

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 conclusions.

16 Second, we need to formalize the objective of the experiment through a utility
17 function. There are already examples of OED in the Psychological literature
18 that implement different utility functions (e.g. Myung and Pitt, 2009; Zhang
19 and Lee, 2010). In both, the objective of the design is to discriminate between
20 competing cognitive models, however, considerations of what makes a model
21 “good” drive 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 data generating model. For example, in their paper, Myung and
 51 Pitt (2009) use the tools of OED to optimize an experiment to discriminate
 52 between two memory retention models. In this case, the design found will
 53 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 η from a set H , where data \mathbf{y} will be observed
 56 given the design and a model's parameters θ . 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 η , 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, η can be interpreted as a
 63 probability measure over the experimenral space.

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

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

83 When the objective of the experiment is to make an inference about the
84 parameters of a generalized linear model, some authors (e.g. Bernardo, 1979)
85 have proposed to consider the gain in Shanon's Information as a utility func-
86 tion. In this case, the objective is to find the design η that maximizes the ex-
87 pected gain in Shanon Information, or equivalently, maximizes the Kullback-

88 Leibler divergence between the posterior and prior distribution. With this
 89 function, the expected utility of an experimental design is represented by the
 90 following equation:

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

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

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

94 which is the posterior expected Shanon information. The objective of this
 95 function will be to minimize the posterior variance of the model's param-
 96 eters, however, there are other utility functions that can be used for design
 97 optimization for generalized linear models. A good overview of some of this
 98 functions for linear and nonlinear design problems can be found in Chaloner
 99 and Verdinelli (1995).

100 When dealing with generalized linear models, the posterior distribution
 101 of the parameter vector θ is not always tractable, however, in the literature
 102 of design optimization it is common to use the normal approximation to the
 103 posterior distribution

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

104 Where $I(\hat{\theta}, \eta)$ denotes the observed fisher information matrix given an ex-
 105 perimental design and $\hat{\theta}$ is the maximum likelihood estimate of θ . Even with
 106 this, the marginal distribution of the data ($p(y|\eta)$) in equation 1 needs to be
 107 aproximated, however, when the posterior utility only depends on y through

108 some consistent estimate of $\hat{\theta}$, a further approximation is to take the pre-
 109 dictive distribution of $\hat{\theta}$ to be the prior distribution Chaloner and Larntz
 110 (1989).

111 With both approximations, 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:

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

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

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

121 Finally, to check if a particular design is optimal, one needs to use the
 122 concept of the directional derivative (see Chaloner and Larntz, 1989). In this
 123 case, we denote x as a design which puts all the weight on a single point x
 124 in the experimental space X . The derivative of a design η in the direction of

125 the design x is defined as:

$$d(\eta, x) = \lim_{\epsilon \rightarrow 0} \epsilon^{-1} (\phi((1 - \epsilon)\eta + \epsilon x) - \phi(\eta)) \quad (8)$$

126 3. OUTLINE

127 4. Optimal Experimental Design: Example

128 Why is detecting changes important for an organism?

129 Change detection in probabilistic series.

130 Arising problems with experimental design.

131 Research question and its statistical interpretation

132 Assumption about the relationship between a subjects response the depen-
133 dent variable under study

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

136 Utility function and its relationship with the objective of the experiment

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

142 4.1. Using a model to generate prior distributions

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

$$U(\delta\theta^*) = \max_{\delta\theta} \int_{\beta} \log(\det(I(\beta|\delta\theta)))\pi(\beta)d\beta \quad (9)$$

146 The previous integral can be approximated via Monte Carlo sampling

147 **5. Results**

148 *5.1. Construction of the prior distribution*

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

152 *5.2. Optimal design*

153 Aproximating the utility function (integral) throught Monte Carlo simu-
154 lation Utility aproximation for 2 Design points
155 the approximation returns a smooth curve over the 2 point design space.

