Bayesian Optimal Experimental Design for Generalized Linear Models.

Manuel Villarreal, Carlos Velázquez and Arturo Bouzas

Mexico City, Mexico

Abstract

Keywords: Optimal experimental design, Generalized linear models, Psycophysics, Change point detection

1. Introduction

One of the main challenges in scientific research is the design of an experiment. A good experimental design will make the difference between finding an answer to our research question and wasting valuable resources. Optimal Experimental Design (OED) allows us to re-interpret the design of an experiment as a decision problem where the objective is to maximize a utility function. This function is a numeric representation of our preferences over the possible results of the experiment.

To optimize a design, first we need to identify which of our variables are subject to this procedure. For example, we might want to select which values of an independent variable should be tested, or how many times we should test each of those values. Most of the time, these choices are made on the basis of previous research. However, it might be the case that there is not enough information to make these decisions with confidence, or that the values that are commonly used, do not allow for strong conclusions. Second, we need to formalize the objective of the experiment through a utility function. For example, if our objective is to discriminate between the predictions of two cognitive models, we can select a function wich assigns a greater value to designs were their predictions are further appart. There are two examples in the Psychological literature that implement different utility functions (e.g. Myung and Pitt, 2009; Zhang and Lee, 2010). In both, the objective of the design is to discriminate between competing cognitive models, however, considerations of what makes a model "good" drive the authors to choose different functions.

In this paper, we will present a different approach where the problem is not to select between models but to make inferences about the parameters of a single one. In particular, we will present an example of OED with a model that is commonly used in Psychophysics and Decision Making, the logistic response model. In this example, we will optimize an experimental design focused on studying the ability of participants to detect changes in the success rate of a probabilistic series. Additionally, we will show how different prior assumptions can change the results obtained from this process, this will highlight the importance of the choice of prior distributions in the process of designing an experiment. It is important to note that the aproach presented here follows a Bayesian perspective. A good introduction to Bayesian and other approaches to OED for Generalized Linear Models (GLM), and their limitations, can be found in Khuri et al. (2006).

The rest of the paper will be organized as follows. First, we will introduce the concepts of OED particularly in the field of GLM's. Second, we will present a brief summary of the experimental problem being addressed and, a computational model which has been used to account for the behavior of human participants under similar settings. Finally, we will present the results of the optimization procedure under different prior distributions, including one generated from the model responses to simulations of the experiment.

2. Optimal experimental design

The concepts of OED from a decision theoretic perspective were introduced by Lindley (1972). In this work, the author argued that the design of an experiment should be approached as a decision problem. In general, one needs to specify a utility function reflecting the purpose of the experiment and then select the design which maximizes expected utility. It is also worth noting that in practice, this process will rely on the assumption of an underlying data generating model. For example, in their paper, Myung and Pitt (2009) use the tools of OED to optimize an experiment to discriminate between two memory retention models. In this case, the design found will be optimal only in the case that either model is true.

Under the OED framework, the problem of designing an experiment reduces to selecting a design η from a set H, where data $\mathbf{y}$ will be observed given the design and a model's parameters θ. The utility function $U(\eta, \mathbf{y}, \theta)$, maps the values of these variables to the Real numbers. For this example, the objective will be to select the n values of an independent variable x from an experimental space X. This space can be thought of as the numerical variables which can be controlled by the experimenter. The design η, will be the weight of our observations assigned to each of the n design points $\eta_i = n_i/n$, with the constraint that $\sum_{i=1}^{n} \eta_i = 1$. In this case, η can be interpreted as a probability measure over the experimenral space and H as the set of all such measures.

The expected utility of a design η is represented by the following equation:

$$U(\eta) = \int \int U(\eta, \mathbf{y}, \theta) p(\theta|\mathbf{y}, \eta) p(\mathbf{y}|\eta) d\theta d\mathbf{y} \tag{1}$$

where $p(\theta|\mathbf{y}, \eta)$ represents the posterior probability of θ given the design η and data $\mathbf{y}$ and, $p(\mathbf{y}|\eta)$ is the predictive distribution of $\mathbf{y}$. Finding an optimal design, reduces to the problem of finding the values for the variables in η for which equation 1 takes it's maximum value. For many applications the requiered integration is not a trivial problem, therefore, one could rely on analytic or numeric approximations to solve it.

