Optimal Experimental Design for Discriminating Models in Bandit Problems

Shunan Zhang, Michael D. Lee

University of California, Irvine

Question of Interest:

- Discriminating between competing models for decision-making in the bandit problems

Question of Interest:

- Discriminating between competing models for decision-making in the bandit problems
- Intuitively, we want to design the experiment in such a way that the result will support the correct model to the largest extent

- Experimental design includes different aspects

- Experimental design includes different aspects
- In this work, we care about the aspects which can be formally quantified, and could possibly affect the distribution of the outcome

- Experimental design includes different aspects
- In this work, we care about the aspects which can be formally quantified, and could possibly affect the distribution of the outcome
- We follow previous work in optimal experimental design for distinguishing psychological models by Myung and Pitt (to appear in *Psychological Review)*

Bandit problems

- a decision-maker must choose between a set of alternatives

Bandit problems

- a decision-maker must choose between a set of alternatives
- each of which has a fixed but unknown reward rate, drawn from a fixed but unknown environmental distribution

Bandit problems

- a decision-maker must choose between a set of alternatives
- each of which has a fixed but unknown reward rate, drawn from a fixed but unknown environmental distribution
- to maximize their total number of rewards over a short sequence of trials

Bandit problems

- a decision-maker must choose between a set of alternatives
- each of which has a fixed but unknown reward rate, drawn from a fixed but unknown environmental distribution
- to maximize their total number of rewards over a short sequence of trials
- the design for experiments in bandit problems is the set of reward rates

Optimal design

- An optimization problem is always associated with maximizing an appropriate utility function, or minimizing an appropriate loss function (Atkinson and Donev, 1992)

Optimal design

- An optimization problem is always associated with maximizing an appropriate utility function, or minimizing an appropriate loss function (Atkinson and Donev, 1992)
- The maximization or minimization is over the design parameter $d \in \mathcal{D}$

Optimal design

- An optimization problem is always associated with maximizing an appropriate utility function, or minimizing an appropriate loss function (Atkinson and Donev, 1992)
- The maximization or minimization is over the design parameter $d \in \mathcal{D}$
- The outcome of the experiment is defined by $p_d(y|\theta)$, i.e. the observables come from a distribution conditional on some unknown parameter vector θ

Optimal design

- An optimization problem is always associated with maximizing an appropriate utility function, or minimizing an appropriate loss function (Atkinson and Donev, 1992)
- The maximization or minimization is over the design parameter $d \in \mathcal{D}$
- The outcome of the experiment is defined by $p_d(y|\theta)$, i.e. the observables come from a distribution conditional on some unknown parameter vector θ
- The model is complete by a prior distribution $p(\theta)$ for the parameter

- Local utility is denoted by $u(d, \theta, y)$

- Local utility is denoted by $u(d, \theta, y)$
- Since the design parameter d has to be chosen before observing the experiment, we need to maximize the expectation of $u(d, \theta, y)$ with respect to (θ, y)

- Local utility is denoted by $u(d, \theta, y)$
- Since the design parameter d has to be chosen before observing the experiment, we need to maximize the expectation of $u(d, \theta, y)$ with respect to (θ, y)
- The global utility is denoted by $U(d)$.

- Local utility is denoted by $u(d, \theta, y)$
- Since the design parameter d has to be chosen before observing the experiment, we need to maximize the expectation of $u(d, \theta, y)$ with respect to (θ, y)
- The global utility is denoted by $U(d)$.
- We can formally state the design problem as

$$\begin{aligned} d^* &= \arg\max_{d \in \mathcal{D}} U(d), \ \ \mathit{where} \\ U(d) &= \int u(d, \theta, y) p_d(\theta, y) d\theta dy \\ &= \int u(d, \theta, y) p(\theta) p_d(y|\theta) d\theta dy \end{aligned} \tag{1}$$

As a preliminary study, we focus on two-armed bandit problems, and are interested in discriminating two versions of win-stay, lose-shift model

As a preliminary study, we focus on two-armed bandit problems, and are interested in discriminating two versions of win-stay, lose-shift model

- Model A: If win, stay with probability γ; If lose, shift with probability γ

As a preliminary study, we focus on two-armed bandit problems, and are interested in discriminating two versions of win-stay, lose-shift model

- Model A: If win, stay with probability γ; If lose, shift with probability γ
- Model B: If win, stay with probability γ_w; If lose, shift with probability γ_l

As a preliminary study, we focus on two-armed bandit problems, and are interested in discriminating two versions of win-stay, lose-shift model

- Model A: If win, stay with probability γ; If lose, shift with probability γ
- Model B: If win, stay with probability γ_w; If lose, shift with probability γ_l

To start with, we try to find one best pair of reward rates for one game with 8 trials.

