

Bayesian Sequential Experimental Design for Binary Response Data with Application to Electromyographic Experiments

Nammam Ali Azadi ^{*}, Paul Fearnhead [†], Gareth Ridall [‡], and Joleen H. Blok [§]

Abstract. We develop a sequential Monte Carlo approach for Bayesian analysis of the experimental design for binary response data. Our work is motivated by surface electromyographic (SEMG) experiments, which can be used to provide information about the functionality of subjects' motor units. These experiments involve a series of stimuli being applied to a motor unit, with whether or not the motor unit fires for each stimulus being recorded. The aim is to learn about how the probability of firing depends on the applied stimulus (the so-called stimulus-response curve). One such excitability parameter is an estimate of the stimulus level for which the motor unit has a 50% chance of firing. Within such an experiment we are able to choose the next stimulus level based on the past observations. We show how sequential Monte Carlo can be used to analyse such data in an online manner. We then use the current estimate of the posterior distribution in order to choose the next stimulus level. The aim is to select a stimulus level that minimises the expected loss of estimating a quantity, or quantities, of interest. We will apply this loss function to the estimates of target quantiles from the stimulus-response curve. Through simulation we show that this approach is more efficient than existing sequential design methods in terms of estimating the quantile(s) of interest. If applied in practice, it could reduce the length of SEMG experiments by a factor of three.

Keywords: Bayesian design, sequential design, motor unit, particle filtering, generalized linear model, binary response

We present an algorithm for adaptive design where efficient online estimation of parameters of a model is required for a given experiment. In classical experimental design, an optimality criterion is minimised to select optimal design points. For generalised linear models (GLMs), this results in design points which depend on the parameters that we wish to estimate (Khuri and Mukhopadhyay 2006). The classical approach to deal with this problem is to use a current estimate of the parameter to construct the design points. As a consequence, the design is only optimal for the given values of the parameters ('local optimal' design). This is often referred to as the design dependence problem (Khuri et al. 2006).

^{*}i)Department of Epidemiology and Biostatistics, Kurdistan University of Medical Sciences, Sanandaj, Iran. n.azadi@muk.ac.ir

ⁱⁱ)Department of Mathematics and Statistics, Lancaster University, UK

[†]Department of Mathematics and Statistics, Lancaster University, UK p.fearnhead@lancaster.ac.uk

[‡]Department of Mathematics and Statistics, Lancaster University, UK g.ridall@lancaster.ac.uk

[§]Department of Clinical Neurophysiology, University Medical Center Rotterdam, Netherlands j.h.blok@erasmusmc.nl

Bayesian experimental design ([Chaloner and Verdinelli 1995](#)) deals with this problem through the definition of a loss or utility function which is carefully chosen to match the statistical problem in hand. The design points are selected by minimising the expectation of this loss function over a prior distribution. This prior information expresses the uncertainty in the parameters for the data that has been collected so far. In the case of GLMs this expectation is intractable and therefore a good approximation is called for. In this paper we explore the strength of the particle approximation for this purpose.

Particle filters are able to carry out online estimation from systems which can be both non-linear and non-Gaussian. They have been used extensively in a variety of fields such as engineering, finance and genetics (see [Doucet et al. 2001](#), for examples). Most of this research, however, has been in the development of methods of estimation. Applications that use the particle filter for online decision making are much rarer (though see [Gramacy and Polson 2009](#); [Taddy et al. 2011](#), for some examples). The primary purpose of this paper is to illustrate how the particle filter can be used for real-time decision making. We believe that the particle filter is particularly suited for such problems of online decision making.

Our motivation for this problem comes from the field of clinical neurophysiology. In this field, electromyographic (EMG) methods are used to obtain diagnostic information about motor units, the physiological units that constitute our motor system. A motor unit (MU) consists of a single motor neuron cell body in the spinal cord, its long protruding axon, and the tens to thousands of muscle fibers that this neuron innervates. A full characterization of an MU would require an assessment of its morphology and of its excitability, contractile and conduction properties. In current clinical practice, only MU morphology is assessed routinely with needle EMG.

Studies using the so-called threshold tracking technique ([Bostock et al. 1998](#); [Burke et al. 2001](#)) have indicated that data on axonal excitability yield valuable clinical information that cannot be obtained otherwise. Because MUs are the building blocks of the motor system, gaining knowledge about single MU characteristics will increase our insight in the fundamental properties of this system and into pathological processes. It is, therefore, desirable to develop an approach by which excitability information on single MUs can be efficiently added to conventional EMG data.

The excitability of an axon is reflected in the intensity of the electrical stimulus (applied to the surface of the skin above the nerve) that is necessary to evoke an action potential in this axon. Each MU fires if the stimulus exceeds a random threshold, which at a very basic physiological level is determined by chaotic properties of the ion channels in the axon's membrane. Whether a response is present can be determined through monitoring of the muscle connected to the stimulated nerve using a surface electrode over the skin above this muscle (surface EMG or SEMG). If the axon of an MU is activated, so are all of the muscle fibers of this MU. The resulting motor unit action potential (MUAP) has a characteristic shape on SEMG, which is identical from firing to firing.

Excitability testing of a single MU requires the tracking of its threshold (usually defined as the stimulus intensity that elicits a response to 50% of the stimuli). This threshold can be determined by means of a stimulus-response curve. The stimuli are most effectively

delivered by an automated system, where the magnitude of each stimulus is governed by a computer program. At present this program merely sweeps over the complete range with a large fixed number of stimuli (roughly 400). Thus, one important question is how many and what stimuli are needed. This issue can be addressed by means of experimental design.

We describe a method which is able to employ iteratively the information made available by the incoming observations to select subsequent stimulus intensities (design points) using the principles of Bayesian experimental design. The advantages of such a sequential design are obvious. As observations are made, information gained from the data can be used to construct better and more efficient designs. This will reduce the number of applied stimuli and hence the discomfort for the subject and the examination time. Furthermore, Bayesian methods have the advantage that prior information can be used if it is available. In neurological experiments this information can be collected from historical studies.

In this paper we focus our attention on finding an arbitrary quantile, or quantiles, of interest of a single MU. Then through simulation studies we compare the performance of our method with several existing methods taken from the statistical design literature.

In Section 1 we describe existing methods for sequential design and discuss their limitations and advantages. In Section 2 we describe the nature and context of our dataset and use it both to illustrate our model and to elicit priors for the model. We also describe our Bayesian model in this section and discuss our method for updating the posterior and present an algorithm for choosing the appropriate control values (optimal stimulus intensities). We present the results of a comparative simulation study in Section 3. Finally in Section 4 we finish with an evaluation and a discussion.

