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

*Nammam A Azadi**

Department of Mathematics and Statistics, Lancaster University, UK
n.azadi@lancaster.ac.uk

Paul Fearnhead

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

Gareth Ridall

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

Joleen H Blok

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

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 mimimises the expected loss. 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 for choosing the stimulus values. If applied in practice, it could more than halve the length of SEMG experiments.

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

*This work was supported financially by Kurdistan University of Medical Sciences of Iran

1 Introduction

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 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 parameters ('local optimal' design), and is referred to as the design dependence problem (Khuri et al., 2006).

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 on line 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 on-line decision making are much rarer. 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 a 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 has a mean threshold for such stimulation as well as a range of values over which it displays stochastic behavior: it 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 a 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 of interest of a single MU. Then through simulation studies we compare the performance of our method with several existing non-Bayesian methods taken from the statistical design literature.

In Section 2 we describe existing methods for sequential design and discuss their limitations and advantages. In Section 3 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 4. Finally in Section 5 we finish with an evaluation and a discussion.

2 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 an greater height otherwise. The sequential procedure was continued until convergence.

In order to estimate the pth quantile Robbins and Monro (1951) introduced a non-parametric sequential procedure for binary observations (known as RM procedure) which received much attention. A simulation study by Wetherill (1963) showed that while the RM procedure performed well when estimating the median, it performs poorly for extreme quantiles due to a large bias. In order to improve upon the efficiency of the RM procedure, Wu (1985) implemented this procedure within a parametric framework that we refer to as logit-MLE. Wu's method chooses the next design point as the estimate of x_p from fitting a logistic regression model to the data. A known issue with this approach is the uncertainty about choosing the optimal initial design points. In

practice the author suggests a set of 10 to 14 points symmetrically distributed around a guess of the mean, but this is 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) 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.

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 their 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. In this paper we focus on the squared loss function (about a quantile of interest) but this approach can be easily modified to accommodate other loss functions.

3 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 MCMC is carried out. Finally the section ends with a description of the algorithm which defines our optimal data collection mechanism.

3.1 SEMG experiment

As described above, SEMG is a noninvasive 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 recoding was of just a single MU, as evidenced by the all-or-nothing (binary) character of the response. The stimulus-response behaviour of a 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 intensity 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-none state of the MU, in the right hand panel the same data are presented after conversion to binary responses by choosing an appropriate threshold.