Many utility functions have been proposed for both linear and non-linear design problems, however, as was previously mentioned, the choice of a utility function depends on the research question. For example, Myung and Pitt (2009) used the sum of squered errors between the data generated by a model and the predictions of a competing one given an expeirmental design. This is because, the objective is to find a design that can discriminate the predictions of the two. On the other hand, Zhang and Lee (2010) choose a utility function based on the Bayes Factor. Both functions will have their advantages, nontheless, the first, emphasizes disciminability of model's predictions, while the second aims for a design in which the data generating model is more likely.

When the objective of the experiment is to make an inference about the parameters of a generalized linear model, some authors (e.g. Bernardo, 1979) have proposed to consider the gain in Shanon's Information as a utility function. In this case, the objective is to find the design η that maximizes the expected gain in Shanon Information, or equvalently, maximizes the Kullback-Leibler divergence between the posterior and prior distribution. With this function, the expected utility of an experimental design is defined as:

$$U(\eta) = \int \int log \frac{p(\theta|y,\eta)}{p(\theta)} p(y,\theta|\eta) d\theta dy \tag{2}$$

The prior distribution in the denominator of the logarithm can be droped as it does not depend on the experimental design, therefore, the optimal design will be the one that maximizes:

$$U(\eta) = \int \int log\{p(\theta|y,\eta)\} p(y,\theta|\eta) d\theta dy \tag{3}$$

which is the posterior expected Shanon information. The objective of this function is to minimize the posterior variance of the model's parameters, however, there are other functions that can be used for design optimization for GLM's. A good overview of some of this functions for linear and non-linear design problems can be found in Chaloner and Verdinelli (1995).

One of the problems when dealing with design optimization for GLM's is that the posterior distribution of the parameter vector θ is not always tractable. However, it is common to use the normal approximation to the posterior distribution

$$\theta|y,\eta \sim N\left(\hat{\theta}, [nI(\hat{\theta},\eta)]^{-1}\right) \tag{4}$$

Where $I(\hat{\theta},\eta)$ denotes the observed fisher information matrix given an experimental design and $\hat{\theta}$ is the maximum likelihood estimate of θ. Even with this, the marginal distribution of the data $p(\mathbf{y}|\eta)$ in equation 1 needs to be aproximated. In this case, according to Chaloner and Larntz (1989), we can take the prior distribution of θ as the predictive distribution of $\hat{\theta}$ in order to approximate the posterior expected utility.

With both aproximations, the value of $U(\eta)$ becomes:

$$U(\eta) = -\frac{k}{2}log(2\pi) - \frac{k}{2} + \frac{1}{2}\int log\{det[nI(\theta,\eta)]\}p(\theta)d\theta \tag{5}$$

Equation 5 gives the exact expected utility of an experimental design. However, one could drop the constant and multiplier terms giving the following form:

$$\phi(\eta) = \int log\{det[nI(\theta,\eta)]\}p(\theta)d\theta \tag{6}$$

$$= \mathbb{E}_\theta\left(log\{det[nI(\theta,\eta)]\}\right) \tag{7}$$

the function $\phi(\eta)$ is known as a design criterion. An optimal design would be the one maximizing equation 7. It is worth noting that this relies on the normal approximation to the posterior distribution in 4, therefore, for small samples there are some constraints that can help to assure normality (see Clyde and Chaloner, 2002). This kind of criteria can be found for other utility functions. For example, Chaloner and Larntz (1989) presented the derivation and application of two design criteria, one of which was $\phi(\eta)$. The second one, aimed to optimize a design where certain predictions of the model are of interest.

Finally, to test wether a particular design η_0 is optimal, one can use the concept of a directional derivative. The definition and conditions for this test can be found in more detail in Chaloner and Larntz (1989). As the authours suggest, we denote η_x as a design which puts all the weight on a single point x in the experimental space X and $d(\eta, x)$ as the derivative of η

in the direction of η_x. A design η_0 is optimal if the roots of the directional derivative are in the support of η_0 and it is lower for any other point in X. For the selected criterion ϕ, the directional derivative is equal to:

$$d(\eta, x) = \mathbb{E}_\theta \; tr\left(I(\theta, \eta_x) I(\theta, \eta)^{-1}\right) - p \tag{8}$$

where p represents the number of parameters in the logistic model and $tr(\cdot)$ is the represents the trace.

3. OUTLINE

4. Optimal Experimental Design: Example

4.1. Using a model to generate prior distributions

5. Results

5.1. Consruction of the prior distribution

5.2. Optimal design

6. Discussion

References

Bernardo, J., 1979. Expected information as expected utility. The Annals of Statistics 7 (3), 686–690.