1 Existing methods for sequential design

Sequential design for binary data can be traced back to the work of [Dixon and Mood \(1948\)](#) who introduced the so-called *up-and-down* procedure for estimating the dropping height at which an explosive specimen was equally likely to explode as to not. Using an initial guess of this dropping height, the successive specimen was tested at a lower height if the previous outcome was explosive, or at a greater height otherwise. The sequential procedure was continued until convergence.

In order to estimate the p th quantile, [Robbins and Monro \(1951\)](#) introduced a non-parametric sequential procedure for binary observations. A simulation study by [Wetherill \(1963\)](#) showed that while the Robbins-Monro procedure performed well for estimating the median, it had a poor performance with a large bias for estimating extreme quantiles. [Joseph \(2004\)](#) introduced a modification of the Robbins-Monro procedure which improved its performance for extreme quantiles.

Another improvement on the Robbins-Monro procedure is the logit maximum likelihood estimation (logit-MLE) approach of [Wu \(1985\)](#). This method introduces a parametric model for the response curve and calculates the maximum likelihood estimator of the

parameter of this model given current data. It then chooses the next design point as the p th quantile of this model, conditioning on the estimated parameter value. Even if the parametric model is incorrect this approach will produce a consistent estimator of the p th quantile, although the choice of model does affect the convergence rate. [Joseph et al. \(2007\)](#) extend this idea, to make the performance of the approach more robust to a poor choice of parametric model.

A known issue with these approaches is the uncertainty about choosing the optimal initial design points which are necessary to obtain the initial $\hat{\theta}_0$. A common approach is to try and use a set of 10 to 14 points symmetrically distributed around a guessed mean, but this can be difficult to accomplish when little or no information on the mean is available.

Other approaches to sequential design for binary data are based on calculating the optimal design points under some definition of optimality. For example, with the D-optimality criterion ([Chaudhuri and Mykland 1993](#); [Neyer 1994](#)) the aim is to minimise the determinant of the expected information matrix. The optimal design points can then be calculated in terms of the unknown parameters. Often the current MLEs for these parameters are used to give estimates of the optimal design points. As with the method of [Wu \(1985\)](#), implementing these procedures requires a good choice of the initial design points. Recent work has tried to incorporate both D-optimal designs and a Robbins-Monro procedure into an efficient general approach to sequential design ([Wu and Tian 2013](#)).

Partly in an attempt to resolve the issue of finding the best initial design, there has been recent interest in Bayesian methods for sequential design. [Dror and Steinberg \(2008\)](#) proposed a sequential two-stage Bayesian strategy. At the first stage an optimal initial design point is defined and then in the second stage the next design point is chosen as the one that gives the best outcome to a D-optimality criterion. Careful consideration is needed to avoid the non-singularity problem of the design in the early stages of the algorithm. [Dror and Steinberg \(2008\)](#) established a systematic procedure to find the optimal initial design point.

In this paper, we present a sequential Monte Carlo (SMC) technique that successfully eliminates the need for a feasible initial design. In this approach, the initial design points depend on the prior which describes the uncertainty in the parameter estimates in choosing design points. The prior is also useful for incorporating external information (from previous experiments on motor units from other subjects). Sequential Monte Carlo is used to recursively calculate the posterior distribution of the parameters as each new observation is made. Using the current posterior distribution, we can then choose the next stimulus value so as to minimise an appropriate expected loss function. This loss function is related to the purpose of the statistical analysis.

2 Our proposed method

In this section we first give a description of our application, followed by a description of our Bayesian model. The section continues with a discussion of the principles of Bayesian experimental design and introduces the notation and mechanism by which sequential Markov chain Monte Carlo (MCMC) is carried out. Finally the section ends with a description of the algorithm which defines our optimal data collection mechanism.

2.1 SEMG experiment

As described earlier, SEMG is a non-invasive technique which records the nerve's response to a stimulus using a surface electrode placed directly on the skin overlaying the muscle. Figure 1 shows an example data set. This data presents the amplitude of the signals recorded from the thenar (thumb) muscles in response to around 400 stimuli, which were applied to the median nerve. The intensity of the stimuli was low enough to ensure that the recording was of just a single MU, as evidenced by the all-or-nothing (binary) character of the response. The stimulus-response behaviour of an MU can be described by parameters which are henceforth referred to as MU excitability properties (Ridall et al. 2006).

In this experiment, the minimum and maximum stimulus intensities are chosen to include the stimulus values over which the MU's response is stochastic. Electrical stimulation to the nerve is then gradually increased from this minimum to the maximum stimulus intensity. The left hand panel of Figure 1 shows 400 actually observed amplitudes recorded at 82 distinct stimulus intensities. The stimuli were applied at a rate of 2 Hz. The horizontal scale represents the stimulus intensity values, which vary from $s_{\min} = 8.6$ mA to $s_{\max} = 9.4$ mA and the vertical scale represents the signal amplitude in μV . To describe the all-or-nothing state of the MU, in the right hand panel the same data are presented after conversion to binary responses by choosing an appropriate threshold.

2.2 Model specification

Our approach can be used with any parameterisation of the stimulus-response curve. For our SEMG experiment previous data has suggested that a logistic curve is appropriate (Azadi 2011), and we focus on such a curve in the following.

Let $\mathbf{s} = s_{1:n} = (s_1, \dots, s_n)$ denote the vector of fixed and known stimulus levels and $\mathbf{y} = y_{1:n} = (y_1, y_2, \dots, y_n)$ denote the corresponding binary observed responses. Given the stimulus values we assume the observations are realizations of independent Bernoulli distributions with success probabilities depending on stimulus values via the logistic function:

$$\Pr(Y_i = 1 | s_i, m, b) = \frac{1}{1 + \exp\{-b(s_i - m)\}} \quad i = 1, 2, \dots, n. \quad (1)$$

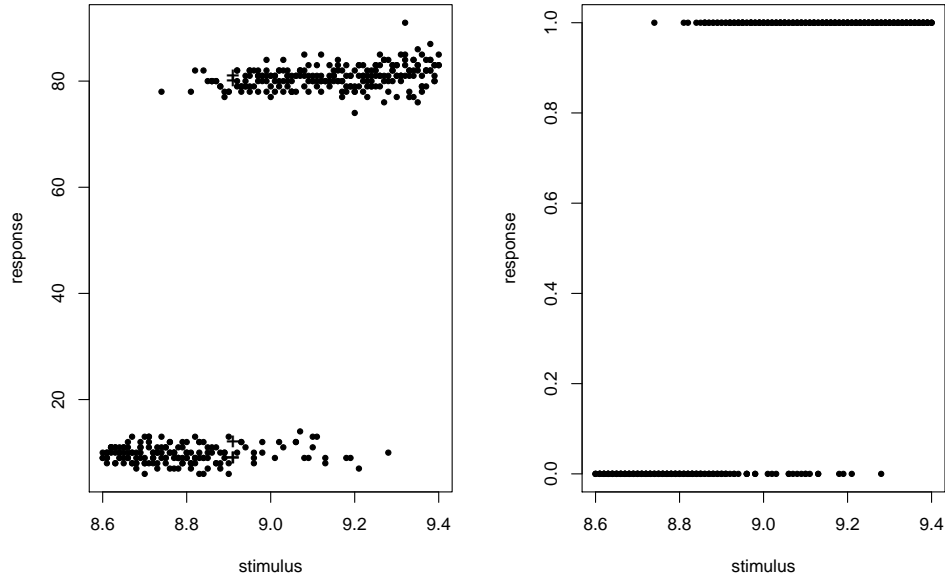


Figure 1: Stimulus-response plots. The left panel is a scatter plot showing the amplitude of the measured signal against stimulus intensity, recorded using surface electromyography. The low-level amplitudes around $10\mu V$ represent background noise (no response), and the values around $80\mu V$ denote the amplitude of the action potential of the motor unit for activation. The right hand panel shows the all or nothing state of the motor unit response after thresholding has been applied.

