_i}{\sigma_j} \right) - 1 \right] \mu_i \sigma_j L(Y_t | \mu_i, \sigma_j) w_i v_j}{\sum_{i=1}^m \sum_{j=1}^n L(Y_t | \mu_i, \sigma_j) w_i v_j} \quad (8)$$

where $\sigma_j = r_j/4$ and $\mu_i = z_i/\sqrt{2}$, r_j, v_j are the roots and associated weights for the Gauss-Laguerre approximation, and z_i, w_i are the roots and associated weights for the Gauss-Hermite approximation. For speed in the simulations, we used $m = 6$ and $n = 3$ this likely affected the efficiency of the procedure.

- GB. An approximation of the algorithm in the software of Gardner and Boice (2003). The algorithm starts with an initial design by replicating 4 design points 10 times each. The $k+1$ block of design points are determined by $\bar{\mu}_{k-1} + \bar{\sigma}_{k-1}(-1.35, -.45, .45, 1.35)$. The multipliers $(-1.35, -.45, .45, 1.35)$ are based on the suggestion given by Watts and Andrews (1985). The k^{th} block estimate of μ and σ is $(\bar{\mu}, \bar{\sigma})_k$ the solution to the the nonlinear least-squares problem with objective function

$$\sum_{i=1}^k \sum_{j=1}^4 \left[\hat{p}_{ij} - \Phi \left(\frac{x_{ij} - \mu}{\sigma} \right) \right]^2 \quad (9)$$

where $\hat{p}_{ij} = \sum_{k=1}^{10} y_{ijk}/10$ and μ and σ are restricted to be within a tolerance of μ_{k-1} and σ_{k-1} respectively. For $\bar{\mu}_0$ and $\bar{\sigma}_0$ we use the medians of the prior distributions.

4 Results

Figure 4 shows contours of the estimated RMSE for estimating the mean μ using the eight different methods with $n = 200$. Figure 4 is a similar plot for estimating σ and Figure 4 shows $\sqrt{MSE_\mu + MSE_\sigma}$. From these plots and Table 4 there is little difference in the MSEs for the locally D-optimal and locally A-optimal designs. This is because in these cases the locally optimal sampling points are at $\bar{\mu} \pm 1.138\bar{\sigma}$ and $\bar{\mu} \pm 1.090\bar{\sigma}$.

We further note that sequentially, A-optimality is generally more efficient than D-optimality. This is particularly true at the early stages ($n < 100$) when the uncertainty in estimates of μ and σ is high. We speculate that this difference is due to the fact that $A(z)$ in (4) has a denominator that can “standardize” gross errors due to estimation, whereas $D(z)$ in (2) has no such standardization.

Table 4 shows the average rank of the MSEs of the eight method for estimating μ, σ and the average rank of the sum of these MSE’s. The average is taken over all 30 design points shown in Figure 3 and each MSE is calculated via simulation with the 400 replicates. Low ranks correspond with low MSE.

We also found that in many cases the sequential A-optimal designs were essentially trying to emulate the locally A-optimal two-point designs. Figure 4 shows the design locations for $21 \leq n \leq 200$ for the sequential A-optimal design denoted by \triangle and for the locally A-optimal 2 point design \circ for the case when $\mu = -0.906$ and $\sigma = 0.634$. The dashed lines show the optimal levels for sampling based on knowledge of μ and σ .