Chaloner, K., Larntz, K., February 1989. Optimal bayesian design applied to logistic regression experiments. Journal of Statistical Planning and Inference 21 (2), 191–208.

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

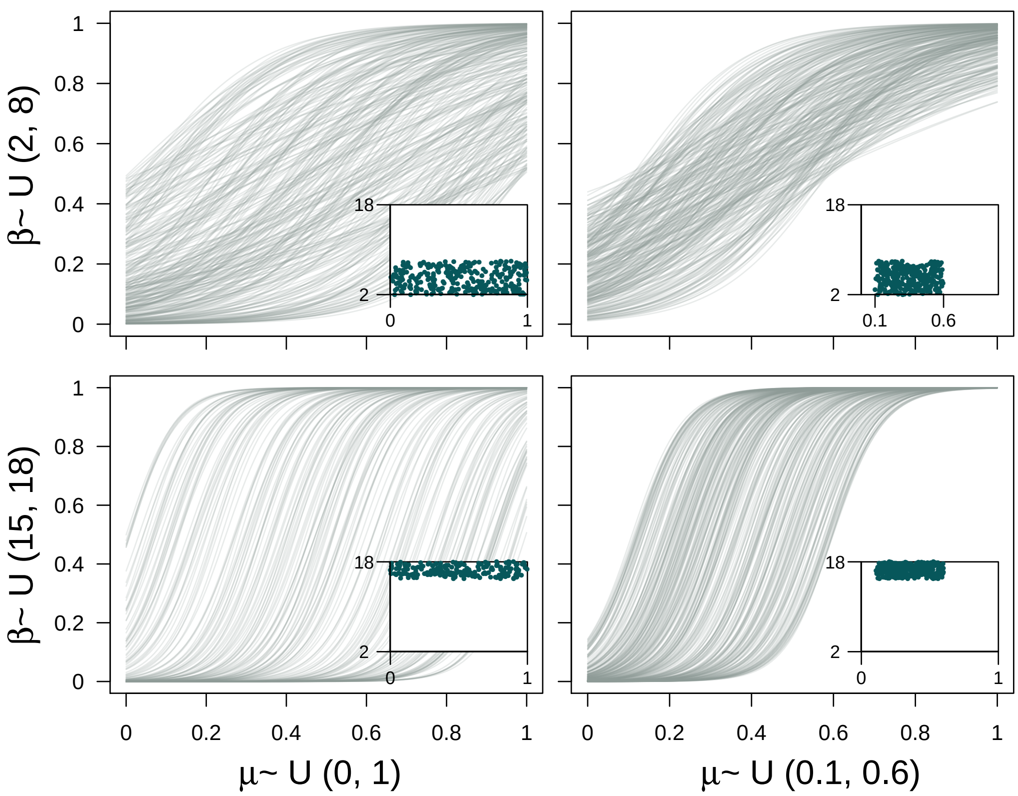

Figure 1: Prior and predictive prior distributions.