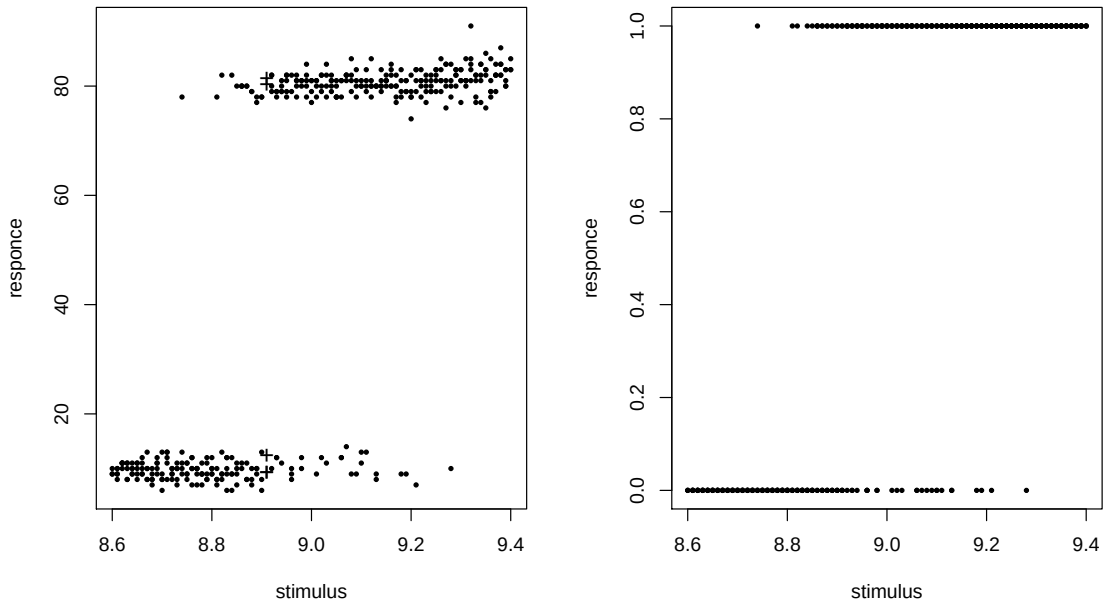

Figure 1: Stimulus-response plots. The left panel is a scatter plot showing the amplitude of the measured signal against stimulus intensity, recorded using surface electromyography. The low-level amplitudes around 10 μV represent background noise (no response), and the values around $80\mu V$ denote 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.

3.2 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 $\mathbf{s} = s_{1:n} = (s_1, \ldots, s_n)$ denote the vector of fixed and known stimulus levels and $\mathbf{y} = y_{1:n} = (y_1, y_2, \ldots, y_n)$ 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:

$$\Pr(Y_i = 1|s_i, m, b) = \frac{1}{1 + \exp\{-b(s_i - m)\}} \qquad i = 1, 2, \ldots, n \tag{1}$$

Here m denotes the mean threshold (the stimulus for which a 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).

3.3 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

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

In our simulations we use a prior for $p(\theta)$ that is mildly informative. Prior information is used from the results of the analysis of earlier SEMG experiments, see Section 4.1. 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:

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

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}.

3.4 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 on-line techniques that provide a flexible framework for the updating of posterior distributions in real time. The idea of SMC is to make an approximation of the posterior using a set of weighted particles, $\{\theta_t^{(i)}, w_t^{(i)}\}$, $i = 1, 2, \ldots, N$. 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 $t+1$

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

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 $i = 1, \ldots, N$, 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).

Algorithm 1 Bayesian updating using the particle approximation

Input: A set of stimuli values $(s_1, s_2, \ldots, s_n)$ where $s_i \in$ s.
A set of observations $(y_1, y_2, \ldots, y_n)$ 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 , $\theta_0^{(1)}, \ldots, \theta_0^{(N)}$, from prior $p(\theta)$.
Assign each particle the initial weight $w_0^{(i)} = 1/N$ for $i = 1, \ldots, N$.

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

1. Given the s_t generate the new observation y_t from a Bernoulli distribution

$$y_t \sim \text{Bernoulli}(1, p)$$

where p is given by

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

2. For the new observation y_t, update the weights using

$$w_t^{(i)} \propto w_{t-1}^{(i)} \times p$$

3. Normalize the new weights $W_t^{(i)} = \dfrac{w_t^{(i)}}{\sum_{i=1}^N w_t^{(i)}}$.

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

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 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. A weight is considered as negligible if it drops below a pre-specified discrimination factor ε_0. The threshold ε_0 is arbitrarily chosen so that it discards the weights that are too small. Secondly, we monitor the effective sample size (ESS) (Liu and Chen, 2008) 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, say 50% of the initial number of particles, 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:

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

were $\mathrm{M}_t^{(i)} = a\theta_t^{(i)} + (1-a)\bar{\theta}_t$, $\Sigma_t = hV_t$, and $\bar{\theta}_t$ and $V(t)$ are the current esimates 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 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 0.974 to 0.995 interval for the resampling to be efficient. The details of the SMS algorithm with resampling step are given in Algorithm 2.

Algorithm 2 :Sequential Monte Carlo algorithm

Input: A set of stimuli values $(s_1, s_2, \ldots, s_n)$ where $s_i \in \mathtt{s}$.
A set of observation $(y_1, y_2, \ldots, y_n)$ where y_i denotes the response of MU at stimulus s_i.
A guess value for m_0 and b_0, the true logit parameter values.
A prior, $p(\theta)$, for $\theta = (m, b)$ and The number of particles N.
Two arbritary thresholds, N_0 and ϵ_0 and a constant $0 < a < 1$.

Initialise: Generate N particles from prior $p(\theta)$, $\theta_0^{(1)}, \ldots, \theta_0^{(N)}$.
Set the initial weights $w_0^{(i)} = 1/N$ for $i = 1, \ldots, N$.