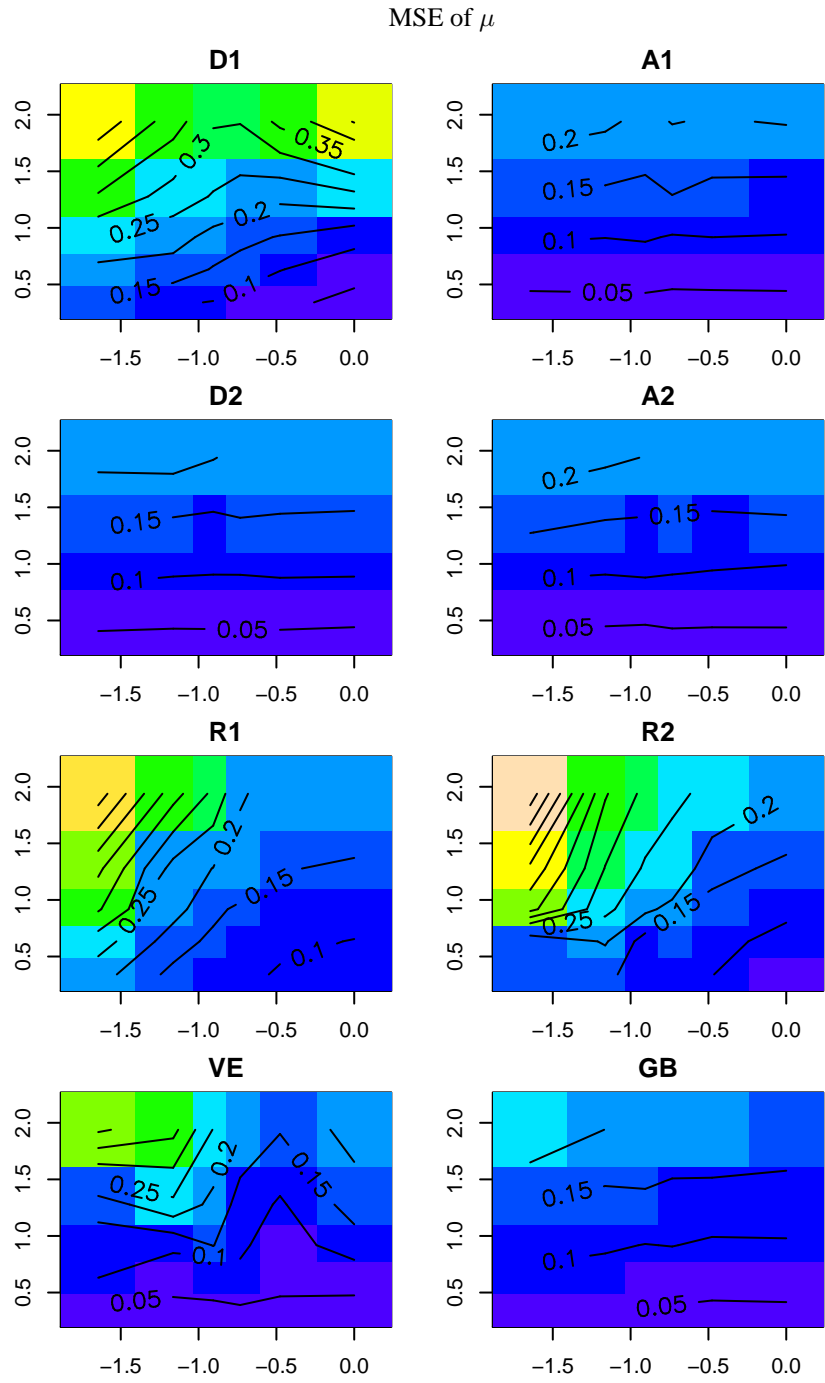


Figure 2: Contours of RMSE after 200 observations as a function of μ and σ for the eight methods tested (Section 3).