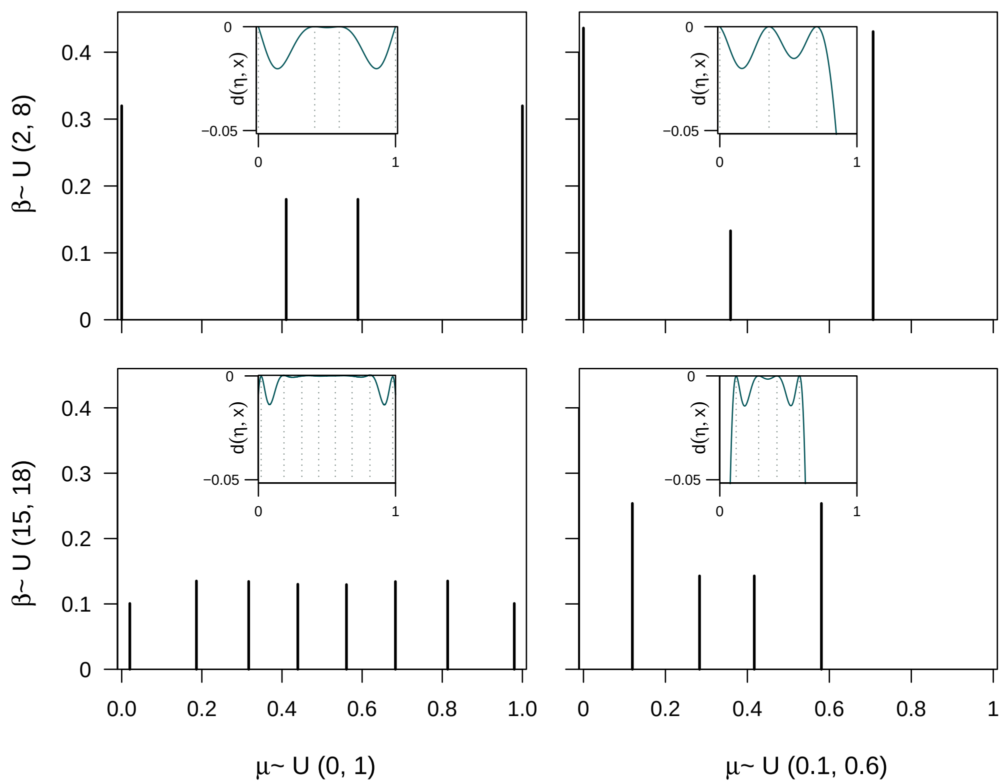

Figure 2: Optimal experimental designs and directional derivatives.

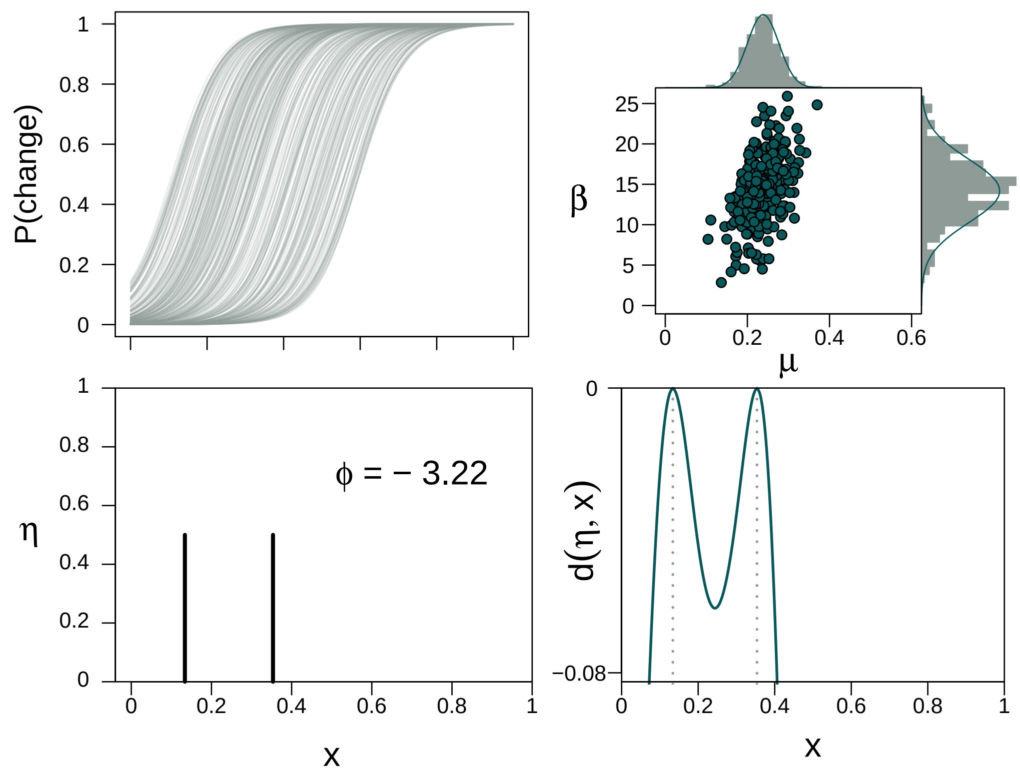

Figure 3: Informative prior distribution, prior predictive, optimal design and directional derivative.

Clyde, M., Chaloner, K., 2002. Constrained design strategies for improving normal approximations in nonlinear regression problems. Journal of Statistical Planning and Inference 104, 175–196.

Khuri, A. I., Mukherjee, B., Sinha, B. K., Ghosh, M., 2006. Design issues for generalized linear models: A review. Statistical Science 21 (3), 376–399.

Lindley, D. V., 1972. Bayesian statistics, a review. Vol. 2. SIAM.

Myung, J., Pitt, M., 2009. Optimal experimental design for model discrimination. Psychological Review 116 (3), 499–518.

Zhang, S., Lee, M., 2010. Optimal experimental design for a class of bandit problems. Journal of Mathematical Psychology 54, 499–508.