Here m denotes the mean threshold (the stimulus for which an MU has a 50% chance of responding) for the MU. This is known by some as the location parameter of the logistic-response curve. The slope parameter b of the logistic curve is a measure of the range of stimulus values for which the firing MU shows stochastic behaviour (the unit fires some of the time).

2.3 Bayesian analysis

To perform inference for the parameters $\theta = (m, b)$, we introduce a prior, $p(\theta)$. If the likelihood of the first t SEMG observations is given by $p(y_{1:t}|\theta, s_{1:t})$, then the posterior density can be expressed as

$$p(\theta|y_{1:t}, s_{1:t}) \propto p(y_{1:t}|\theta, s_{1:t})p(\theta). \quad (2)$$

In our simulations we use a prior for $p(\theta)$ that is mildly informative. Prior information can be obtained from the results of the analysis of earlier SEMG experiments. The

posterior density (2) is updated as observations arrive in real time. The relationship between the posterior after $t + 1$ observations and that after t observations can be expressed as:

$$p(\theta|y_{1:t+1}, s_{1:t+1}) \propto p(y_{t+1}|\theta, s_{t+1})p(\theta|y_{1:t}, s_{1:t}). \quad (3)$$

That is the new posterior density at time $t + 1$ is the likelihood of the new observation, y_{t+1} , multiplied by the old posterior at time t . The likelihood for the new observation is obtained from (1) which represents the response of the MU after having being exposed to the current stimulus intensity s_{t+1} .

2.4 The particle approximation to the posterior

Due to the non-linearity of the logistic transformation, the posterior (3) becomes difficult to sample from directly and an approximation is called for. Sequential Monte Carlo (SMC) methods are a set of online techniques that provide a flexible framework for the updating of posterior distributions in real time (see Liu and Chen 1998; Doucet et al. 2001, 2000, for an introduction). The idea of SMC is to make an approximation of the posterior using a set of weighted particles, $\{\theta_t^{(i)}, w_t^{(i)}\}$, $i = 1, 2, \dots, N$. The approximation is a discrete distribution whose support points are the set of particles, with the probability assigned to a particular support point, $\theta_t^{(i)}$, being proportional to the weight associated with the corresponding particle, $w_t^{(i)}$. SMC algorithms then determine how to generate the set of weighted particles at time $t + 1$ from those at time t .

Given a set of weighted particles at time t , by using (3) we get the following approximation to the posterior at time $t + 1$,

$$p(\theta|y_{1:t+1}, s_{1:t+1}) \approx \frac{\sum_{i=1}^N w_t^{(i)} p(y_{t+1}|\theta_t^{(i)}, s_{t+1}) \delta_{\theta_t^{(i)}}(\theta)}{\sum_{i=1}^N w_t^{(i)} p(y_{t+1}|\theta_t^{(i)}, s_{t+1})}, \quad (4)$$

where $\delta_{\theta_t^{(i)}}(\theta)$ is a point-mass at $\theta = \theta_t^{(i)}$. This gives a simple recursion where the particles are fixed, $\theta_{t+1}^{(i)} = \theta_t^{(i)}$ for $i = 1, \dots, N$, and the change in the posterior distribution is captured by the update of the weights. The approximation is initialised by generating N draws from prior densities of parameters θ . The initial weight $1/N$ is considered for each draw and then, following a new observation, the weights are renewed according to the likelihood of the new observation. At time $t + 1$ the weights are updated as $w_{t+1}^{(i)} \propto w_t^{(i)} p(y_{t+1}|\theta_t^{(i)}, s_{t+1})$. Algorithm 1 outlines this approach. It should be noted that rather than using a random sample from the prior, we can get more accurate results using a stratified sample or with a quasi-Monte Carlo sample (Fearnhead 2005; Dror and Steinberg 2008).

The problem with Algorithm 1 is that after a number of iterations a large number of updated weights become negligible (Petris et al. 2009). This makes the algorithm inefficient and reduces the accuracy of the particle approximation. We use two strategies

Algorithm 1 : Bayesian updating using the particle approximation

Input: A set of stimuli values (s_1, s_2, \dots, s_n) where $s_i \in \mathcal{S}$.
 A set of observations (y_1, y_2, \dots, y_n) where y_i is the response of MU at stimulus s_i .
 A prior, $p(\theta)$, for $\theta = (m, b)$.
 The number of particles N .

Initialise: Generate N particles, $\theta_0^{(1)}, \dots, \theta_0^{(N)}$, from prior $p(\theta)$.
 Assign to each particle the initial weight $w_0^{(i)} = 1/N$ for $i = 1, \dots, N$.

Iterate For $t = 1, 2, \dots, n$:

(i) Define

$$p(Y_t = 1 | s_t, \theta_0^{(i)}) = \frac{1}{1 + \exp\{-b_0^{(i)}(s_t - m_0^{(i)})\}},$$

$$\text{and } p(Y_t = 0 | s_t, \theta_0^{(i)}) = 1 - p(Y_t = 1 | s_t, \theta_0^{(i)}).$$

(ii) For the new observation y_t , update the weights using

$$w_t^{*(i)} \propto w_{t-1}^{(i)} \times p(Y_t = y_t | s_t, \theta_0^{(i)}).$$

(iii) Normalize the new weights $w_t^{(i)} = \frac{w_t^{*(i)}}{\sum_{j=1}^N w_t^{*(j)}}$.

