

# 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, Psychophysics, 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 representation of our preferences over the possible results of the experiment.

To optimize the design of an experiment, 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

16 conclusions. Second, we need to formalize the objective of the experiment  
17 through a utility function. There are two examples in the literature that use  
18 different functions (Myung and Pitt, 2009; Zhang and Lee, 2010, e.g.), in  
19 both examples the objective is to discriminate between competing cognitive  
20 models. Nonetheless, considerations of what makes a model “good” drive  
21 the authors to choose different functions.

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

35 The rest of the paper will be organized in the following way. First, we  
36 will introduce the concepts of OED particularly in the field of Generalized  
37 linear models. Second, we will present a brief summary of the experimental  
38 problem being addressed and, a computational model which has been used  
39 to account for the behavior of human participants under similar settings.  
40 Finally, we will present the results of the optimization procedure under dif-

41 ferent prior distributions, including one generated from the model responses  
 42 to simulations of the experiment.

## 43 **2. Optimal experimental design**

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

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

64 The expected utility of a design  $\eta$  is represented by the following equation:

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

65 Finding the design that is optimal given a utility function, reduces to the  
66 problem of finding the values for the variables in  $\eta$  for which equation 1  
67 takes it's maximum value. However, for many applications the requiered  
68 integration is not a trivial problem, therefore, one could rely on analytic or  
69 numeric approximations to solve it.

70 Many utility functions have been proposed for both linear and non-linear  
71 design problems, however, as was previously mentioned, the choice of a utility  
72 function depends on the research question. For example, Myung and Pitt  
73 (2009) utilizes the sum of squered errors between the data generated by a  
74 model and the predictions of a competing one given an expeirmental design.  
75 This is because, the objective is to find a design that can discriminate the  
76 predictions of the two. On the other hand, Zhang and Lee (2010) choose a  
77 utility function based on the Bayes Factor. Both of this functions will have  
78 their advantages, nontheless, the first utility emphasizes discriminability of  
79 model's predictions, while the second one aims for a design in which the data  
80 generating model is more likely.

81 When the objective of the experiment is to make an inference about the  
82 parameters of a generalized linear model, some authors (e.g. Bernardo, 1979)  
83 have proposed to consider the gain in Shanon's Information as a utility func-  
84 tion. In this case, the objective is to find the design  $\eta$  that maximizes the ex-  
85 pected gain in Shanon Information, or equivalently, maximizes the Kullback-  
86 Leibler divergence between the posterior and prior distribution. With this  
87 function, the expected utility of an experimental design is represented by the

88 following equation:

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

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

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

92 Which is the posterior expected Shanon information.

93 When dealing with generalized linear models, the posterior distribution  
 94 of the parameter vector  $\theta$  is not always tractable, however, in the literature  
 95 of design optimization it is common to use the following aproximation to the  
 96 posterior distribution

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

97 Where  $I(\hat{\theta}, \eta)$  denotes the observed fisher information matrix given an ex-  
 98 perimental design and  $\hat{\theta}$  is the maximum likelihood estimate of  $\theta$ . Even with  
 99 this, the marginal distribution of the data ( $p(y|\eta)$ ) in equation 1 also needs  
 100 to be aproximated, however, when the posterior utility only depends on  $y$   
 101 through some constistent estimate of  $\hat{\theta}$ , a further approximation is to take  
 102 the predictive distribution of  $\hat{\theta}$  to be the prior distribution Chaloner and  
 103 Larntz (1989).

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

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

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

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

The function  $\phi(\eta)$  is known as a design criterion. An optimal design would be the one maximizing equation 6, nonetheless, 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 help to assure normality (see Clyde and Chaloner, 2002).

### 3. OUTLINE

#### 4. Optimal Experimental Design: Example

Why is detecting changes important for an organism?

Change detection in probabilistic series.

Arising problems with experimental design.

Research question and its statistical interpretation

Assumption 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,

127 however, we could use other cognitive models in order to propose a prior  
128 distribution.

