Bayesian Adaptive Optimal Design of Psychology Experiments

Jay I. Myung, Mark I. Pitt, Yun Tang, and Daniel R. Cavagnaro

Department of Psychology,
Ohio State University,
1835 Neil Avenue,
Columbus, OH 43210-1351, USA
{myung.1, pitt.2, tang.162, cavagnaro.2}@osu.edu

Abstract. Experimentation is fundamental to the advancement of science, whether one is interested in studying the neuronal basis of a sensory process in cognitive neuroscience or assessing the efficacy of a new drug in clinical trials. Adaptive methodologies in experimentation, in which the information learned from each experiment is used to inform subsequent experiments, are particularly attractive because they can have the potential to reduce the time required for data collection while simultaneously increasing the informativeness of the knowledge learned in the experiment. More concretely, the problem to be solved in adaptive sequential design optimization for model discrimination is to identify an experimental design under which one can infer the underlying model, among a set of candidate models of interest, in the fewest possible steps. This problem is challenging because of the many, sometimes arbitrary, choices that must be made when designing an experiment. Nevertheless, it is generally possible to find a design that is optimal in a defined sense. In this paper, addressing the design optimization problem in discrimination of formal models in psychology, we apply a simulated-based Bayesian approach that was recently introduced in statistics (Müller et al, 2004) and present simulation results from its application.

Keywords. Bayesian adaptive design, model discrimination, particle filter, psychology experiment, stochastic optimization.

1 Introduction

Optimal experimental design has been considered at length in statistics (e.g., Chernoff, 1959; Kiefer, 1959; Box & Hill, 1967; Chaloner & Verdinelli, 1995; Atkinson & Donev, 1992) as well as in other science and engineering disciplines (e.g., El-Gamal & Palfrey, 1996; Bardsley, Wood & Melikhova, 1996; Allen, Yu & Schmitz, 2003). Among a variety of questions about design optimization that can be addressed, the one that has received most of the attention is that of identifying a design that makes the variances of parameter estimates of a given model as small as possible, thereby allowing the model to make the most accurate predictions. This goal is achieved by what is known as the D-optimum criterion, under which the design that maximizes the determinant of the variance-covariance matrix is to be chosen, formally speaking. In the case of multiple models being entertained, what is the so called T-optimum criterion (Ponce de Leon & Atkinson, 1991; Ucinski & Bogacka, 2005) is sought.

Design optimization can also be formulated from a sequential (i.e., adaptive) decision perspective. To give a relevant example, in sequential design optimization, the optimal design is obtained after repeating the experiment several times. Specifically, using the response outcome from each experimental stage, one then seeks the design for the next stage that is maximally informative, appropriately defined. The adaptive design process is repeated over a series of stages until a stopping criterion is met. This is illustrated in Figure 1.

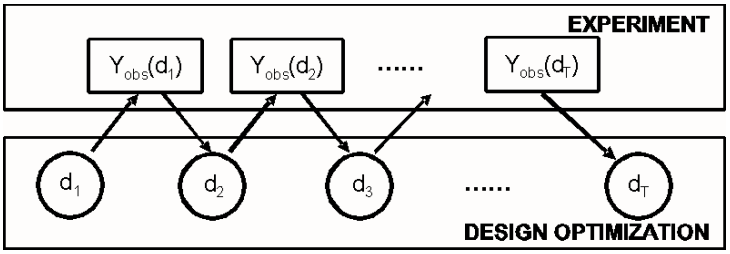

Fig. 1. Schematic illustration of adaptive design optimization.

Recently, a promising new approach has been proposed in statistics to solve the design optimization problem (Müller, Sanso & De Iorio, 2004; Amzal, Bois, Parent & Robert, 2006). It is a simulation-based

Bayesian approach based on Markov chain Monte Carlo (Gilks, Richardson & Spiegelhalter, 1996). What makes this approach an important break-through is that it is a fully general, albeit approximate, solution to optimizing any experimental design, including those that are most challenging, such as those involving nonlinear models with non-normal errors.

In this paper, extending recent work of Myung and Pitt (2008) and Cavagnaro, Myung, Pitt and Kujala (2009), we apply the simulation Bayesian approach and demonstrate its effectiveness in designing optimal designs for the purpose of discriminate quantitative models in psychology. We begin by discussing a formal framework of adaptive design optimization.