Output A set of weighted particles that approximate $p(\theta | y_{1:n}, s_{1:n})$.

to address this problem. Firstly, to speed up the algorithm and save unnecessary computations, we remove particles with negligible weights at the end of each iteration. We can do this in a principled way using rejection control (Liu et al. 1998). This involves setting a threshold ε_0 . Any particle whose weight, $w_t^{(i)}$, drops below ε_0 is removed with probability $w_t^{(i)}/\varepsilon_0$; otherwise its weight is increased to ε_0 . This procedure removes particles, but ensures the algorithm remains asymptotically unbiased as $N \rightarrow \infty$. In practice we set the threshold at $\varepsilon_0 = N/10$. Secondly, we monitor the effective sample size (ESS) (Liu et al. 1998) which is defined as

$$N_{ESS} = \frac{1}{\sum_{i=1}^N (w_t^{(i)})^2}.$$

If the effective sample size drops below a pre-defined threshold N_0 , a resample and refresh step is used to rejuvenate or jitter the particles (Liu and West 2001). The resampling step on its own, no matter which kind of resampling is used, will only replace a large number of particles, some with very small weights, by a small number of distinct particles. This is referred to as the problem of particle depletion. To address the problem of the lack of particle diversity, Liu and West (2001) suggest transforming and jittering the particles from a normal kernel density in such a way that the first two moments of the posterior distribution are preserved. Resampling is performed by drawing particle values from a kernel density estimate of the posterior. This estimate

takes the form of the following multivariate normal:

$$p(\theta|y_t, s_t) \propto \sum_{i=1}^N w_t^{(i)} \mathcal{N}(\theta; M_t^{(i)}, \Sigma_t), \tag{5}$$

where $M_t^{(i)} = a\theta_{t-1}^{(i)} + (1-a)\bar{\theta}_t$, $\Sigma_t = hV_t$, and $\bar{\theta}_t$ and V_t are the current estimates of, respectively, the mean and variance of the posterior. The parameters a and h are chosen to satisfy the equation $a^2 + h^2 = 1$, so that the mean and variance of this kernel approximation is equal to the current estimate of the mean and variance of the posterior. Liu and West (2001) suggest that the parameter a should lie in the interval 0.974 to 0.995 for the resampling to be efficient. The details of the SMC algorithm with the resampling step are given in Algorithm 2.

Algorithm 2 : Sequential Monte Carlo algorithm with resampling

Input: A constant a , with $0 < a < 1$, and threshold for resampling N_0 . Set $h = \sqrt{1-a^2}$. Add steps (4)--(6) to each iteration of Algorithm 1.

4. For $i = 1, \dots, N$, if $w_t^{(i)} < \varepsilon_0$ then with probability $w_t^{(i)}/\varepsilon_0$ remove particle i , otherwise set $w_t^{(i)} = \varepsilon_0$. Renormalise the weights.
5. Compute $N_{ESS} = \left(\sum_{i=1}^N (w_t^{(i)})^2 \right)^{-1}$.
6. If $N_{ESS} > N_0$ then set $\theta_t^{(i)} = \theta_t^{(i-1)}$; otherwise:
 - Calculate the posterior mean and variance from the current set of weighted particles; denote these $\bar{\theta}_t$ and V_t .
 - Draw N new particles, $\{\theta_t^{(i)}\}_{i=1}^N$, from

$$p(\theta|y_t, s_t) \propto \sum_{i=1}^N w_t^{(i)} \mathcal{N}(\theta; M_t^{(i)}, \Sigma_t),$$

where $M_t^{(i)} = a\theta_{t-1}^{(i)} + (1-a)\bar{\theta}_t$ and $\Sigma_t = hV_t$.

- Assign to particles the new weight $1/N$.
-

2.5 Sequential design procedure

We now focus on how to choose a new stimulus, s_{t+1} , given the observations to date, $y_{1:t}$ and $s_{1:t}$. For $j = 1, \dots, J$, let $\phi_j = \phi_j(\theta)$ denote a function of the parameters that is of interest, such as a quantile. To choose the best new stimulus value we minimise the expected value of a loss function for estimating ϕ_j for $j = 1, \dots, J$. For the sake of simplicity we focus on a quadratic loss. Let $\mathbf{CE} = (\phi_1, \dots, \phi_J)$, θ_0 be the true parameter values and $\phi_0 = \phi(\theta_0)$. Then our loss associated with an estimate $\hat{\phi}$ is

$$L(\hat{\mathbf{CE}}, \mathbf{CE}_0) = \sum_{j=1}^J \gamma_j (\phi_j(\theta_0) - \hat{\phi}_j)^2. \tag{6}$$

This is the sum of quadratic losses for estimating each ϕ_j , with $\{\gamma_j\}_{j=1}^J$ being a set of positive constants that define the relative weight given to the error in estimating different ϕ_j s. It can be easily shown that the estimate that minimises the quadratic loss for ϕ_j is the posterior mean. The corresponding expected loss is a weighted average of the posterior variance for each $\phi_j(\theta)$.

To select the optimal stimulus intensities s_1, s_2, \dots, s_T , we want to construct a rule which if applied recursively minimises the expected posterior variance of ϕ . The minimisation is with respect to the rule that, given the observations and past stimuli, selects the next stimulus value. The expectation is with respect to both the parameter (from the prior) and observations we may observe. Such an approach has been called Bayesian global optimization (Mockus 1989). Readers are referred to Benassi et al. (2011) for a fully Bayesian formulation of the approach.

Obtaining such an optimal rule is intractable. Instead we propose choosing s_{t+1} to minimise the expected variance of ϕ after the $(t+1)$ st observations (see Voelkel 1999, for a related approach). That is, given $s_{1:t}$ and $y_{1:t}$, we want to choose

$$s_{t+1} = \arg \min_s \left\{ \mathbb{E}_{Y_{t+1}} \left(\sum_{j=1}^J \gamma_j^2 \text{Var}(\phi_j | y_{1:t}, Y_{t+1}, s_{1:t}, s) \right) \right\}. \quad (7)$$

As observations are binary, for any value of s we can easily evaluate the right-hand side of (7) in terms of an expectation with respect to the posterior distribution of θ given $s_{1:t}$ and $y_{1:t}$. In particular we obtain, for each term in the sum,

$$\begin{aligned} \mathbb{E}_{Y_{t+1}}[\text{Var}(\phi_j | y_{1:t}, Y_{t+1}, s_{1:t}, s)] &= \Pr(Y_{t+1} = 0 | s_{1:t}, y_{1:t}, s) \text{Var}(\phi_j | s_{1:t}, y_{1:t}, s, Y_{t+1} = 0) \\ &\quad + \Pr(Y_{t+1} = 1 | s_{1:t}, y_{1:t}, s) \text{Var}(\phi_j | s_{1:t}, y_{1:t}, s, Y_{t+1} = 1), \end{aligned}$$

and the terms on the right-hand side can be estimated using our weighted particles. The details of how to calculate the expected loss associated with a given stimulus value is given in Algorithm 3. We numerically minimise this expected loss over stimulus values to find the optimal stimulus value to use at the next iteration.