- We use Bayes factor as the local utility.

- We use Bayes factor as the local utility.

$$\begin{aligned} U(d) = p(A) \int BF_{A/B} p(y_A|\gamma, d) p(\gamma) dy_A d\gamma \\ + p(B) \int BF_{B/A} p(y_B|\gamma_w, \gamma_l, d) p(\gamma_w) p(\gamma_l) dy_B d\gamma_w d\gamma_l \end{aligned} \quad (2)$$

- We use Bayes factor as the local utility.

$$\begin{aligned} U(d) = p(A) \int & BF_{A/B} p(y_A|\gamma, d) p(\gamma) dy_A d\gamma \\ + p(B) \int & BF_{B/A} p(y_B|\gamma_w, \gamma_I, d) p(\gamma_w) p(\gamma_I) dy_B d\gamma_w d\gamma_I \end{aligned} \tag{2}$$

Where

$$BF_{A/B} = \frac{\int p(y_A|\gamma) p(\gamma) d\gamma}{\int \int p(y_A|\gamma_w, \gamma_I) p(\gamma_w) p(\gamma_I) d\gamma_w d\gamma_I} \tag{3}$$

$$BF_{B/A} = \frac{\int \int p(y_B|\gamma_w, \gamma_I) p(\gamma_w) p(\gamma_I) d\gamma_w d\gamma_I}{\int p(y_B|\gamma) p(\gamma) d\gamma} \tag{4}$$

- We use Bayes factor as the local utility.

$$
\begin{aligned}
U(d) = p(A) \int & BF_{A/B}\, p(y_A|\gamma, d)p(\gamma)dy_A d\gamma \\
& + p(B) \int BF_{B/A}\, p(y_B|\gamma_w, \gamma_l, d)p(\gamma_w)p(\gamma_l)dy_B d\gamma_w d\gamma_l
\end{aligned}
\tag{2}
$$

Where

$$
BF_{A/B} = \frac{\int p(y_A|\gamma)p(\gamma)d\gamma}{\int\int p(y_A|\gamma_w, \gamma_l)p(\gamma_w)p(\gamma_l)d\gamma_w d\gamma_l} \tag{3}
$$

$$
BF_{B/A} = \frac{\int\int p(y_B|\gamma_w, \gamma_l)p(\gamma_w)p(\gamma_l)d\gamma_w d\gamma_l}{\int p(y_B|\gamma)p(\gamma)d\gamma} \tag{4}
$$

- With a prior distribution of the model parameters, and the formal mathematical set up of the models, we can solve this optimization task by numeric methods.

- Get the utility for each design at a grid of 0.05. The mode of the utility surface is the optimal design with respect to the prior knowledge about the model parameters

- Get the utility for each design at a grid of 0.05. The mode of the utility surface is the optimal design with respect to the prior knowledge about the model parameters
- We start with trying various priors for the model parameters to find the optimal design under these priors

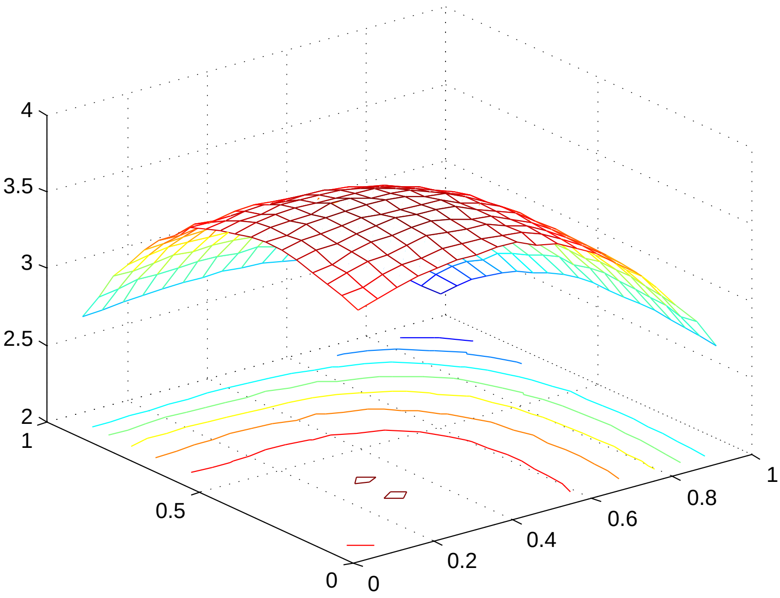

Figure: Utility surface corresponding to $\gamma \sim beta(5,1)$, $\gamma_w \sim beta(5,1)$, $\gamma_l \sim beta(4,2)$.