2 Adaptive Design Optimization for Model Discrimination

The problem of adaptive design optimization for model discrimination is identifying an experimental design under which one can infer the underlying model, among a set of models under consideration, as well as the model's parameter values, in the fewest possible steps. Bayesian decision theory offers a principled approach to this optimization problem. Each potential design is treated as a gamble whose payoff is determined by the outcome of an experiment carried out with that design. The idea is to estimate the utilities of hypothetical experiments carried out with a given design, so that an "expected utility" of that design can be computed. This is done by considering every possible observation that could be obtained from an experiment with a given design and then evaluating the relative likelihoods and statistical values of these observations. The design with the highest expected value is then chosen as the optimal design.

2.1 A Simulation-based Bayesian Approach

Formally, the problem of adaptive design optimization for model discrimination is to identify an optimal design d^* that maximizes the expected utility over all possible designs (e.g., Amzal et al, 2006)

$$d^* = \underset{d}{\operatorname{argmax}}\{U(d)\}. \tag{1}$$

In the above equation, $U(d)$ is referred to as the *global utility* function and is defined as

$$U(d) = \sum_{m=1}^{K} p(m) \int\int u(d, \theta_m, y)\, p(y|\theta_m, d)\, p(\theta_m)\, dy\, d\theta_m, \tag{2}$$

where $m = \{1, 2, \ldots, K\}$ is one of a set of K formal models being considered, d is a design, y is the outcome of an experiment with design d under model m, and θ_m is a parameterization of model m. We refer to the function $u(d, \theta_m, y)$ as the *local utility* of the design d. It measures the utility of a *hypothetical* experiment carried out with design d when the outcome y is observed from model m with its parameter value θ_m. Consequently, $U(d)$ represents the expected local utility over all models under consideration, the full parameter space of each model, and all possible observations given a particular model-parameter pair, with respect to the model prior probability $p(m)$, the parameter prior distribution $p(\theta_m)$, and the sampling distribution $p(y|\theta_m, d)$, respectively.

At each stage $s = 1, 2, \ldots$ of experimentation, an optimal design is sought by maximizing the global utility function $U(d)$ in Eq. (2) given the model and parameter priors. Once a specific outcome z_s is observed at stage s of an *actual* experiment carried out with the optimal design d_s, the outcome is then used to update the model and parameter priors for the next stage $s + 1$ using Bayes rule and Bayes factor calculation (e.g., Gelman et al, 2004) as follows

$$\begin{aligned} p_{s+1}(m) &= \frac{p_s(m)}{\sum_{k=1}^{K} p_s(k)\, BF_{(k,m)}(z_s)} \\ p_{s+1}(\theta_m) &= \frac{p(z_s|\theta_m, d_s)\, p_s(\theta_m)}{\int p(z_s|\theta_m, d_s)\, p_s(\theta_m)\, d\theta_m}, \end{aligned} \tag{3}$$

where $BF_{(k,m)}(z_s)$ denotes the Bayes factor defined as the ratio of the marginal likelihood of model k to that of model m given the realized outcome z_s. The above updating scheme is applied successively

on each stage of experimentation, after an initialization with equal model priors $p_{(s=0)}(m) = 1/K$ and a suitable parameter prior $p_{(s=0)}(\theta_m)$.

2.2 Computation by Density Simulation

To find the optimal design d^* in a general setting is a highly nontrivial problem. The computation requires simultaneous optimization and high-dimensional integration, which can be analytically intractable for the complex, nonlinear models as often used in many real-world problems. A new approach to this problem has been proposed in statistics (Müller et al, 2004). It is a simulation-based approach that includes a computational trick that allows one to find the optimal design without having to directly evaluate the integration in Eq. (2).

According to the computational trick, the design d is treated as a random variable and an auxiliary distribution $h(d, \cdot)$ is defined, which admits $U(d)$ as its marginal density. Specifically, suppressing the subscript s for simplicity, we define

$$h(d, y_1, \theta_1, \ldots, y_K, \theta_K) = \alpha \left[\sum_{m=1}^{K} p(m)\, u(d, \theta_m, y_m) \right] p(y_1, \theta_1, \ldots, y_K, \theta_K | d), \tag{4}$$

where $\alpha(> 0)$ is the normalizing constant of the auxiliary distribution and $p(y_1, \theta_1, \ldots, y_K, \theta_K|d) = \prod_{m=1}^{K} p(y_m|\theta_m, d)p(\theta_m)$. Note that the subscript m in the above equations refers to model m, not the stage of experimentation. For instance, y_m denotes an experimental outcome generated from model m with design d and parameter θ_m. Marginalizing $h(d, \cdot)$ over $(y_1, \theta_1, \ldots, y_K, \theta_{MK})$ yields

$$\begin{aligned} h(d) &= \int \ldots \int h(d, y_1, \theta_1, \ldots y_K, \theta_K) dy_1\, d\theta_1 \ldots dy_K\, d\theta_K \\ &= \alpha\, U(d). \end{aligned} \tag{5}$$