156 **6. Discussion**

157 Optimal design for the example Properties of the most useful points (they
158 land on the points of the curve where the steepness changes most dramati-
159 cally)

160 Advantages of Optimal Design

161 Using models to generte prior distributions.

162 **References**

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

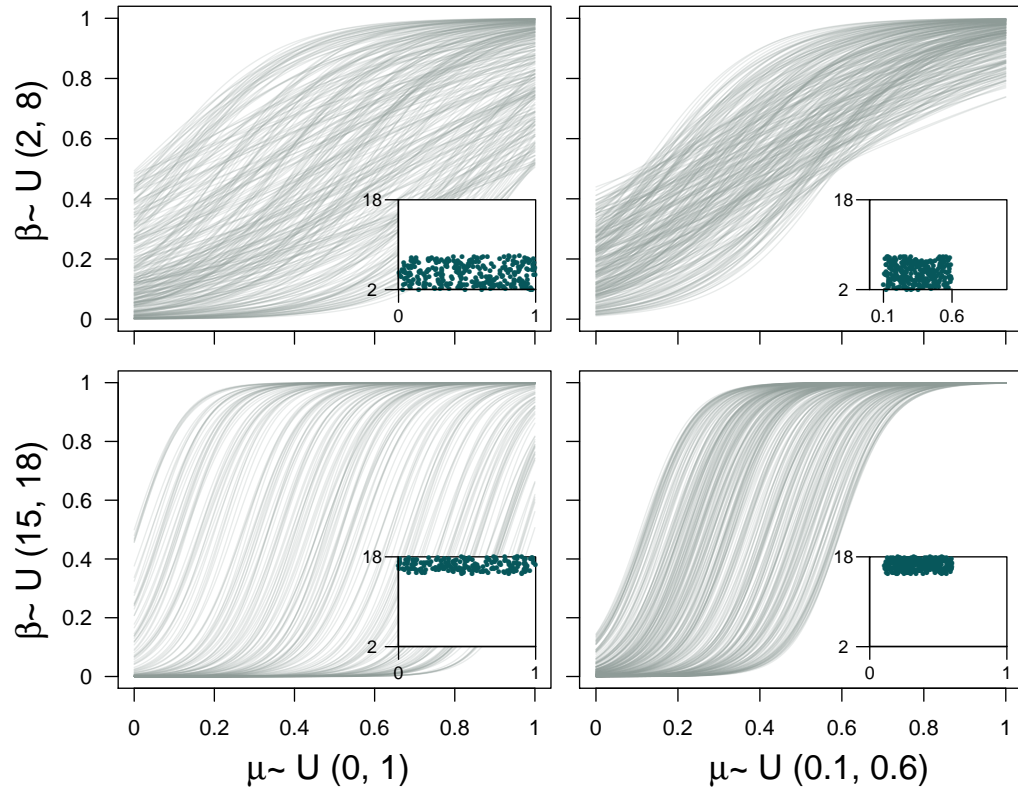


Figure 1: Prior and predictive prior distributions.

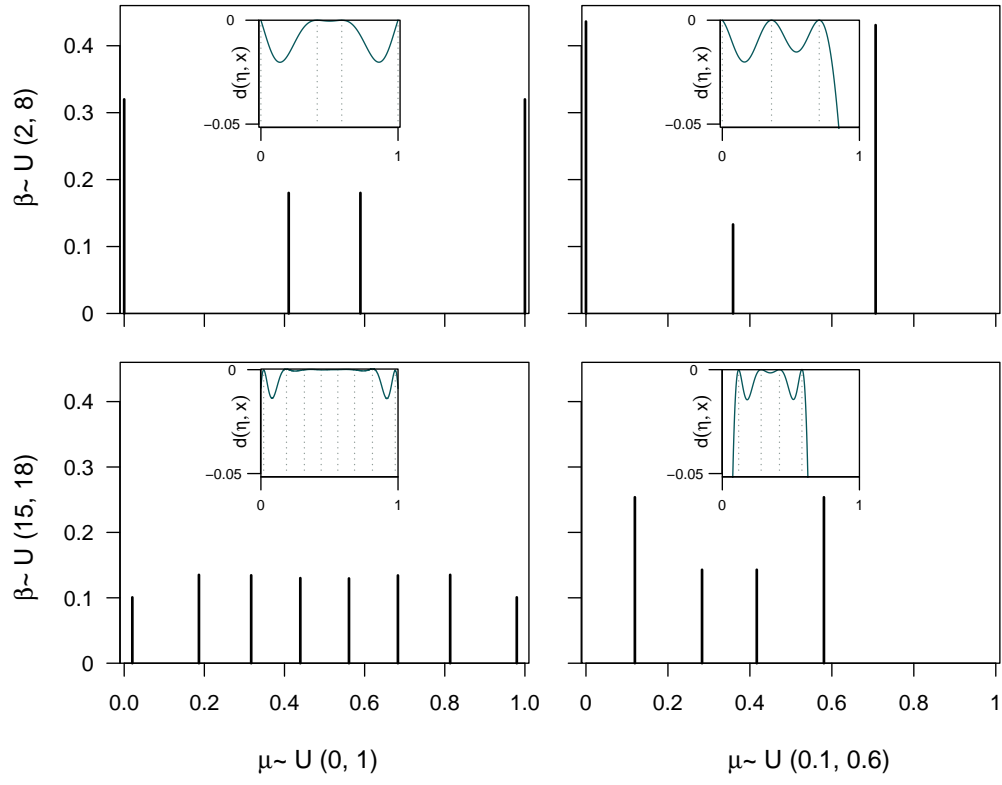


Figure 2: Optimal experimental designs and directional derivatives.