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

1. Given the s_t generate the new observation y_t from a Bernoulli distribution

$$y_t \sim \text{Bernoulli}(1, p)$$

   where p is given by

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

2. For the new observation y_t, update the weights using

$$w_t^{(i)} \propto w_{t-1}^{(i)} p(y_t|s_t, \theta_{t-1}^{(i)})$$

3. Normalize the new weights $W_t^{(i)} = \dfrac{w_t^{(i)}}{\sum_{i=1}^{N} w_t^{(i)}}$.
4. Discard those weights that are smaller than ϵ_0.
5. Compute $N_{ESS} = \left(\sum_{i=1}^{N} (W_t^{(i)})^2\right)^{-1}$
6. If $N_{ESS} < N_0$ do Liu and West resampling
   - Calculate the posterior mean and variance from the current set of weighted particles; denote these $\bar{\theta}$ and Σ_{t-1}.
   - Draw N new particles, $\{\theta_t^{(i)}\}_{i=1}^{N}$, from $\mathcal{N}(M_{\substack{(i)\\t-1}}, h^2\Sigma_{t-1})$ where $M^{(i)} = a\theta_{t-1}^{(i)} + (1-a)\bar{\theta}$.
   - Assign to each new particles the new and equal weight $1/N$.
   - Replace $\theta_{t-1}^{(i)} = \theta_t^{(i)}$.

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

3.5 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 $s_{1:t}$. Let $\phi = \phi(\theta)$ denote a function of the parameters that is of interest. To choose the best new stimulus value we minimise the expected value of a loss function. For the sake of simplicity we focus on a quadratic loss $L(\hat{\phi}, \phi_0) = (\phi_0 - \hat{\phi})^2$, which represents our perceived penalty in estimating ϕ_0 by $\hat{\phi}$. However our method is not restricted to the squared loss function and can be easily generalised to other types of loss. It can be easily shown that the estimate that minimises the quadratic loss function is the posterior mean, with a corresponding expected loss given by the posterior variance.

To select the optimal stimulus intensities $s_1, s_2, \ldots, s_T$, we want to construct a rule so that if applied recursively, it 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) or the expected improvement criterion (Jones et al., 1998).

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. That is, given $s_{1:t}$ and $y_{1:t}$, we want to choose

$$s_{t+1} = \arg\min_s \{\mathrm{E}_{Y_{t+1}}(\mathrm{Var}(\phi | y_{1:t}, Y_{t+1}, s_{1:t}, s))\}. \tag{6}$$

As observations are binary, for any value of s we can easily evaluate the right-hand side of (6) 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

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

and the terms on the right-hand side can be estimated using our weighted particles. The details are given in Algorithm 3 which summarises our strategy in selecting optimal design points for the SEMG experiment.

4 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.

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. This method is described in Algorithm 3.

Algorithm 3 Algorithm for Choice of Stimulus

Input:

- A set of possible stimulus levels, $(s_1, s_2, \ldots, s_K)$ where $s_i \in \mathbf{S}$.
- A set of observation $(y_1, y_2, \ldots, y_n)$ where y_i denotes the response of MU at stimulus s_i.
- A guess value for m_0 and b_0, the true logit parameter values.
- Prior distribution, $p(\theta)$, for $\theta = (m, b)$.
- A sample size, n, a constant $0 < a < 1$, the number of particles N, thresholds N_0 and ϵ_0.
- A quantity of interest $\phi(\theta)$, i.e. median or any quantile.

Initialise:

- Generate N particles $\{\theta_0^{(i)}\}_{i=1}^N$ from prior density $p(\theta)$.
- Set the initial weight of each particle $w_0^{(i)} = 1/N$ for $i = 1, \ldots, N$.