Consequently, the design with the highest utility can be found by taking the mode of a sufficiently large sample from the marginal distribution $h(d)$. To improve computational efficiency, we augment the auxiliary distribution with independent samples of y's and θ's given design d as follows, assuming $h(d, y_{1,j}, \theta_{1,j}, \ldots, y_{K,j}, \theta_{K,j})$ is non-negative and bounded,

$$h_J(d, \cdot) = \alpha_J \prod_{j=1}^{J} h(d, y_{1,j}, \theta_{1,j}, \ldots, y_{K,j}, \theta_{K,j}), \tag{6}$$

for a positive integer J and $\alpha_J(> 0)$. The marginal distribution of $h_J(d)$ obtained after integrating out model parameters and outcome variables will then be equal to $\alpha_J\, U(d)^J$. Hence, as J increases, the distribution $h_J(d)$ will become more highly peaked around its (global) mode corresponding to the optimal design d^*, thereby making it easier to identify the mode.

The distribution $h_J(d, \cdot)$ in Eq. (6) is simulated using a sequential Monte Carlo particle filter algorithm (Amzal et al, 2006; Doucet, de Freitas & Gordon, 2001), initially for say $J = 10$ and then by gradually increasing J on subsequent stages of experimentation on an appropriate simulated annealing schedule (e.g., increase J by one every five stages). From the sample of draws, we then empirically estimate the desired marginal distribution $\hat{U}(d)$, up to a constant proportionality, by collecting all d's but disregarding y's and θ's.

3 Experiments

This section presents an application of the Bayesian adaptive design optimization approach described in the previous section in simulated experiments. The goal of the simulation is to assess whether the Bayesian approach is sufficiently robust to solve optimization problems that arise in psychology. Specifically, we are interested in knowing how efficiently and effectively an optimal design can help identify

the data-generating model, among a set of competing models under consideration, and also its true parameter values.

A theoretical issue in psychology we chose to demonstrate the design optimization approach is that of children's numerical estimation. Developmental psychologists have been interested in the question of how children perceive and represent numerical magnitudes such as amount of money, distance, number of discrete objects, and location of integers on number lines, as well as the question of how their mental representations change as they get older (Opfer & Siegler, 2007). It is well established that 5- to 10-year olds' numerical estimation is highly inaccurate and distorted. For example, children often judge the difference between \$1 and \$100 as being larger than the difference between \$901 and \$1000 (Siegler & Booth, 2004), thus suggesting a logarithmic scale representation. On the other hand, they estimate the difference between \$1 and \$6 as being equal to the difference between \$11 and \$16, thereby correctly relying upon a linear representation as it should be. Opfer and Siegler (2007) summarize these findings by saying, "Children between 5 and 10 years of age are believed to rely on linear representations with small numbers but to only gradually extend the linear representations to larger numbers..." (p. 171).

With the goal of identifying the function underlying numerical estimation, Opfer and Siegler (2007) employed a number-line experimental task in which children were presented with an integer between 1 and 999 and asked, without feedback, to indicate how large the number is. They responded by marking a vertical hatch mark on a horizontal 25-cm number line. with its left end labelled 0 and its right end labelled 1000. In particular, Opfer and Siegler were interested in evaluating the following two models:

$$\begin{aligned} LIN: &\quad y = ax + b + e \\ LOG: &\quad y = a\log(x) + b + e, \end{aligned} \tag{7}$$

where x is the stimulus value between 1 and 999, y is the response, and e is a normal error with mean 0 and standard deviation c, and finally, a, b and c are parameters, i.e., $\theta = (a, b, c)$.