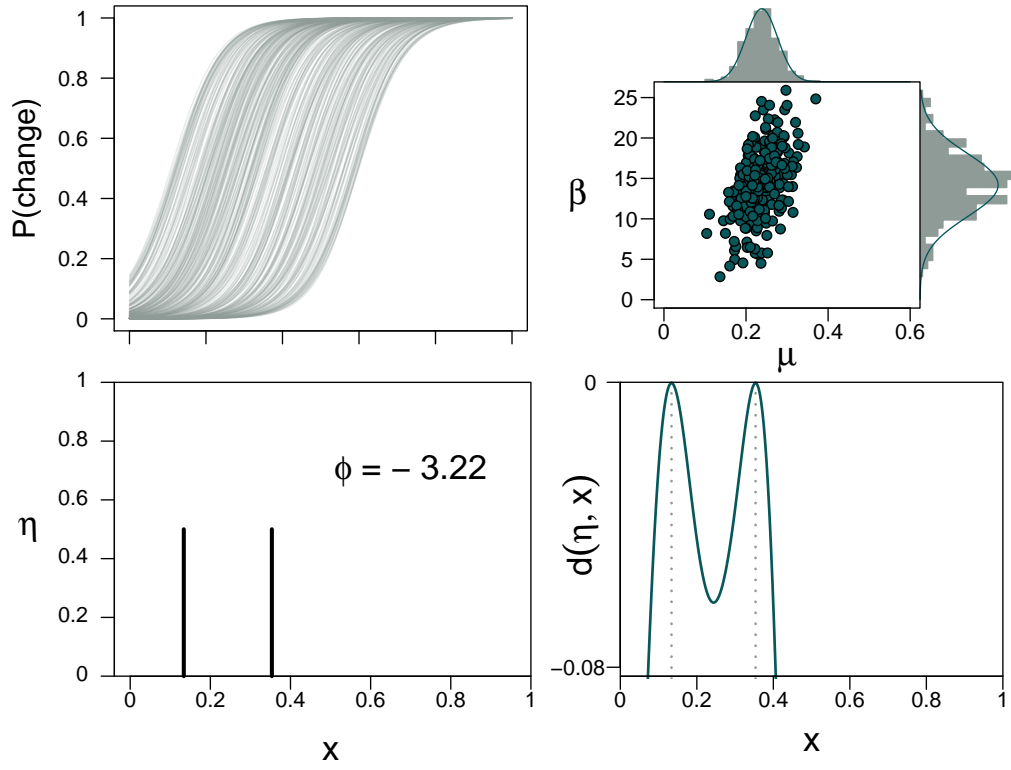


Figure 3: Informative prior distribution, prior predictive, optimal design and directional derivative.

- 165 Chaloner, K., Larntz, K., February 1989. Optimal bayesian design applied
166 to logistic regression experiments. *Journal of Statistical Planning and In-*
167 *ference* 21 (2), 191–208.
- 168 Chaloner, K., Verdinelli, I., 1995. Bayesian experimental design: A review.
169 *Statistical Science* 10 (3), 273–304.
- 170 Clyde, M., Chaloner, K., 2002. Constrained design strategies for improving
171 normal approximations in nonlinear regression problems. *Journal of Sta-*
172 *tistical Planning and Inference* 104, 175–196.
- 173 Khuri, A. I., Mukherjee, B., Sinha, B. K., Ghosh, M., 2006. Design issues for
174 generalized linear models: A review. *Statistical Science* 21 (3), 376–399.
- 175 Lindley, D. V., 1972. Bayesian statistics, a review. Vol. 2. SIAM.
- 176 Myung, J., Pitt, M., 2009. Optimal experimental design for model discrimi-
177 nation. *Psychological Review* 116 (3), 499–518.
- 178 Zhang, S., Lee, M., 2010. Optimal experimental design for a class of bandit
179 problems. *Journal of Mathematical Psychology* 54, 499–508.