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 like time and money.

When designing an experiment, often one starts by making decisions about the number of participants, how many and what values of our independent variable we should test or how many times should we test those values. Most of the time, these choices are made on the basis of previous research on the field. 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. Optimal Experimental Design (OED) offers an alternative to solve this kind of problems through the formalization of the design problem.

OED allows us to re-interpret the problem of designing an experiment as a decision problem, in which the purpose is to maximize a utility function. This function is a numeric representation of our preference over the possible consequences of running an experiment. Therefore, the optimization of an experimental design requires us to have a formal interpretation of the purpose of the experiment.

The concept of optimizing an experimental design is not new to psychological research. There are already examples of design optimization in the literature (e.g. Myung and Pitt, 2009; Zhang and Lee, 2010). Both of this examples discuss and demonstrate the advantages of OED for model comparison in psychology.

In this paper, we will present a different approach where the problem is not to select between two models but to make inferences about the parameters of a single one.

In particular, we will present an example of OED in the context of Generalized Linear Models. These models are widely used in psychophysics. The problem that will be treated here is fairly new, however, the methodological aspects remain the same even with more straight-forward psychophysical experiments. In order to do this, we will use a particular parametrization of the logistic model which is primarily used in statistics.

2. Optimal experimental design

To apply the concepts of OED to a particular problem, first, we need to define the elements of the design space. This space is conformed by the variables that we can manipulate during the experiment, for example, the values that our independent variable might take or the weight (proportion of observations) assigned to each of those values. These elements are the ones that we can modify in order to optimize the design.

The second step would be to formalise the objective of the experiment. For example, in the case of a logistic model, we might want to find the values of our independent variable that minimize the variance of the model parameters, or we might be interested in the magnitude of the physical stimulus for which the probability of a response takes on a certain value. The formalization of the research question will define a utility function.

The last step is to specify the prior information that we have about the problem at hand. This last step can be carried out in two ways, first, we can try to optimize the experiment for a particular guess about the parameter values of the model of interest, or we could use a probability distribution to account for the uncertainty in the values that we are interested in. This last step is primarily important to the optimization process for generalized linear models, because the optimal design will depend on the values of the parameters.

Once we have defined the design space η, the objective of the experiment and the prior informartion $p(\theta)$, the expected utility of a design is represented by the following equation:

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

Therefore, finding the experimental design that is optimal given a utility function, reduces to the problem of finding the values for the variables in η for which equation 1 takes it's maximum value.

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) utilizes 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 of this functions will have their advantages, nontheless, the first utility emphasizes disciminability of model's predictions, while the second one 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 represented by the following equation:

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

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.

When dealing with generalized linear models, the posterior distribution of the parameter vector θ is not always tractable, however, in the literature of design optimization it is common to use the following aproximation 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 ($p(y|\eta)$) in equation 1 also needs to be aproximated, however, when the posterior utility only depends on y through some constistent estimate of $\hat{\theta}$, a further approximation is to take the predictive distribution of $\hat{\theta}$ to be the prior distribution Chaloner and Larntz (1989).

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

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

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

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

The function $\phi(\eta)$ is known as a design critirion. An optimal design would be the one maximizing equation 6, nontheless, it is worth noting that this relies on the normal approximation to the posterior distribution in 4, therefore, for small smaples there are some constraints that help to assure normality (see Clyde and Chaloner (2002)).

3. OUTLINE

4. Optimal Experimental Design: Example

Why is is detecting changes important for an organism?

Change detection in probabilistic series.

Arising problems with experimental design.

Research question and its statistichal interpretation

Assumtion about the relationship between a subjects response the dependent variable under study

Design space for this problem and how to reduce the dimensionality of the space by assuming experimental constraints.

Utility function and its relationship with the objective of the experiment

Arising problems with utility function and the proposed response function. Bayesian solution, assigning a prior distribution to the parameters, the less research in a field the more difficult it is to assign an informative prior, however, we could use other cognitive models in order to propose a prior distribution.