It is straightforward to extend this approach to account for some other loss functions, such as mean absolute error loss, for which the optimal Bayesian estimators for ϕ_j are known. For other loss functions we would need extra computation devoted to calculating the optimal estimators of ϕ_j , and hence the minimum expected loss at the next time step, for each possible next observation value.

3 Simulation studies

Here we describe the details of our simulations which compare the efficiency of several methods for finding percentiles of a stimulus-response curve. A brief description of each method is given below.

Algorithm 3 : Algorithm for calculating expected loss

Input:

- A set of particles, $\{\theta^{(i)}\}_{i=1}^N$, and normalised weights, $\{w^{(i)}\}$, where $\theta^{(i)} = (m^{(i)}, b^{(i)})$.
- A set of quantities of interest $\phi_j(\theta)$, for $j = 1, \dots, J$.
- A set of weights γ_j , $j = 1, \dots, J$.
- A stimulus value s .

(i) For $i = 1, \dots, N$:

(a) Calculate

$$p(Y = 1|s, \theta^{(i)}) = \frac{1}{1 + \exp\{-b^{(i)}(s - m^{(i)})\}},$$

and $p(Y = 0|s, \theta^{(i)}) = 1 - p(Y = 1|s, \theta^{(i)})$.

(b) Update the weights $w_y^{(i)} = w^{(i)}p(Y = y|s, \theta^{(i)})$ for $y = 0, 1$.

(ii) Calculate $p_1 = \sum_{i=1}^N w_1^{(i)}$ and $p_0 = 1 - p_1$ as the probability of response and non-response respectively.

(iii) For $j = 1, \dots, J$, and $y = 0, 1$, compute the posterior mean of $\phi_j(\theta)$:

$$\bar{\phi}_{j,y} = \frac{1}{p_y} \sum_{i=1}^N w_y^{(i)} \phi_j(\theta^{(i)})$$

and the posterior variances

$$\Sigma_{j,y} = \frac{1}{p_y} \sum_{i=1}^N w_y^{(i)} (\phi_j(\theta^{(i)}) - \bar{\phi}_{j,y})^2.$$

(iv) Estimate the expected loss by

$$E(Loss) = \sum_{j=1}^J \gamma_j^2 (p_0 \Sigma_{j,0} + p_1 \Sigma_{j,1}).$$

Output: Output the estimated minimum expected loss associated with stimulus s : $E(Loss)$.

N-Opt Non-Optimal: To act as a baseline comparison, stimuli are taken sequentially from smallest to largest intensity without any optimization being carried out. This is the current approach used in SEMG experiments.

New Our method which uses a quadratic loss function to select optimal stimulus levels. We implement three versions of this, corresponding to loss functions based on targeting different percentiles: *New 0.5* for estimating the 50th percentile; *New 0.75* for estimating the 75th percentile; and *New (0.5,0.75)* for a loss defined as an equal-weighted sum of the losses for estimating the 50th and 75th percentiles.

2DP 2-point augmentation design (Mathew and Sinha 2001): In this method, at each iteration, two stimuli are selected symmetrically around the current estimate of a target quantity according to the expression $s_i = \hat{m} \pm c/\hat{b}$, $i = 1, 2$. The authors derive, in closed form, an expression for the value of $c > 0$ that maximises the determinant of the joint information matrix of logistic regression parameters (D-optimality criterion) assuming \hat{m} and \hat{b} are the true parameter values.

L-MLE The logit MLE method of Wu (1985). If we are targeting the p th quantile, then the next stimulus, S_{t+1} , is chosen to satisfy the relationship $S_{t+1} = \hat{m}_t - \hat{b}_t^{-1} \log(p^{-1} - 1)$ where \hat{m}_t and \hat{b}_t are the MLEs of the logistic parameters at time t and p is the percentile of interest. Instead of using the MLEs, we use the current estimates of the posterior means. We implement two versions: *L-MLE 0.5* for estimating the 50th percentile; and *L-MLE 0.75* for estimating the 75th percentile.

DS The Dror and Steinberg (2008) algorithm. Here, stimulus values are selected sequentially that maximise an expectation of a D-optimal criterion with respect to the current posterior for the parameters.

The accuracy of the different methods is judged based on the mean square error of the estimates of the target quantile, averaged over results from analysing 400 simulated data sets. In each case we plot a relative accuracy of estimates of quantiles as a function of the number of observations. This is a ratio of the mean square error of the best performing method after analysing 200 observations, to the mean square error of the current method with t observations.

When implementing the sequential Monte Carlo algorithm we used $N = 4,000$, $\varepsilon_0 = N/10$, $a = 0.9$ and we resampled when the effective sample size was less than 250. Similar results were obtained using larger N , smaller ε_0 and larger a .

We first considered inference under a known logistic response curve. This has been shown to be an appropriate model for SEMG data. We analysed simulated data, using priors that are consistent with SEMG data. We then considered how the performance of each algorithm is affected by the information in the prior, and how robust methods are to deviations from a logistic response curve. We finally look at applying these methods to the case where we model the response curve as skew-logistic.

SEMG-based simulations

As mentioned, we can use results from previous SEMG experiments to construct an appropriate prior distribution. We have results from the analysis of eight SEMG experiments which we used to construct a prior for b . The maximum likelihood estimates of b for these experiments ranged between 9 and 36. For simplicity we chose a prior distribution for $\log(b)$ that is uniform on $(\log(5), \log(200))$.

Whilst we could take a similar approach to get a prior for m , in practice there is an initial stage to SEMG experiments. Most motor units have thresholds that are very close to one another, too close in fact to be able to be suitable for the SEMG experiment. This initial stage involves searching for a unit that remains isolated even during manipulation of its threshold (e.g., by adjusting stimulus duration or applying a superimposed hyperpolarizing current). Because of this search process there is a good initial guess of m for the motor unit under experimentation, and this information can give a relatively informative prior for m . In our simulations we took this prior to be normally distributed with mean 8 and variance 0.2², which is an appropriate level of uncertainty in m after this initial search phase.