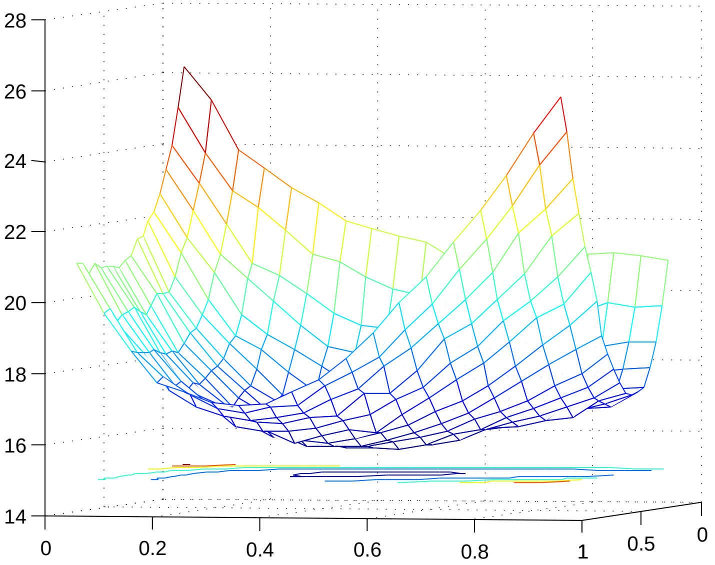

Figure: Utility surface corresponding to $\gamma \sim beta(5,1)$, $\gamma_w \sim beta(1,1)$, $\gamma_l \sim beta(1,1)$.

MCMC Sampling Approach

- We can always use prior simulation Monte Carlo to find the entire utility surface and find the mode

MCMC Sampling Approach

- We can always use prior simulation Monte Carlo to find the entire utility surface and find the mode
- But this is usually expensive

MCMC Sampling Approach

- We can always use prior simulation Monte Carlo to find the entire utility surface and find the mode
- But this is usually expensive
- Another approach (Muller, 1999) is to recast the problem as a problem of simulation from an augmented probability model $h(d, \theta, y)$, where

$$h(d, \theta, y) \propto u(d, \theta, y)p(\theta)p(y|\theta) \tag{5}$$

MCMC Sampling Approach

- We can always use prior simulation Monte Carlo to find the entire utility surface and find the mode
- But this is usually expensive
- Another approach (Muller, 1999) is to recast the problem as a problem of simulation from an augmented probability model $h(d, \theta, y)$, where

$$h(d, \theta, y) \propto u(d, \theta, y) p(\theta) p(y|\theta) \tag{5}$$

- Under h, the marginal distribution in d is proportional to $U(d)$, i.e., $h(d)$ is proportional to $U(d)$, as desired

MCMC Sampling Approach

- We can always use prior simulation Monte Carlo to find the entire utility surface and find the mode
- But this is usually expensive
- Another approach (Muller, 1999) is to recast the problem as a problem of simulation from an augmented probability model $h(d, \theta, y)$, where

$$h(d, \theta, y) \propto u(d, \theta, y) p(\theta) p(y|\theta) \tag{5}$$

- Under h, the marginal distribution in d is proportional to $U(d)$, i.e., $h(d)$ is proportional to $U(d)$, as desired
- Equivalent to sampling from a distribution which is proportional to the utility

Annealing Procedure

- Expected utility surfaces are flat over a wide range of designs

Annealing Procedure

- Expected utility surfaces are flat over a wide range of designs
- We follow an annealing process introduced by Muller(1995), consider

$$h_J(d, \theta_1, y_1, \ldots, \theta_J, y_J) \propto \prod_{j=1}^{J} u(d, \theta_j, y_j) p(\theta_j) p(y_j | \theta_j) \quad (6)$$

Annealing Procedure

- Expected utility surfaces are flat over a wide range of designs
- We follow an annealing process introduced by Muller(1995), consider

$$h_J(d, \theta_1, y_1, ..., \theta_J, y_J) \propto \prod_{j=1}^{J} u(d, \theta_j, y_j) p(\theta_j) p(y_j|\theta_j) \quad (6)$$

- The implied marginal in d is proportional to the J-th power of the expected utility, $h_J(d) \propto U^J(d)$.