4.1. Using a model to generate prior distributions

Using the prior distribution, the utility function and the definition of a design space we can otimize the experimental design in this case we are looking for $\delta\theta^*$ that maximizes the following equation:

$$U(\delta\theta^*) = \max_{\delta\theta} \int_{\beta} log(det(I(\beta|\delta\theta)))\pi(\beta)d\beta \tag{7}$$

The previous integral can be approximated via Monte Carlo sampling

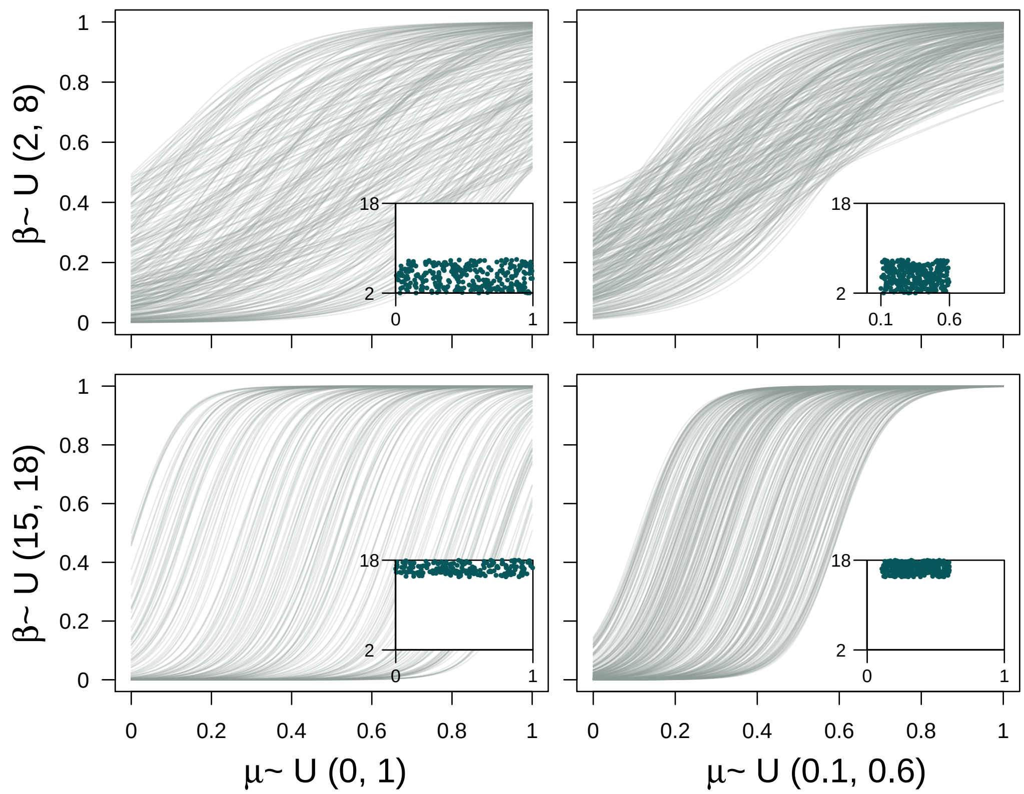

Figure 1: Prior and predictive prior distributions.

5. Results

5.1. Consruction of the prior distribution

Prior over model parameters(Gallistel et al 2014) Results Constructing the prior: we take a multivariate normal distribution with mean and covariance equal to the unbiased estimators for both parameters.

5.2. Optimal design

Aproximating the utility function (integral) throught Monte Carlo simulation Utility aproximation for 2 Design points