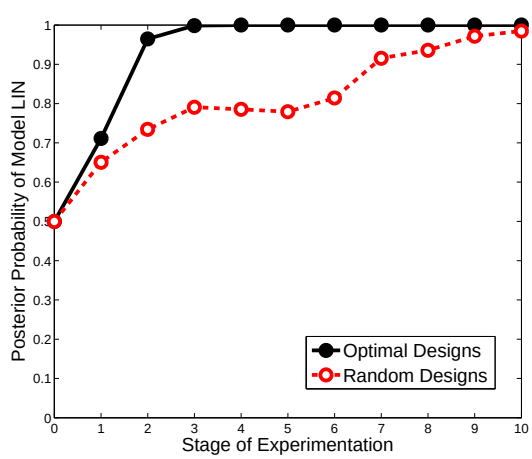

Fig. 2. Posterior model probability of model LIN as a function of the stage of experimentation. The solid curve represents the result of optimal designs and the dotted curve represents the result of random (non-optimized) designs. LIN is the data generating model.

We conducted a computer simulation to illustrate the adaptive design optimization procedure for discriminating between the two models in Eq. (7), in which optimal designs consisting of x values were sought over a series of stages of experimentation. On each stage, we used the particle filter algorithm (Amzal, 2006, p. 776) to find an optimal stimulus value $x = d^*$ $(0 < x < 1)$[1], that maximizes $U(d)$ in Eq. (2). We then generated a simulated response z from model LIN according to the following model

[1] In our simulations, for mathematical convenience, both the stimulus value x and response y were re-scaled to lie between 0 and 1, instead of 0 and 1000 in the original experiment of Ofper and Siegler (2007).

equation: $z = (0.9)d^* + 0.1 + e$, where $e \sim N(0, 0.1^2)$. This was followed by updating the model and parameter priors according to Eq. (3). This procedure was repeated for 10 stages of experimentation. For comparison, we also conducted simulated experiments in which no optimal designs were sought. Instead, on each stage, a design value d was randomly generated between 0 and 1, independent of the model and parameter priors as well as simulated responses. Regarding the local utility function $u(\cdot)$ in Eq. (2), we chose to use a Bayes factor based function of the form: $u(d, \theta_m, y_m) = BF_{(m,m')}(y_m)$, where the subscript m' refers to the other competing model (e.g., $m = LIN, m' = LOG$).[2] Uniform priors defined over the following parameter ranges were used: $0 < a < 5, -1 < b < 1, 0 < c < 1$ for model LIN and $0 < a < 1, 0 < b < 2, 0 < c < 1$ for model LOG. All results below are based on 20 particles, each with 50 iterations.

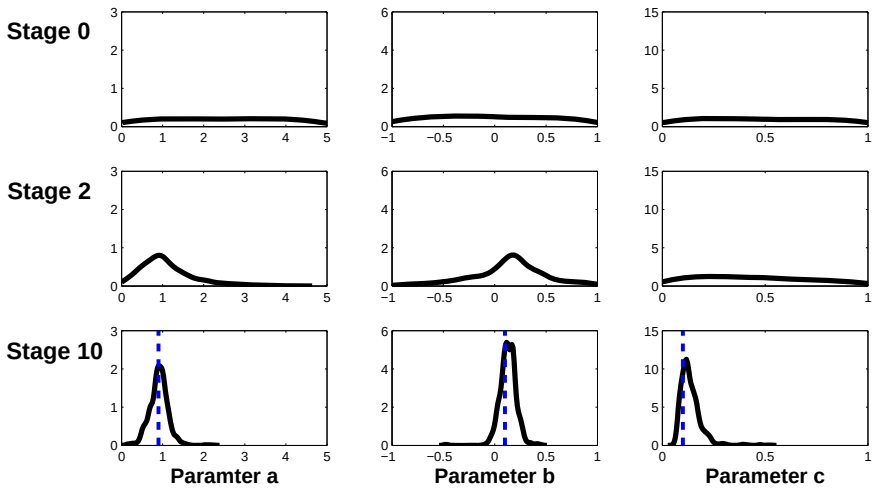

Fig. 3. Posterior parameter distributions for the three parameters of the data-generating model LIN, shown for three selected stages of experimentation. The distributions are approximated with kernel smoothing densities using the normal kernel function. The vertical dotted lines in the bottom row indicate the true parameter values of the data-generating model.

Figure 2 shows the posterior model probability of LIN plotted against the stage of experimentation. As can be seen in this figure, with optimal designs, the true data generating model (LIN) was rapidly identified after just two or three stages of experimentation. In contrast, with non-optimal, random designs, it took almost five times more observations before the true model could be confidently identified.