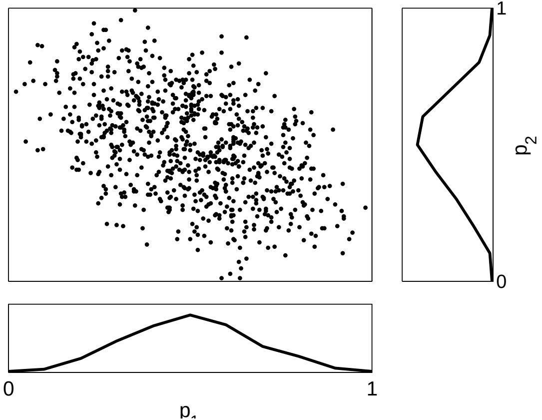

Figure: Samples of optimal designs. $\gamma \sim beta(1,1)$, $\gamma_w \sim beta(1,1)$, $\gamma_l \sim beta(1,1)$.

The mode is $p_1 = .50, p_2 = .50$.

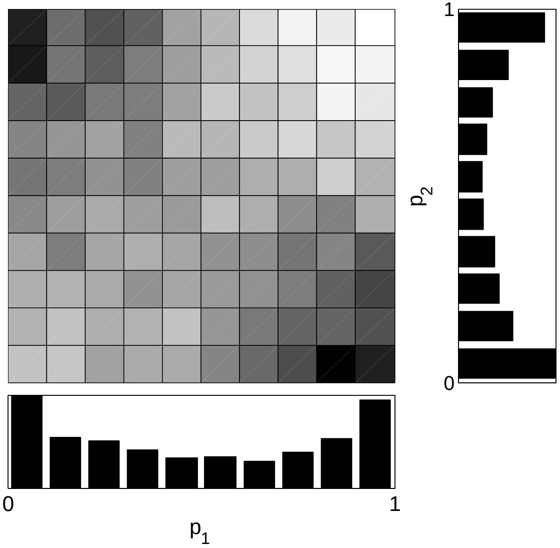

Figure: Samples of optimal designs. $\gamma \sim beta(18.87, 7.791)$, $\gamma_w \sim beta(2.872, 0.680)$, $\gamma_l \sim beta(2.318, 1.818)$.

Optimal design to distinguish the WSLS model and the ϵ-greedy model

Optimal design to distinguish the WSLS model and the ϵ-greedy model

- ϵ-greedy model is one standard model of the bandit problems from machine learning

Optimal design to distinguish the WSLS model and the ϵ-greedy model

- ϵ-greedy model is one standard model of the bandit problems from machine learning
- On each trial, the reward rates are estimated by the decision-maker

Optimal design to distinguish the WSLS model and the ϵ-greedy model

- ϵ-greedy model is one standard model of the bandit problems from machine learning
- On each trial, the reward rates are estimated by the decision-maker
- With probability ϵ, choose randomly

Optimal design to distinguish the WSLS model and the ϵ-greedy model

- ϵ-greedy model is one standard model of the bandit problems from machine learning
- On each trial, the reward rates are estimated by the decision-maker
- With probability ϵ, choose randomly
- With probability 1-ϵ, choose the alternative with the highest estimated reward rate

Optimal design to distinguish the WSLS model and the ϵ-greedy model

- ϵ-greedy model is one standard model of the bandit problems from machine learning
- On each trial, the reward rates are estimated by the decision-maker
- With probability ϵ, choose randomly
- With probability 1-ϵ, choose the alternative with the highest estimated reward rate
- Find optimal design to distinguish this model and WSLS model, using empirical priors for the model parameters

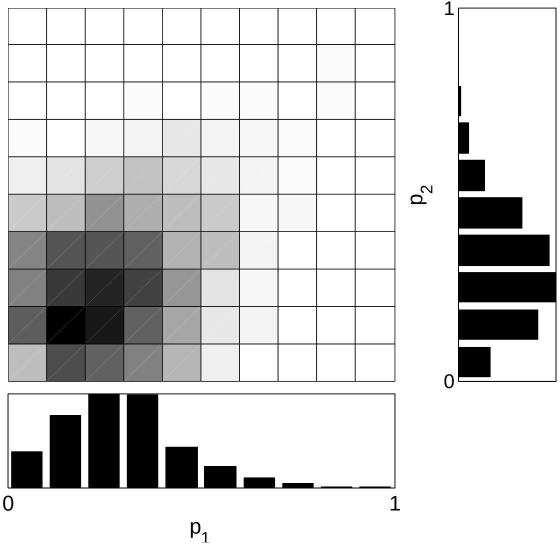

Figure: Samples of optimal designs. $\gamma \sim \mathit{beta}(18.87, 7.791)$, $\epsilon \sim \mathit{beta}(5.8, 11.3)$.