In Figure 2 we show the choice of stimulus for different methods for one data set. There are similarities between our new method and the logit-MLE method when targeting the same quantile. For both cases the choice of stimulus values is concentrated around estimates of the quantile of interest. By comparison the method of Dror and Steinberg alternates between high and low quantiles. Its choice of stimulus values is very similar to the 2-point augmentation (results not shown), which is based on alternately choosing stimulus values that are current estimates of the 18th and 82nd quantile. Results

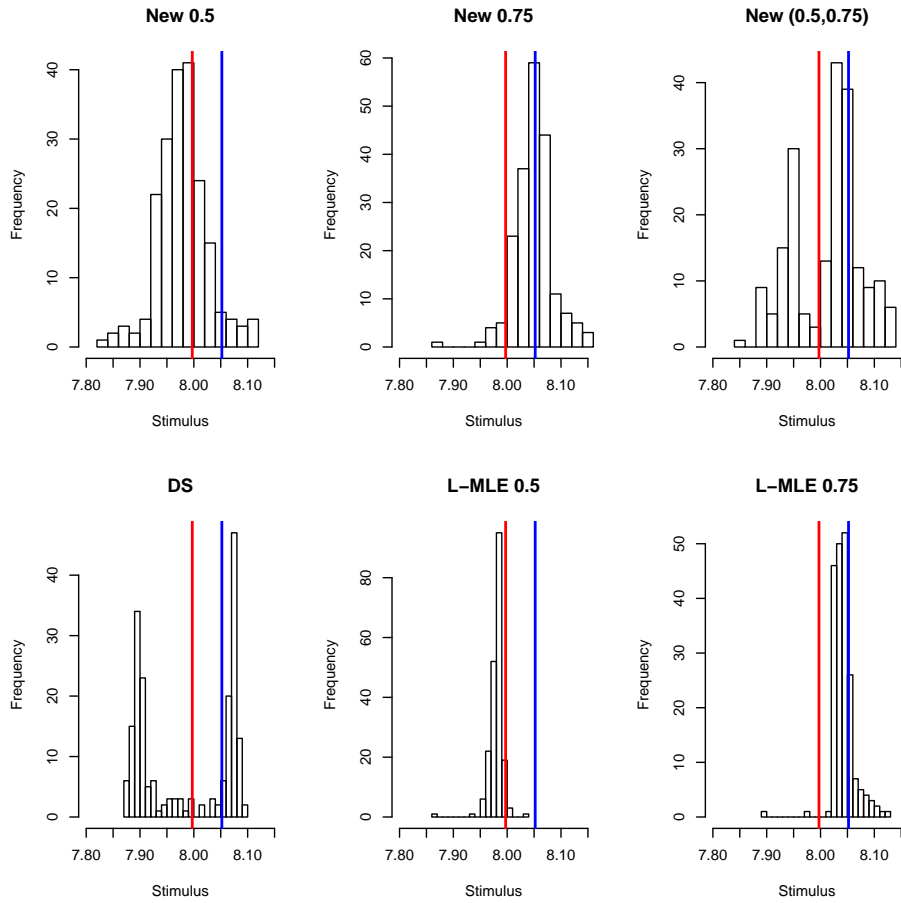


Figure 2: An example of the stimulus chosen by 6 of the different methods. Vertical lines show the true values of the 50th and 75th percentiles.

of the relative accuracy of the different designs for estimating the 50th, 75th and 95th percentiles are given in Figure 3. The new method performs best in terms of estimating the quantile that is the focus of the design. However, if we implement the new design for estimating one quantile, it performs worse than other methods, such as the 2 data-point

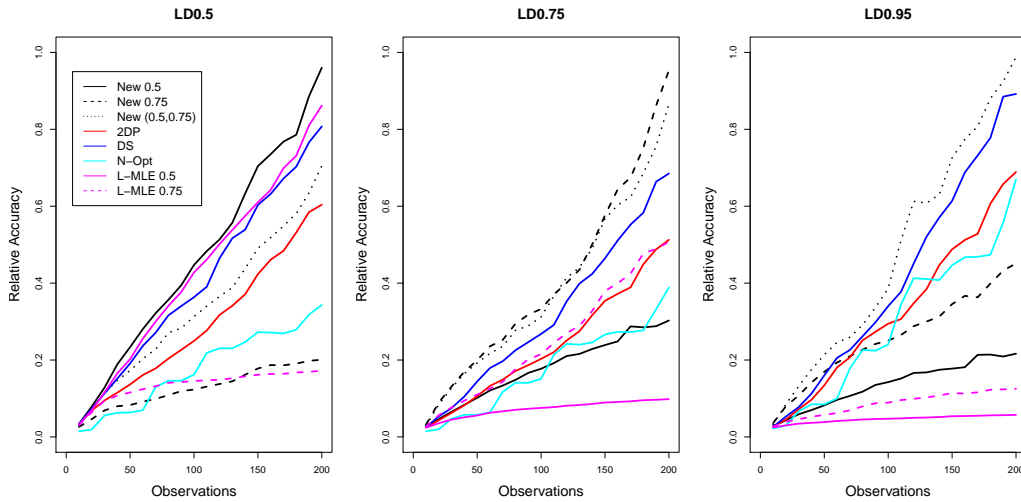


Figure 3: Relative accuracy of different methods at estimating 50^{th} (left), 75^{th} (middle) and 95^{th} (right-hand figure) percentile of the response curve. Prior variance on m is 0.2^2 .

augmentation (2DP), or the method of Dror and Steinberg (DS) in terms of accuracy of estimates of other quantiles. Similar behaviour is noticed for the logit-MLE approach, but the new design method out-performs the logit-MLE approach in all cases. By implementing the new method, but targeting two quantiles, we get similar performance to the 2DP and DS methods. Note that the current method used in SEMG (N-Opt) performs poorly for estimating the 50^{th} and 75^{th} quantiles. In practice it is estimates of the 50^{th} quantile that are of main interest to those conducting SEMG experiments. Our method gives a roughly 3-fold increase in accuracy over N-Opt in that case.

Effect of prior

We next looked at how the results we obtained were robust to the choice of prior. We repeated our analysis with a larger prior variance on m . Results for a prior variance of 2^2 are shown in Figure 4. These results are very similar to those in Figure 3 for the substantially more informative prior. We also looked at how the accuracy of inferences depended on the true value of b , and the prior for b . We got similar results to those in Figures 3 and 4 in all cases we investigated (results not shown).

Robustness to model error

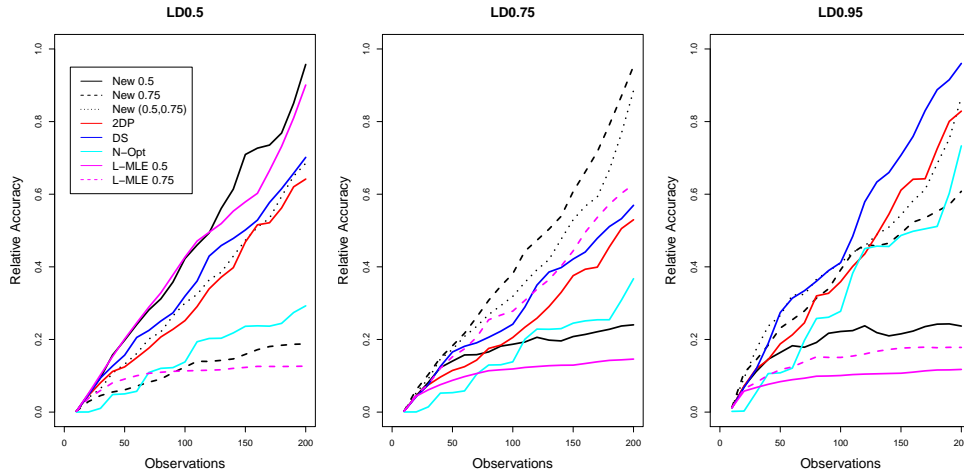


Figure 4: Relative accuracy of different methods at estimating 50th (left), 75th (middle) and 95th (right-hand figure) percentile of the response curve. Prior variance on m is 2^2 .