#### 129 4.1. Using a model to generate prior distributions

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

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

133 The previous integral can be approximated via Monte Carlo sampling

## 134 5. Results

### 135 5.1. Construction of the prior distribution

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

### 139 5.2. Optimal design

140 Aproximating the utility function (integral) throught Monte Carlo simu-  
141 lation Utility aproximation for 2 Design points

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

## 143 6. Discussion

144 Optimal design for the example Properties of the most useful points (they  
145 land on the points of the curve where the steepness changes most dramati-  
146 cally)

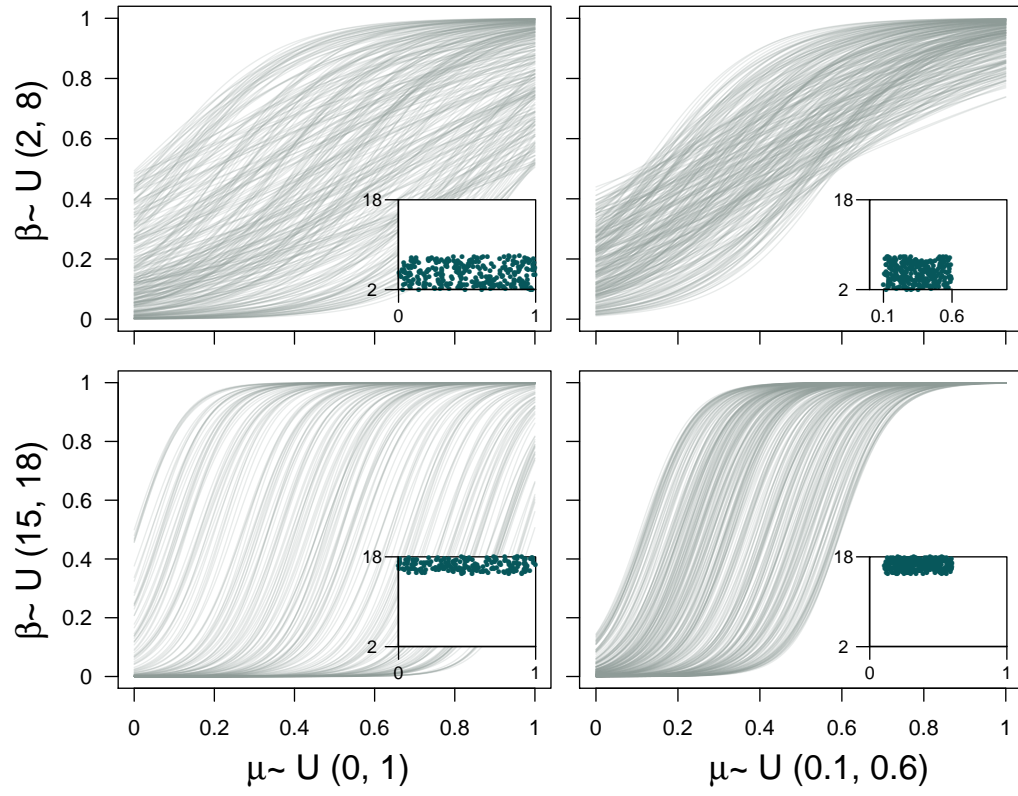


Figure 1: Prior and predictive prior distributions.



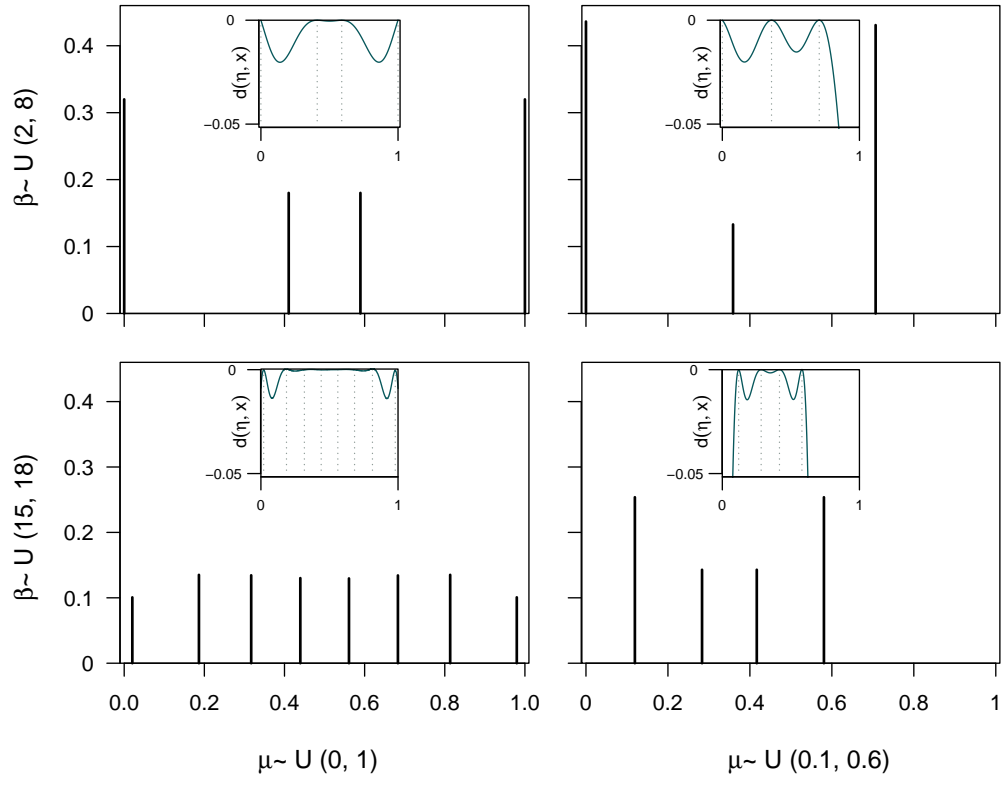


Figure 2: Optimal experimental designs and directional derivatives.