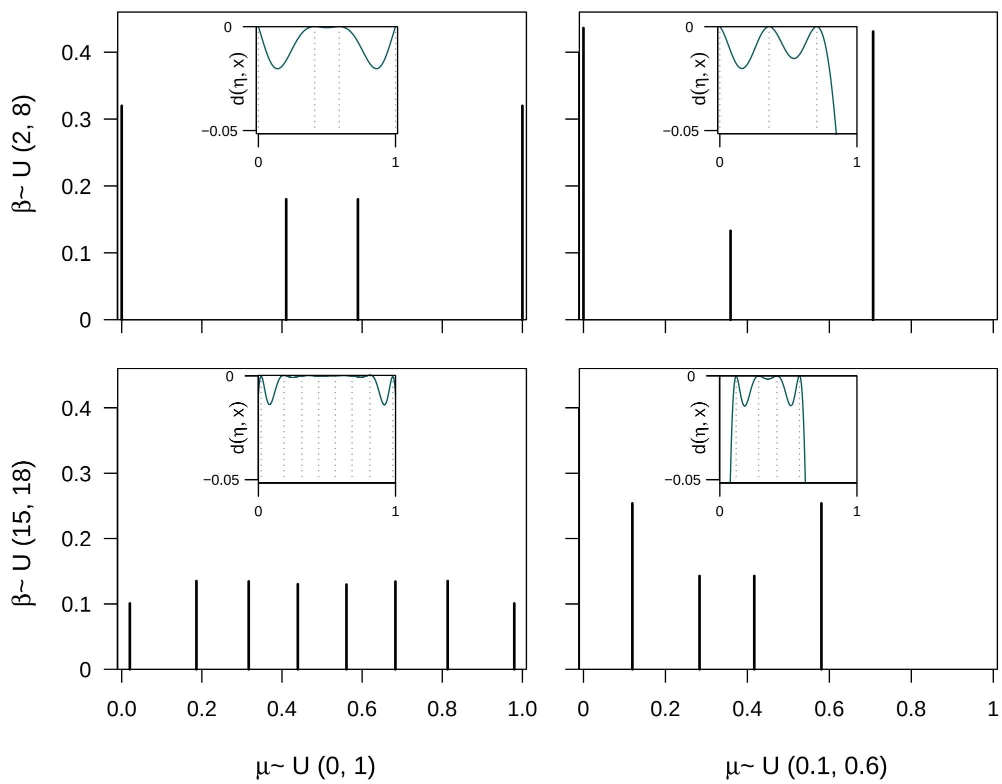

Figure 2: Optimal experimental designs and directional derivatives.

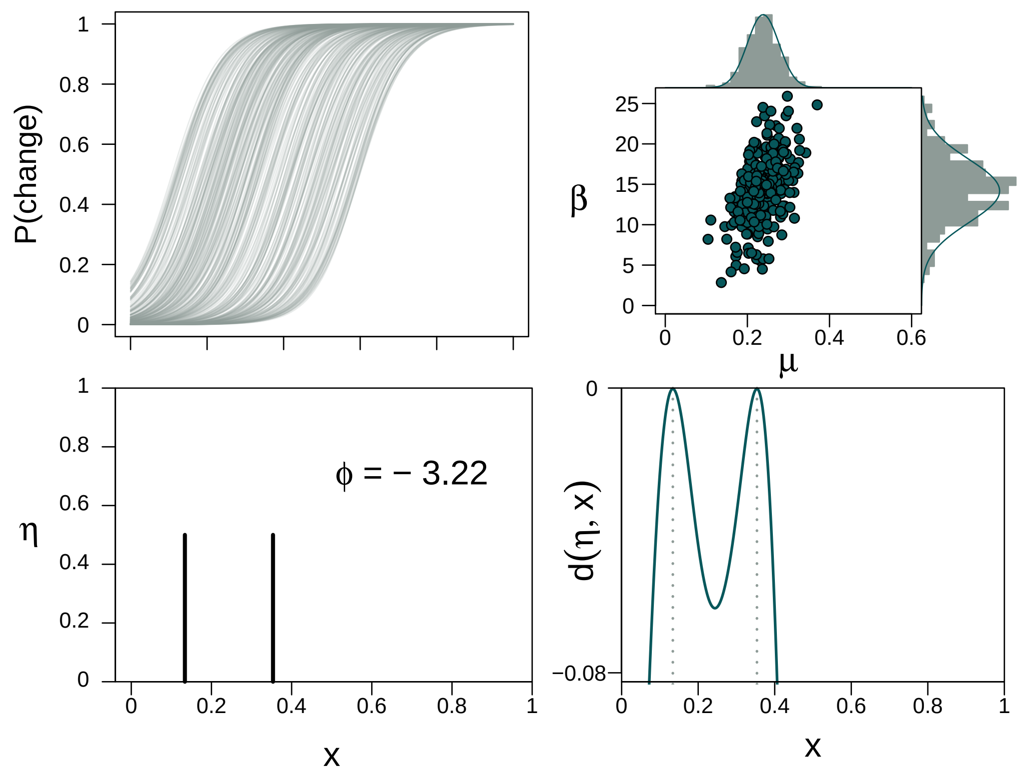

Figure 3: Informative prior distribution, prior predictive, optimal design and directional derivative.

the approximation returns a smooth curve over the 2 point design space.

6. Discussion

Optimal design for the example Properties of the most useful points (they land on the points of the curve where the steepness changes most dramatically)

Advantages of Optimal Design

Using models to generte prior distributions.

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.

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.

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.