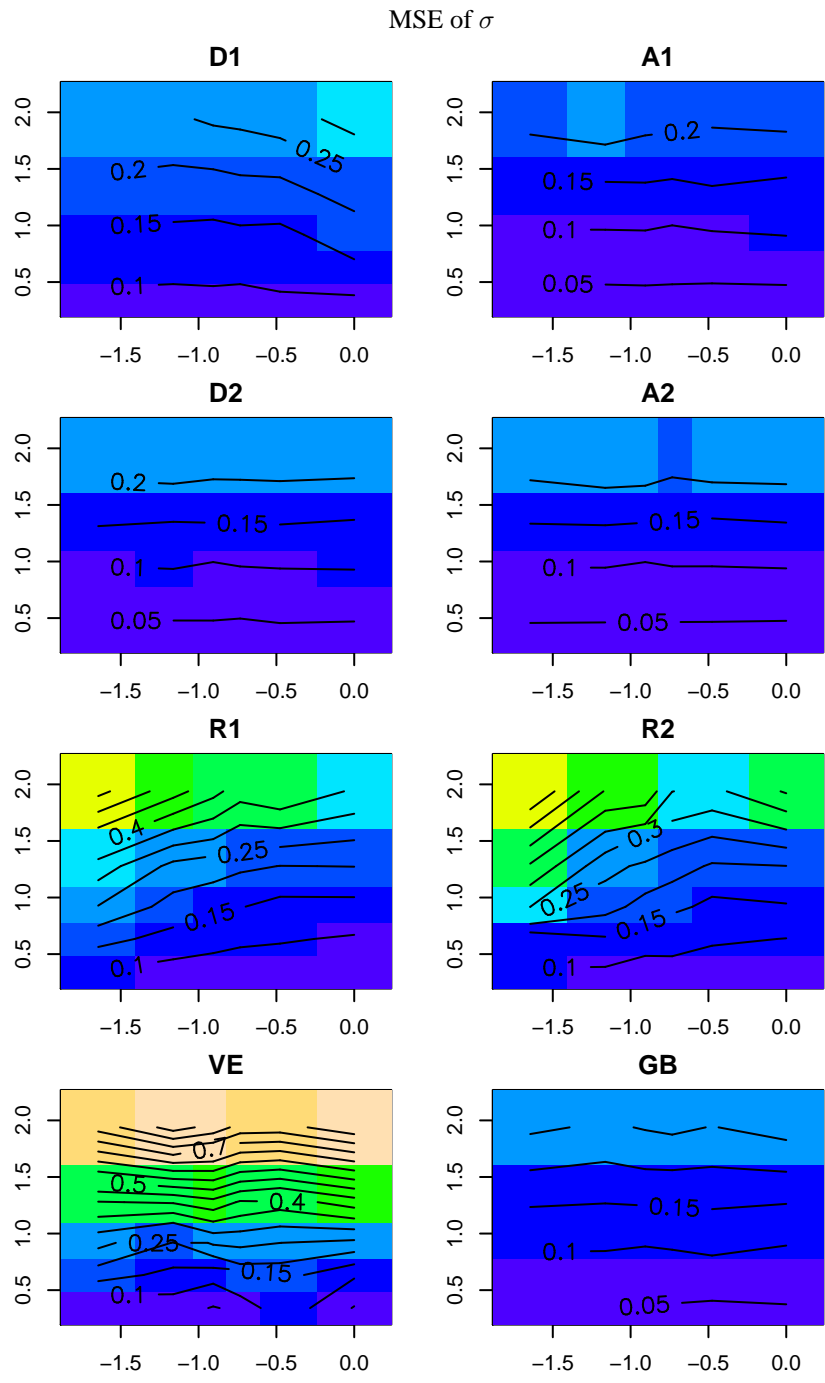


Figure 3: Contours of RMSE after 200 observations as a function of μ and σ for the eight methods tested (Section 3).