Next we investigated the robustness of inferences based on different designs to model error. We simulated data from a skew-logistic model

$$p(Y_t = 1 | s_t, \theta) = \left(\frac{1}{1 + \exp\{-b(s_t - m)\}} \right)^g, \tag{8}$$

but analysed the data under the assumption of a logistic response curve. Methods such as logit-MLE are known to be robust to model errors, at least in terms of the quantile that is the target of the design. The value of g in this model determines the amount of skewness in the response (curve) with $g = 1$ being the logistic curve, and values of g away from 1 producing increasingly skewed response curves. Figure 5 gives results for $g = 2$ (similar results were observed for $g = 1.5$ and $g = 1.2$). The prior variance for m was 1, and we again assumed a uniform prior on $\log b$ with range $[\log 5, \log 200]$. We again see similar patterns in terms of accuracy, with the new method performing best in terms of estimating the quantile that it targets, the performance of the new method targeting two quantiles more robust across a range of quantiles, and the DS and 2DP methods more accurate for estimating other quantiles. In terms of the bias of methods due to the model mis-specification, we see that the methods that target specific quantiles tend to have smaller bias for estimating those quantiles. This is linked to these methods choosing stimulus values close to the quantile of interest (see Figure 2).

Application to skew-logistic model

We finally extended our analysis to an inference for the skew-logistic model (8), and allowed for model uncertainty. With probability 0.5 $g = 1$, and the response curve is logistic; otherwise we had a uniform prior on $\log g$ with range $[-1, 1]$. We implemented

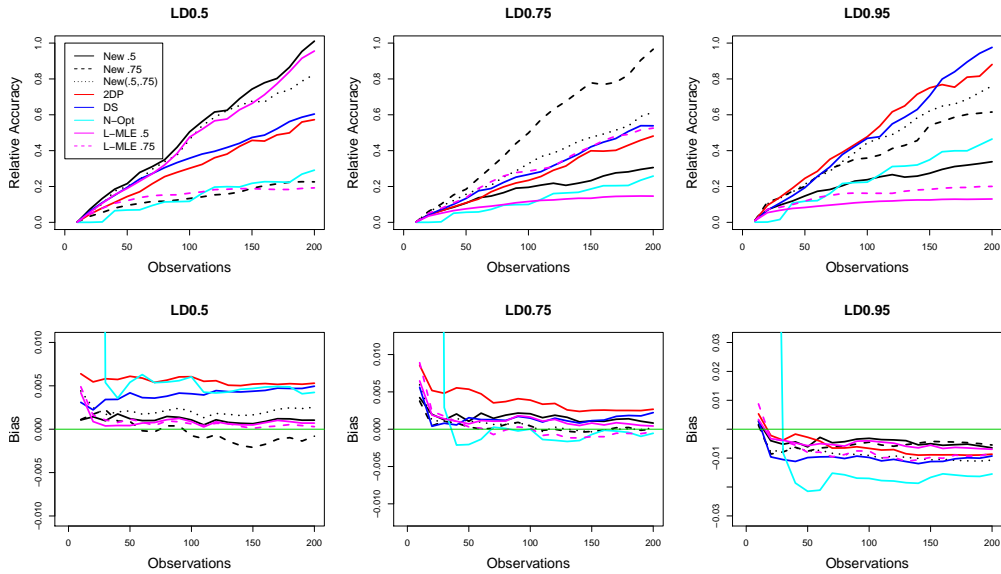


Figure 5: Relative accuracy (top) and Bias (bottom) of different methods at estimating 50th (left), 75th (middle) and 95th (right-hand figure) percentile of the skew-logistic response curve with $g = 2$. Inferences are based on the incorrect assumption that the response curve is logistic.

the new method, the logit-MLE, both of which are simple to apply in this scenario, and an extension of the 2DP approach to the skew-logistic model. This extension involves numerically calculating the optimal design for the skew-logistic model for a grid of g values. This design depends on the unknown parameters of the model, and we implement the design based on our current estimates of these parameters. The optimal design for the skew-logistic model was a 3-point design for all values of g , and involved alternating between stimulus values at a low, middle and high quantile value with the precise value of these quantiles depending on g . We denote this method by 3DP. Results are given in Figure 6. We see a similar pattern to before. Both the new method and logit-MLE do well at estimating the quantile their design is implemented for, but substantially less well for other quantiles. The new method outperforms logit-MLE. Both 3DP and the new method targeting two quantiles give a more robust performance across different quantiles, with 3DP best for the 95th quantile: one which none of the other designs target.

4 Conclusions

In this paper we presented an algorithm for carrying out sequential online design of experiments of binary data when an accurate estimate of a target quantile is sought.

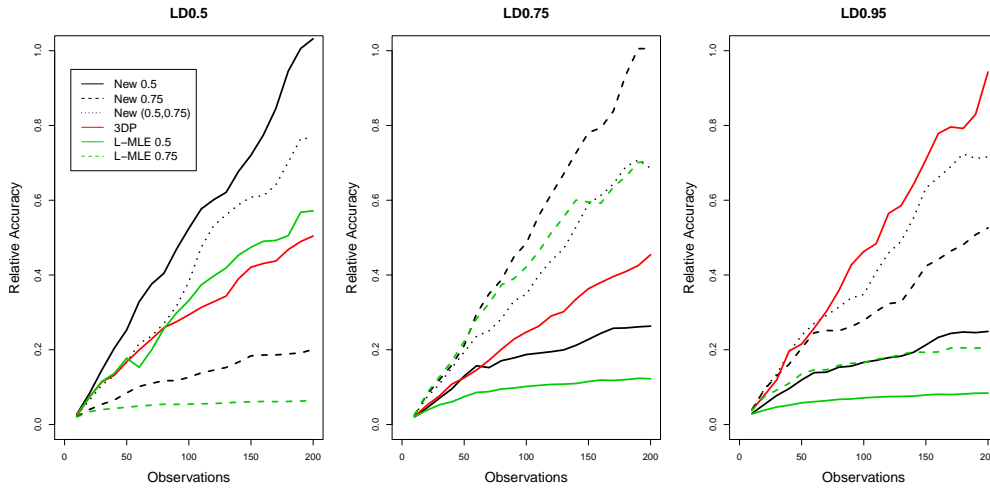


Figure 6: Relative accuracy of different methods at estimating 50^{th} (left), 75^{th} (middle) and 95^{th} (right-hand figure) percentile of the response curve in the case of model uncertainty.