Iterate:

1. Set $t = 1$
2. For stimulus intensity $s_{t,k}$, $k = 1, \ldots, K$:
   (a) Find from a logistic model
   $$p(y_{t,k} = y | s_{t,k}, \theta_{t-1}^{(i)}) = \frac{1}{1 + \exp\{-b_0(s_{t,k} - m_0)\}} \qquad y = 0, 1$$
   (b) Update the weights with respect to above probabilities $w_{t,k,y}^{(i)} = w_{t-1}^{(i)} p(y_{t,k} = y | \theta_{t-1}^{(i)}, s_{t,k})$
   (c) Take $p_{t,k,1} = \sum_{i=1}^N w_{t,k,1}^{(i)}$ and $p_{t,k,0} = 1 - p_{t,k,1}$ as the probability of response and non-response respectively.
   (d) Normalise the weights $W_{t,k,y}^{(i)} = w_{t,k,y}^{(i)} / \sum_{i=1}^N w_{t,k,y}^{(i)}$.
   (e) Compute the posterior mean of $\phi(\theta)$:
   $$\bar{\phi}_{t,k,y} = \sum_{i=1}^N W_{t,k,y}^{(i)} \phi_{t,k,y}(\theta_{t-1}^{(i)})$$
   and the posterior variances
   $$\mathbf{\Sigma}_{t,k,y} = \sum_{i=1}^N W_{t,k,y}^{(i)} (\phi_{t,k,y}(\theta_{t-1}^{(i)}) - \bar{\phi}_{t,k,y})^2$$
   (f) Estimate the expected variance of $\phi(\theta)$ by
   $$\bar{\mathbf{\Sigma}}_{t,k} = p_{t,k,0} \mathbf{\Sigma}_{t,k,0} + p_{t,k,1} \mathbf{\Sigma}_{t,k,1}$$
3. The optimal stimulus $s_{opt} \in s_{1:K}$, is the stimulus intensity that minimises $\bar{\Sigma}_{t,k}$.
4. Simulate the observation y_{opt} for the optimal stimulus s_{opt}.
5. For the y_{opt}, update and normalize the weights using one iteration of Algorithm 2.

Output *A set of optimal stimuli (design points).*

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).

L-MLE The Wu method (Wu, 1985): In this method 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.

For each sequential design method, we use sequential Monte Carlo to perform online inference for the parameters. To compare the different methods we analyse 1000 simulated data sets with each method. Each simulated data set uses a different pair of parameter values, drawn from our prior. The efficiency of the different methods is judged based on the mean squared error of the estimates of the target quantile.

4.1 Formulation of the prior

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 that is found; and this information can give a relatively informative prior for m. In our simulations we took this prior to be uniform across $(8.6, 9.4)$, which is an appropriate level of uncertainty in m after this initial search phase.

4.2 Results of our simulation

Our simulations compare the efficiencies of our method in finding the 50^{th}, the 75^{th}, the 85^{th}, and the 95^{th} percentiles. For each method and each percentile we initialise the Algorithm 3 with $N = 4000$ particles obtained by taking 100 samples from b's prior and 40 samples from m's prior distribution. These particles are used to approximate posterior density. They increase to $N = 10000$ particles when the number of effective sample size falls below 2000 particles ($N_0 = 2000$). We also use $\epsilon_0 = 0.00003$ and the constant value $a = 0.98$. At each stage of the simulation, we calculate the mean squared error of the estimated quantile.

We compare each method based on the mean square error of its estimates, and look at how this depends on the number of observations made. To make this comparison clearer, we look at the relative accuracy of each method against the accuracy of our new approach. That is for a given method, the relative accuracy of the estimator based on n observations is defined to be the mean square error of the estimates using the new method and 200 observations divided by the mean square error of the estimates for that method after n observations. For example, a value of 0.5

would mean that the MSE for the method after n observations is twice that of our method using 200 observations.