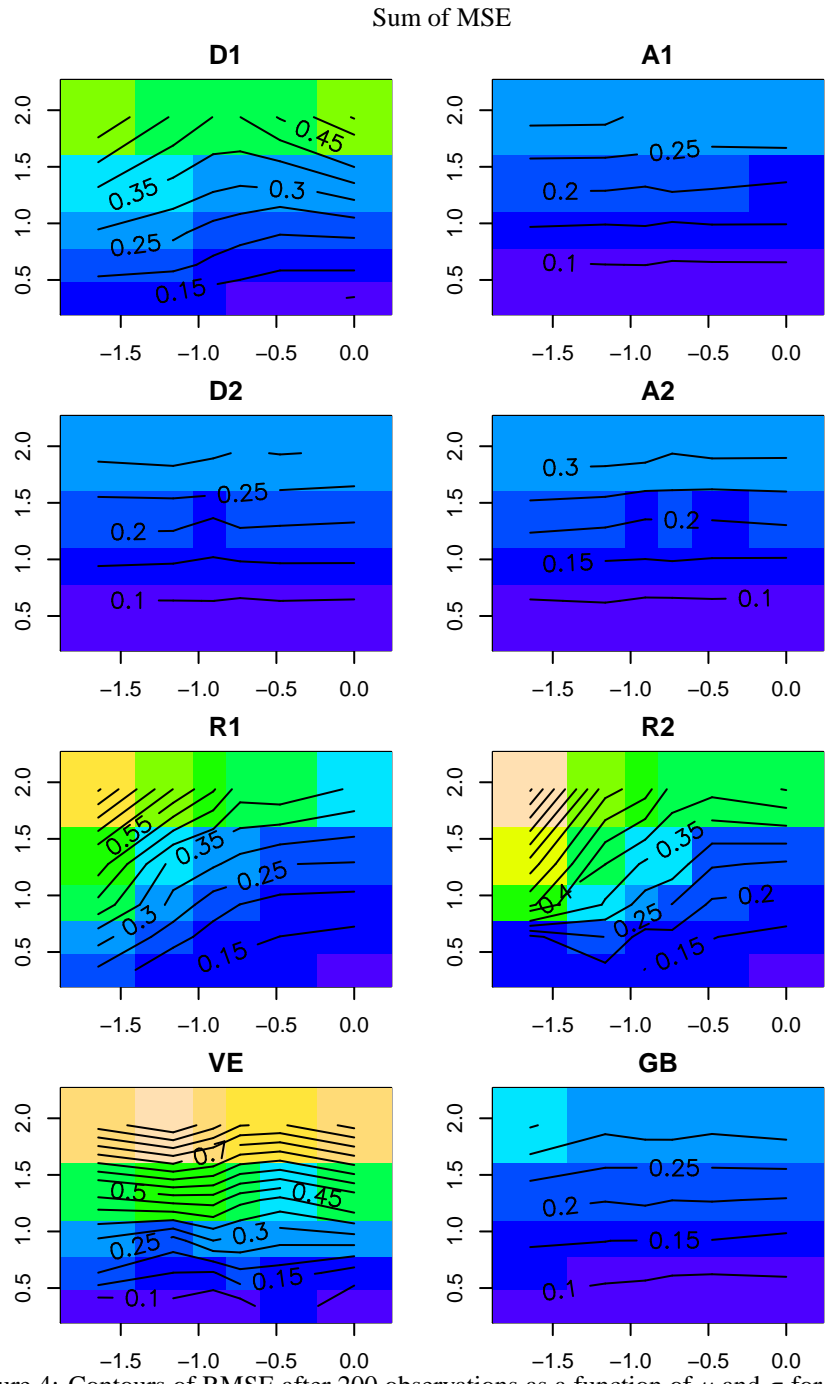


Figure 4: Contours of RMSE after 200 observations as a function of μ and σ for the eight methods tested (Section 3).

μ								
n	D1	A1	D2	A2	R1	R2	VC	GB
80	6.23	2.87	3.10	2.67	6.57	6.87	4.00	3.70
120	6.47	2.60	3.17	2.60	6.43	7.07	4.07	3.60
200	6.77	2.67	3.23	2.90	6.43	7.10	3.87	3.03
σ								
n	D1	A1	D2	A2	R1	R2	VC	GB
80	4.07	1.97	2.07	2.73	5.30	5.47	6.40	8.00
120	5.37	1.87	1.97	2.57	6.23	6.43	7.33	4.23
200	5.83	1.60	2.00	2.47	6.07	6.67	7.30	4.07
Sum								
n	D1	A1	D2	A2	R1	R2	VC	GB
80	4.73	1.67	2.10	2.37	5.33	5.77	6.03	8.00
120	5.97	1.67	2.27	2.33	6.33	6.67	6.90	3.87
200	6.13	1.70	2.20	2.27	6.23	6.80	6.73	3.93

Table 1: Average Rank of MSE for estimating μ , σ and the average rank of the sum of these MSE's. The average is taken over all 30 design points shown in Figure 3 and each MSE is calculated via simulation with 400 replicates. Low ranks correspond with low RMSE.