Discussion

Discussion

- Given quantitative models and prior knowledge about the distribution of model parameters, we are able to design an experiment in order to get data which will discriminating between these models to the largest extent.

Discussion

- Given quantitative models and prior knowledge about the distribution of model parameters, we are able to design an experiment in order to get data which will discriminating between these models to the largest extent.
- Our application of the design optimization algorithm had the following findings:

Discussion

- Given quantitative models and prior knowledge about the distribution of model parameters, we are able to design an experiment in order to get data which will discriminating between these models to the largest extent.
- Our application of the design optimization algorithm had the following findings:
  - To discriminate between the two versions of WSLS model, the best type of experimental design is to pick two extreme reward rates, one large and one small

Discussion

- Given quantitative models and prior knowledge about the distribution of model parameters, we are able to design an experiment in order to get data which will discriminating between these models to the largest extent.
- Our application of the design optimization algorithm had the following findings:
  - To discriminate between the two versions of WSLS model, the best type of experimental design is to pick two extreme reward rates, one large and one small
  - to discriminate between the WSLS model and ϵ-greedy model, the best type of experimental design is to pick up two moderately small reward rates

Discussion

- Given quantitative models and prior knowledge about the distribution of model parameters, we are able to design an experiment in order to get data which will discriminating between these models to the largest extent.
- Our application of the design optimization algorithm had the following findings:
  - To discriminate between the two versions of WSLS model, the best type of experimental design is to pick two extreme reward rates, one large and one small
  - to discriminate between the WSLS model and ϵ-greedy model, the best type of experimental design is to pick up two moderately small reward rates
- Individual difference in the prior for model parameters, as well as the prior for models, are interesting questions to be addressed in future works

Thank You!

We assess this integral by prior simulation Monte Carlo. With prior knowledge about the model parameters, for one p_i pair, i.e. one design d,

1. draw γ from $p(\gamma)$; draw γ_w from $p(\gamma_w)$; draw γ_l from $p(\gamma_l)$;
2. suppose model A is the true model and let it generate data for one game with 8 trials, with parameter γ. This realizes a draw from $p(y_A|d,\gamma)$;
3. calculate the likelihood of the simulated data under model A over that under model B. This gives the Bayes factor $BF_{A/B}$.
4. realizes a draw from $p(y_B|d,\gamma_w,\gamma_l)$;
5. calculate $BF_{B/A}$.
6. sum up utility $= BF_{A/B} + BF_{B/A}$
7. repeat 1-6 for J times, calculate the average utility associated with the current design d.

Muller Algorithm with Annealing: MCMC scheme with stationary distribution $h(d, \theta, y)$.

1. start with a design d^0;
2. at d^t, simulate (γ_j^t, y_{Aj}^t) and $(\gamma_{wj}^t, \gamma_{lj}^t, y_{Bj}^t)$, $j = 1, ...J$, for each simulated experiment evaluate $BF_{jA/B}^t$ and $BF_{jB/A}^t$;
3. evaluate $w^t = \sum_{j=1}^J \log(BF_{jA/B}^t + BF_{jB/A}^t)$;
4. propose a new candidate design $\tilde{d}^t$ from a symmetric proposal distribution (e.g. $N(d^t, \sigma^2 I)$);
5. at $\tilde{d}^t$, simulate $(\tilde{\gamma}_j^t, \tilde{y}_{Aj}^t)$ and $(\tilde{\gamma}_{wj}^t, \tilde{\gamma}_{lj}^t, \tilde{y}_{Bj}^t)$, $j = 1, ...J$, for each simulated experiment evaluate $\tilde{BF}_{jA/B}^t$ and $\tilde{BF}_{jB/A}^t$;
6. evaluate $\tilde{w}^t = \sum_{j=1}^J \log(\tilde{BF}_{jA/B}^t + \tilde{BF}_{jB/A}^t)$;
7. evaluate the acceptance probability defined as $AP = min(1, e^{\tilde{w}^t - w^t})$, accept candidate with AP;
8. set $t = t + 1$, repeat 2-7 until convergence, all accepted d^t's thereafter should represent an optimal design solution.