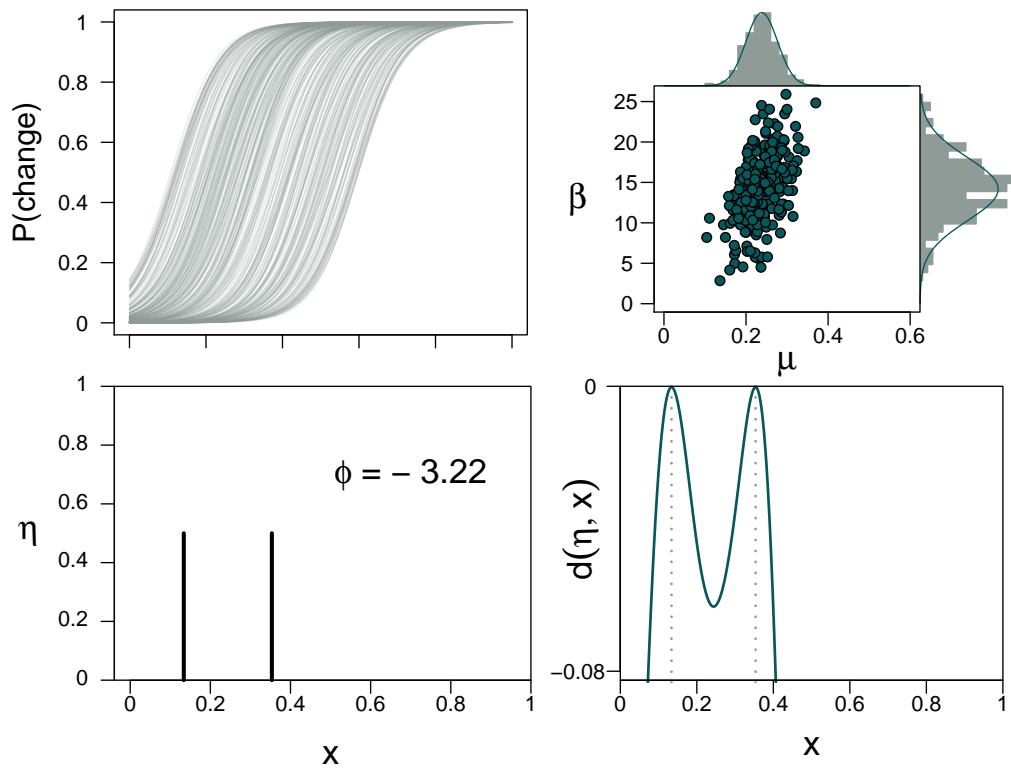


Figure 3: Informative prior distribution, prior predictive, optimal design and directional derivative.

147 Advantages of Optimal Design  
148 Using models to generate prior distributions.

## 149 **References**

- 150 Bernardo, J., 1979. Expected information as expected utility. *The Annals of*  
151 *Statistics* 7 (3), 686–690.
- 152 Chaloner, K., Larntz, K., February 1989. Optimal bayesian design applied  
153 to logistic regression experiments. *Journal of Statistical Planning and In-*  
154 *ference* 21 (2), 191–208.
- 155 Clyde, M., Chaloner, K., 2002. Constrained design strategies for improving  
156 normal approximations in nonlinear regression problems. *Journal of Sta-*  
157 *tistical Planning and Inference* 104, 175–196.
- 158 Khuri, A. I., Mukherjee, B., Sinha, B. K., Ghosh, M., 2006. Design issues for  
159 generalized linear models: A review. *Statistical Science* 21 (3), 376–399.
- 160 Lindley, D. V., 1972. *Bayesian statistics, a review*. Vol. 2. SIAM.
- 161 Myung, J., Pitt, M., 2009. Optimal experimental design for model discrimi-  
162 nation. *Psychological Review* 116 (3), 499–518.
- 163 Zhang, S., Lee, M., 2010. Optimal experimental design for a class of bandit  
164 problems. *Journal of Mathematical Psychology* 54, 499–508.