The posterior parameter distributions of the three parameters (a, b, c) for model LIN are shown in Figure 3 for three selected stages of experimentation. Uniform distributions were used for all three parameters as initial priors at stage 0 as shown in the top row. Examination of the posterior distributions in the figure indicate that as the stage of experimentation accumulates with more observations, the posterior distributions gradually move toward the correct parameter values (i.e., $a = 0.9, b = 0.1, c = 0.1$), at the same time becoming more peaked around the mode.

To summarize, these simulation results clearly demonstrate the efficiency of the adaptive design optimization procedure for model discrimination. With the optimal design, the correct model was identified with over 0.95 probability based on just a fraction of the observations that would be required to achieve the same level of confidence with the non-optimal, random design.

4 Conclusion

In the all-too-common situation where experimentation is costly and resources are limited, researchers must endeavor to get the most out of each and every observation. To that end, adaptive design opti-

[2] Obviously, other sensible choices of the local utility function are also possible. They include the sum of squared errors between observed and predicted data, model selection criteria, and information theoretic measures such as mutual information between model and data (see, e.g., Cavagnaro et al, 2009).

mization is a useful tool for obtaining maximially informative results from the fewest possible trials and participants. The computationally challenging problem of finding optimal designs, which was once prohibitively difficult, can now be solved practically with modern stochastic optimization techniques. We are currently appliying the method in experimentation and expanding its application to more complex experimental designs.

Acknowledgements

This research is supported by National Institute of Health Grant R01-MH57472 to JIM and MAP. Parts of this work are based on manuscripts (Myung & Pitt, 2008; Cavagnaro et al, 2009) that have been submitted for publication as journal articles. We thank Michael Rosner for the implementation of the design optimization algorithm in C++.

References

Allen, T., Yu, L., and Schmitz, J. (2003). An experimental design criterion for minimizing meta-model prediction errors applied to die casting process design. *Applied Statistics*, 52:103–117.

Amzal, B., Bois, F., Parent, E., and Robert, C. (2006). Bayesian-optimal design via interacting particle systems. *Journal of the American Statistical Association*, 101(474):773–785.

Atkinson, A. and Donev, A. (1992). *Optimum Experimental Designs*. Oxford University Press.

Bardsley, W., Wood, R., and Melikhova, E. (1996). Optimal design: A computer program to study the best possible spacing of design points for model discrimination. *Computers & Chemistry*, 20:145–157.

Box, G. and Hill, W. (1967). Discrimination among mechanistic models. *Technometrics*, 9:57–71.

Cavagnaro, D. R., Myung, J. I., Pitt, M. A. and Kujala, J. V. (2009). Adaptive design optimization: A mutual information based approach to model discrimination in cognitive science. *Submitted for publication.*

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

Chernoff, H. (1959) Sequential design of experiments. *Annals of Mathematical Statistics*, 30(3):755-770.

Doucet, A., de Freitas, N., and Gordon, N. (2001). *Sequential Monte Carlo Methods in Practice*. Springer.

El-Gamal, M. and Palfrey, T. (1996). Economical experiments: Bayesian efficient experimental design. *International Journal of Game Theory*, 25:495–517.

Gelman, A., Carlin, J. B., Stern, H., and Rubin, D. B. (2004). *Bayesian Data Analysis (2nd edition)*. Chapman & Hall/CRC.

Gilks, W. R., Richardson, S. and Spielgelhalter, D. J. (1996). *Markov Chain Monte Carlo in Practice*. Chapman & Hall.

Kiefer, J. (1959). Optimum experimental designs. *Journal of the Royal Statistical Society. Series B*, 21(2):272–319.

Müller, P., Sanso, B., and De Iorio, M. (2004). Optimal Bayesian design by inhomogeneous Markov chain simulation. *Journal of the American Statistical Association*, 99(467):788–798.

Myung, J. and Pitt, M. (2008). Optimal experimental design for model discrimination. *Submitted for publication.*

Opfer, J. and Siegler, R. (2007). Representational change and children's numerical estimation. *Cognitive Psychology*, 55:165–195.

Ponce de Leon, A. and Atkinson, A. C. (1991). Optimal experimental design for discriminating between two rival models in the presence of prior information. *Biometrika*, 78:601–608.

Siegler, R. and Booth, J. L. (2004). Development of numericall estimation in young children. *Child Development*, 75:438–444.

Ucinski, D. and Bogacka, B. (2005). T-optimum design for discrminating between two multiresponse dynamic models. *Journal of the Royal Statistical Society, Series B*, 67:3–18.