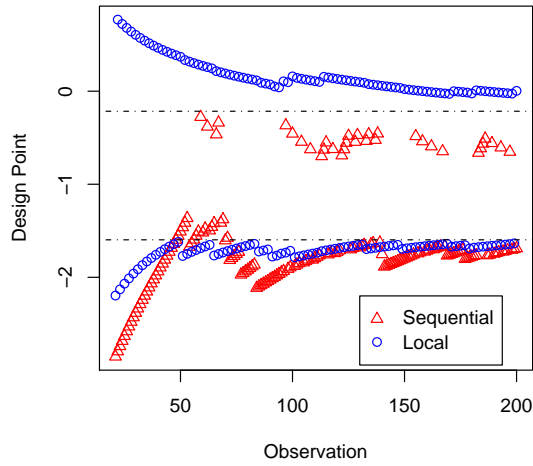


Figure 5: The design points for the locally A-optimal design are shown with the “o” symbol, and the sequential A-optimal design points are shown with \triangle where the initial design is a random from the prior distribution of μ and $\mu = -0.906$ and $\sigma = 0.634$. The dashed lines show the true A-optimal two-point design location.

5 Discussion

We have shown that locally optimal designs based on either D or A-optimality are efficient for binary response experiments aimed at estimating probit parameters. Implementing these designs for real-time data collection is with the modern computing abilities.

There is much to be done in the area of statistical design for body image testing. In this paper, and those referenced herein, design problems are formulated under the hypothesis that observations are independent. While this is certainly the case in many studies, it is clearly not so in body image testing. A more sophisticated model which incorporates learning and dependence among responses is needed.

Acknowledgments

The authors wish to thank Professor Rick Gardner for several comments and ideas that helped direct this work.

References

Chaloner, K. and K. Larntz, (1989), "Optimal Bayesian Design Applied to Logistic Regression Experiments," *Journal of Statistical Planning and Inference*, **21**, pp 191-208.

Chaloner, K. and I. Verdinelli, (1989), "Bayesian Experimental Design: A review," *Statistical Science*, **10**, pp 273-304.

Chung, K. L., (1954), "On Stochastic Approximation Method," *The Annals of Mathematical Statistics*, **25**, 463-483.

Clyde, M., Muller P. and Parmigiani, G. (1996). "Inference and Design Strategies for a hierarchical logistic regression Model." In: Berry, D. and Stangl, D. (Eds.), *Bayesian Biostatistics*, Marcel Dekker, New York.

Fuh, C. D., J. S. Lee and C. M. Liaw, (2003), "The Design Aspect of the Bruceton Test for Pyrotechnics Sensitivity Analysis," *Journal of Data Science*, **1**, pp. 83-101.

Gardner, R. M. and R. Boice, (2003), "A Computer Program for Measuring Body Size Distortion and Body Dissatisfaction," *Behavior Research Methods, Instruments, and Computers*, in press.

Khan, M. K. and A. A. Yazdi, (1998), "On D-Optimal Designs for Binary Data," *Journal of Statistical Planning and Inference*, **18**, pp. 83-91.

Natarajan R. and C. E. McCulloch, (1998), "Gibbs Sampling With Diffuse Proper Priors: A Valid Approach to Data Driven Inference?," *Journal of Computational and Graphical Statistics*, **7**, pp 267-277.

Sebastiani, P. and R. Settimi, (1997), "A Note on D-Optimal Designs for a Logistic Regression Model," *Journal of Statistical Planning and Inference*, **59**, pp. 359-368.

Sun, D. and R. K. Tsutakawa,(1997), “ Bayesian Design for Dose-Response Curves with Penalized Risk, ” *Biometrics*, **53** pp. 1262-1273.

Tsutakawa, R. K., (1967), “ Asymptotic Properties of the Block Up-and-Down Method in Bio-assay”, *The Annals of Mathematical Statistics*, Vol. 38, No. 6, pp. 1822-1828

Watt, R. J. and D. P. Andrews, (1983), “APE: Adaptive Probit Estimation of Psychometric Functions,” *Curr. Psych. Rev.*, **1**, pp. 205-214.