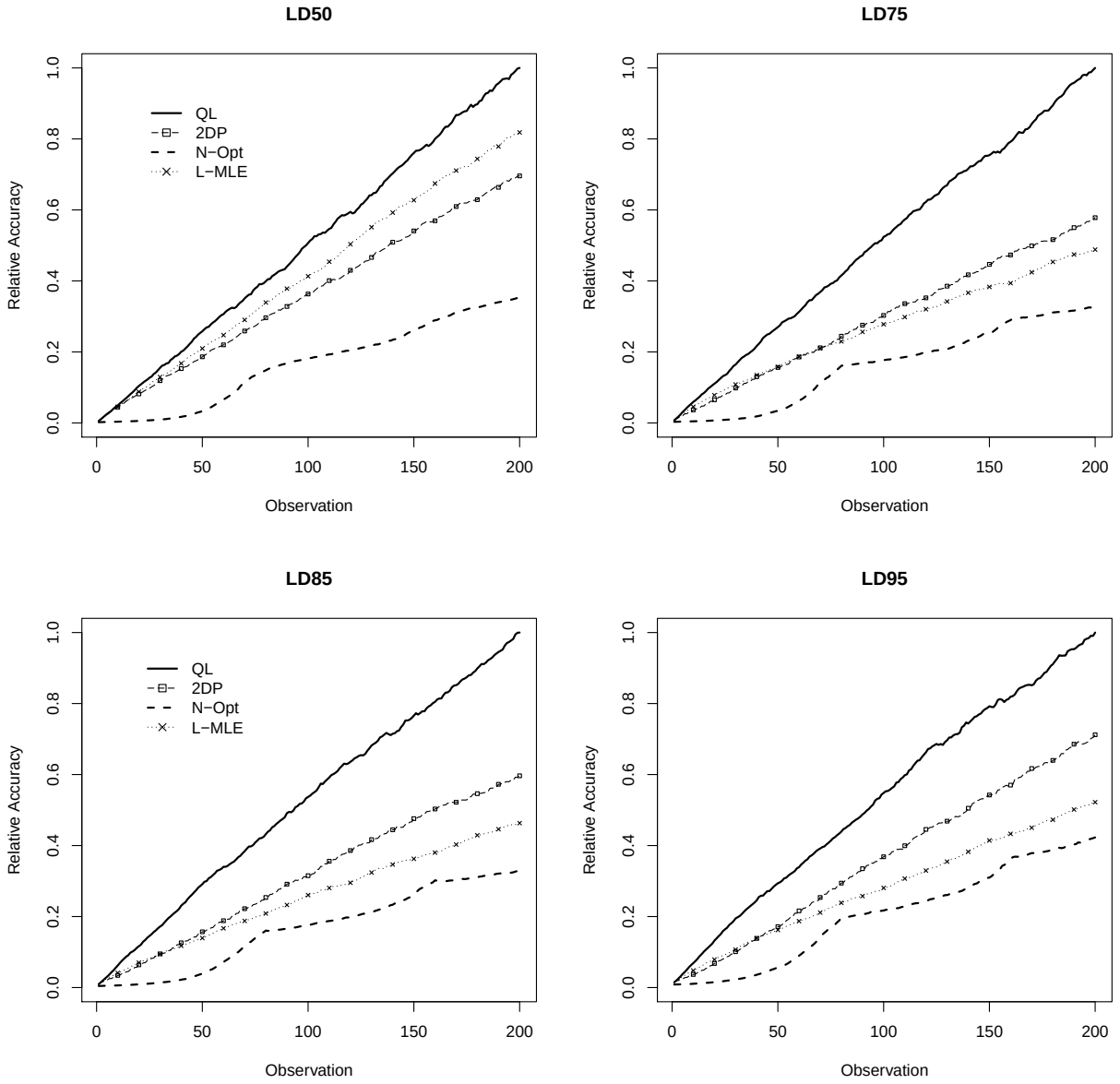

Figure 2: The relative gain in the mean squared error of our method over other approaches in the study for estimating the 50^{th}, 75^{th}, 85^{th} and 95^{th} quantiles. The $y-$scale represent the relative accuracy and $x-$scale the observation arriving over time.

The black solid lines in Figure 2 represent the method of the authors. The plots illustrate the larger relative accuracy using our method than the others at all quantiles. The accuracy is more substantial when estimating higher quantiles. The non-optimal procedure, as expected, is the least efficient and the lines showing its relative accuracy are consistently lower than those of the other schemes.

At LD50, the efficiency of Wu's method is close to that of our method but better than that of the

2-point design scheme. However the logit-MLE loses its efficiency dramatically when the target quantile is not the median. We can illustrate this by noting that for the estimation of, say, LD75, our method achieves a 50% in relative acuracy after collecting roughly 100 observations whereas with the 2-point design more than 190 observations are required. For the logit-MLE more than 200 observations require to gain a 50% in the relative accuracy.