Our simulation results suggest that our approach is more efficient than any of the alternative methods at estimating this target quantile. Methods based on, for example, D-optimality perform better at estimating quantiles away from this target. Our method can give more robust performance across a range of quantiles by targeting more than one quantile. One advantage of our method is that it is simple to apply to a range of models. For example extending the method to a skew-logistic model was much easier than for approaches based on D-optimal designs, where we need to re-calculate what the D-optimal design was for each new model.

This work was motivated by SEMG experiments. The results in Figure 3 show a substantive improvement in accuracy over the existing procedure for choosing stimulus levels in SEMG experiments. We obtain the same accuracy for estimating the median of the response curve with about 60 to 70 observations as the existing approach obtains using 200 observations. Thus using our sequential procedure for choosing stimulus levels could reduce the length of SEMG experiments to about one third their current length. This would have significant benefits for both the cost of such experiments, and the level of discomfort of the patients involved.

References

- Azadi, N. A. (2011). “Modeling the excitability behaviour of a single motor unit and experimental design for the surface EMG test.” Ph.D. thesis, Mathematics and Statistics Department, Lancaster University, Lancaster, UK. [291](#)

- Benassi, R., Bect, J., and Vazquez, E. (2011). “Robust Gaussian process-based global optimization using a fully Bayesian expected improvement criterion.” *Learning and Intelligent Optimization*, 6683: 176–190. [296](#)
- Bostock, H., Cikurel, K., and Burke, D. (1998). “Threshold tracking techniques in the study of human peripheral nerve.” *Muscle Nerve*, 21(2): 137–158. [288](#)
- Burke, D., Kiernan, M. C., and Bostock, H. (2001). “Excitability of human axons.” *Clinical Neurophysiology*, 112(9): 1575–1585. [288](#)
- Chaloner, K. and Verdinelli, I. (1995). “Bayesian experimental design: A review.” *Statistical Science*, 10(3): 273–304. [288](#)
- Chaudhuri, P. and Mykland, P. A. (1993). “Nonlinear experiments: Optimal design and inferences based on likelihood.” *Journal of the American Statistical Association*, 88: 938–546. [290](#)
- Dixon, W. J. and Mood, A. M. (1948). “A method for obtaining and analyzing sensitive data.” *Journal of the American Statistical Association*, 43: 109–126. [289](#)
- Doucet, A., de Freitas, J. F. G., and Gordon, N. J. (eds.) (2001). *Sequential Monte Carlo Methods in Practice*. New York: Springer-Verlag. [288](#), [293](#)
- Doucet, A., Godsill, S. J., and Andrieu, C. (2000). “On sequential Monte Carlo sampling methods for Bayesian filtering.” *Statistics and Computing*, 10: 197–208. [293](#)
- Dror, H. A. and Steinberg, D. M. (2008). “Sequential experimental designs for generalized linear models.” *Journal of the American Statistical Association*, 103(481): 288–198. [290](#), [293](#)
- Fearnhead, P. (2005). “Using random quasi-Monte-Carlo within particle filters, with application to financial time series.” *Journal of Computational and Graphical Statistics*, 14: 751–769. [293](#)
- Gramacy, R. B. and Polson, N. G. (2009). “Particle learning of Gaussian process models for sequential design and optimization.” *Journal of Computational and Graphical Statistics*, 20(1): 102–118. [288](#)
- Joseph, V. R. (2004). “Efficient Robbins-Monro Procedure for Binary Data.” *Biometrika*, 91: 461–470. [289](#)
- Joseph, V. R., Tian, Y., and Wu, C. F. J. (2007). “Adaptive Designs for Stochastic Root-Finding.” *Statistica Sinica*, 17: 1549–1565. [290](#)
- Khuri, A., Mukherjee, B., Sinha, B., and Ghosh, M. (2006). “Design issues for generalized linear models: A review.” *Statistical Science*, 21: 373–399. [287](#)
- Khuri, A. and Mukhopadhyay, S. (2006). *GLM designs: The dependence on unknown parameters dilemma.*, 203–223. Singapore: World Scientific. [287](#)

- Liu, J. and West, M. (2001). “Combined parameter and state estimation in simulation based filtering.” In Doucet, A., de Freitas, J. F. G., and Gordon, N. J. (eds.), *Sequential Monte Carlo in Practice*, 197–223. New York: Springer-Verlag. 294, 295
- Liu, J. S. and Chen, R. (1998). “Sequential Monte Carlo methods for dynamic systems.” *Journal of the American Statistical Association.*, 93: 1032–1044. 293
- Liu, J. S., Chen, R., and Wong, W. H. (1998). “Rejection control and sequential importance sampling.” *Journal of the American Statistical Association*, 93(443): 1022–1031. 294
- Mathew, T. and Sinha, B. K. (2001). “Optimal designs for binary data under logistic regression.” *Journal of Statistical Planning and Inference*, 93: 295–307. 297
- Mockus, J. (1989). *Bayesian approach to global optimization: Theory and applications.* New York: Springer. 296
- Neyer, B. T. (1994). “D-Optimality-Based Sensitivity Test.” *Technometrics*, 36: 61–70. 290
- Petris, G., Petrone, S., and Campagnoli, P. (2009). *Dynamic linear models with R.* New York: Springer. 293
- Ridall, P. G., Pettitt, A. N., McCombe, P. A., and Henderson, R. D. (2006). “Motor unit number estimation, a Bayesian approach.” *Biometrics*, 16: 1235–1250. 291
- Robbins, H. and Monro, S. (1951). “A stochastic approximation method.” *Annals of Mathematical Statistics*, 29: 400–407. 289
- Taddy, M. A., Gramacy, R. B., and Polson, N. G. (2011). “Dynamic Trees for Learning and Design.” *Journal of the American Statistical Association*, 106(493): 109–123.
URL <http://EconPapers.repec.org/RePEc:bes:jnlasa:v:106:i:493:y:2011:p:109-123> 288
- Voelkel, J. G. (1999). “Sequential Experimental Designs For Sensitivity Experiments.” Technical report, Center for Quality and Applied Statistics, Rochester Institute of Technology. <http://www.rit.edu/kgcoe/cqas/about/TR1999-1.pdf>. 296
- Wetherill, J. B. (1963). “Sequential estimation of quantal response curves.” *Journal of the Royal Statistical Society. Series B*, 25(1): 1–48. 289
- Wu, C. F. J. and Tian, Y. (2013). “Three-phase sequential design for sensitivity experiments.” Technical report, Georgia Institute of Technology. <http://www2.isye.gatech.edu/~jeffwu/~3pod/optimal.design.pdf>. 290
- Wu, J. (1985). “Efficient sequential designs with binary data.” *Journal of the American Statistical Association*, 80: 974–984. 289, 290, 298