5 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. Our simulation results suggest that our approach is more efficient than any of the alternative methods. Whilst developed for binary data, it is straightforward to extend this approach to other types of GLM.

This work was motivated by SEMG experiments. The results in Figure 2 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.

Our algorithm is close to that of Dror and Steinberg (2008) idea but is more general in applicability. The authors method is limited to the D-optimality criterion. Although it makes use of the particle approximation, it fails to address the problem of particle depletion (Doucet et al., 2000) which can reduce its accuracy. In contrast to the Dror and Steinberg (2008) method, we allow our particles to be rejuvenated in the presence of particle depletion. Furthermore our algorithm is straightforward and easy to implement. In addition, it does not required initial points and works well at small sample sizes.

References

Azadi, N. A. (2011). *Modeling the excitability behaviour of a single motor unit and conducting an efficient experimental design for the surface EMG test.* Ph. D. thesis, Mathematics and Statistics Department, Lancaster University, Lancaster, UK.

Bostock, H., K. Cikurel, and D. Burke (1998). Threshold tracking techniques in the study of human peripheral nerve. *Muscle Nerve 21*(2), 137–158.

Burke, D., M. C. Kiernan, and H. Bostock (2001). Excitability of human axons. *Clinical Neurophysiology 112*(9), 1575–1585.

Chaloner, K. and I. Verdinelli (1995). Bayesian experimental design: A review. *Statistical Science 10*(3), 273–304.

Chaudhuri, P. and P. A. Mykland (1993). Nonlinear experiments: Optimal design and inferences based on likelihood. *Journal of the American Statistical Association 88*, 938–546.

Dixon, W. J. and A. M. Mood (1948). A method for obtaining and analyzing sensitive data. *Journal of American Statistical Association 43*, 109–126.

Doucet, A., N. D. Freitas, and N. J. Gordon (2001). *Sequential Monte Carlo methods in practice.* New York: Springer.

Doucet, S., J. Godsill, and C. Andrieu (2000). On sequential monte carlo sampling methods for Bayesian filtering. Technical report.

Dror, H. A. and D. M. Steinberg (2008). Sequential experimental designs for generalized linear models. *Journal of the American Statistical Association 103*(481), 288–198.

Fearnhead, P. (2005). Using random quasi-monte-carlo within particle filters, with application to financial time series. *Journal of Computational Graphical Statistics 14*.

Jones, D. R., M. Schonlau, and J. William (1998). Efficient global optimization of expensive black-box functions. *Journal of Global Optimization 13*(4), 455492.

Khuri, A., B. Mukherjee, B. Sinha, and M. Ghosh (2006). Design issues for generalized linear models: A review. *Statistical Science 21*, 373–399.

Khuri, A. and S. Mukhopadhyay (2006). *GLM designs: The dependence on unknown parameters dilemma.*, pp. 203–223. Singapore: World Scientific.

Liu, J. S. and R. Chen (2008). Sequential Monte Carlo methods for dynamic systems. *Journal of the American Statistical Association 93*(443), 10321044.

Liu, J. S. and M. West (2001). *Combined parameter and state estimation in simulation-based filtering.* New York.

Mathew, T. and B. K. Sinha (2001). Optimal designs for binary data under logistic regression. *Journal of Statistical Planning and Inference 93*, 295–307.

Mockus, J. (1989). *Bayesian approach to global optimization: Theory and applications.* New York: Springer.

Petris, G., S. Petrone, and P. Campagnoli (2009). *Dynamic linear models with R.* New York: Springer.

Ridall, P. G., A. N. Pettitt, P. A. McCombe, and R. D. Henderson (2006). Motor unit number estimation, a bayesian approach. *Biometrics 16*, 1235–1250.

Robbins, H. and S. Monro (1951). A stochastic approximation method. *Annals of Mathematical Statistics 29*, 400407.

Wetherill, J. B. (1963). Sequential estimation of quantal response curves. *Journal of the Royal Statistical Society. Series B 25*(1), 1–48.

Wu, J. (1985). Efficient sequential designs with binary data. *Journal of American Statistical Association 80*, 974–984.