Sequential Experimentation: Theory and Principles

Bobby Kleinberg

Cornell University

Yale Econ 421: Designing the Digital Economy
Guest lecture, 17 October 2017

Sequential Experimentation

Categories of problems

- Sequential experimental design
  Given two or more hypotheses, and one or more experiments whose outcome distributions differ under various hypotheses, design a procedure to test which hypothesis is true.
  E.g., *Hodgkin or non-Hodgkin lymphoma?*

Sequential Experimentation

Categories of problems

- Sequential experimental design
- Multi-armed bandit
  Given two or more actions, each of which produces stochastic payoffs sampled from an unknown stationary distribution, design a procedure to maximize average payoff over time.
  E.g., *Choose a color for the "donate" button on our site.*

Sequential Experimentation

Categories of problems

- Sequential experimental design
- Multi-armed bandit
- Best arm identification
  Given two or more actions as in the multi-armed bandit problem, design a procedure to find the one with the highest average payoff.
  E.g., *Which of these drugs is most effective at treating high blood pressure?*

Sequential Experimentation

Categories of problems

- Sequential experimental design
- Multi-armed bandit
- Best arm identification

Modes of analysis

- Bayesian: Optimize average-case performance under some prior distribution on the true state of the world.
- Minimax: Optimize worst-case performance over all potential states of the world.

Sequential Experimentation

Categories of problems

- Sequential experimental design
- Multi-armed bandit
- Best arm identification

Modes of analysis

- Bayesian: Optimize average-case performance under some prior distribution on the true state of the world.
- Minimax: Optimize worst-case performance over all potential states of the world.

Key Techniques

Chernoff bound
a non-asymptotic form of the law of large numbers, used to justify that certain procedures have high probability of success

Kullback-Leibler divergence
an information theoretic measure of the distinguishability of probability distributions, used to prove that certain procedures are (nearly) optimal

Example #1: Biased coin testing

Hypothesis A: fair coin
Hypothesis B: $\Pr(\text{heads}) = \frac{1+\varepsilon}{2}$

Design a procedure to:

- toss coin repeatedly
- eventually stop and guess A or B
- ensure $\Pr(\text{error}) < \delta$ in both cases.

Try to minimize expected coin tosses.

Fixed Design Procedure: Toss coin s times, guess A unless empirical frequency of heads exceeds $\frac{1}{2} + \frac{\varepsilon}{4}$.

Analysis of fixed design procedure

Theorem (Chernoff-Hoeffding)

If $X_1, \ldots, X_s$ are independent random variables supported in $[0, 1]$, and $\bar{X} = \frac{1}{s}\sum_{i=1}^{s} X_i$, then $\Pr(\bar{X} - \mathbb{E}\bar{X} > \gamma) < \exp(-2\gamma^2 s)$.

Analysis of fixed design procedure with s samples.

An error (under either hypothesis) requires empirical frequency to differ from its expected value by more than $\gamma = \varepsilon/4$.

Hence $\Pr(\text{error}) < \exp(-\frac{1}{8}\varepsilon^2 s)$.

To make this less than δ, set $s > 8\log(1/\delta)/\varepsilon^2$.

E.g., for $\varepsilon = 0.1, \delta = 0.05$, 2400 samples suffice.

Analysis of fixed design procedure

Theorem (Chernoff-Hoeffding)

If $X_1, \ldots, X_s$ are independent random variables supported in $[0, 1]$, and $\bar{X} = \frac{1}{s}\sum_{i=1}^{s} X_i$, then $\Pr(\bar{X} - \mathbb{E}\bar{X} > \gamma) < \exp(-2\gamma^2 s)$.

Two quantitative hallmarks of optimal experimentation.

- $1/\varepsilon^2$ independent samples suffice to distinguish distributions that differ by ε,
- Inflating sample complexity by $\log(1/\delta)$ boosts confidence to $1 - \delta$.

Kullback-Leibler divergence: definition

Definition

KL-divergence If p, q are two distributions on a finite set Ω,

$$D(p \| q) = \sum_{x \in \Omega} p(x) \log \left(\frac{p(x)}{q(x)} \right).$$

More generally, if p and q are probability measures on a measure space Ω,

$$D(p \| q) = \int \log \left(\frac{dp}{dq} \right) dp(x),$$

where dp/dq denotes the Radon-Nikodym derivative.

Kullback-Leibler divergence: key properties

- For all p, q, $D(p \| q) \geq 0$ with equality if and only if $p = q$.
- If $\mathrm{Ber}(r)$ denotes the Bernoulli distribution with parameter r, then $D(\mathrm{Ber}(\frac{1}{2}) \,\|\, \mathrm{Ber}(\frac{1+\varepsilon}{2})) < \varepsilon^2$ for $\varepsilon < 8/9$.
- **(High-probability Pinsker inequality)** If $\Omega = A \cup B$ then $p(B) + q(A) \geq \frac{1}{2}\exp(-D(p \| q))$.
- **(Chain rule for KL-divergence)** If p, q are probability distributions on a sequence $\mathbf{x} = (x_1, \ldots, x_n) \in \Omega_1 \times \cdots \times \Omega_n$,

$$D(p \| q) = \sum_{k=1}^{n} \mathop{\mathbb{E}}_{\mathbf{x} \sim p} \left[D(p(x_k | x_1, \ldots, x_{k-1}) \,\|\, q(x_k | x_1, \ldots, x_{k-1})) \right].$$

Divergence decomposition lemma

Let $\mathfrak{I}$ be a set of experiments and let p_i, q_i denote the distribution of outcomes for experiment $i \in \mathfrak{I}$ under hypotheses p, q, resp.

Let π be a sequential experimentation protocol and let p^π, q^π denote the distributions of sequences produced under p, q, resp.

Lemma (Divergence decomposition lemma)

If $S_i(\pi)$ is the random variable denoting the number of times experiment i is performed under protocol π, then

$$D(p^\pi \| q^\pi) = \sum_{i \in \mathfrak{I}} \mathbb{E}_p[S_i(\pi)] \cdot D(p_i \| q_i).$$

Proof is an application of the chain rule.

Kullback-Leibler divergence: key properties

- For all p, q, $D(p\|q) \geq 0$ with equality if and only if $p = q$.
- If $\mathrm{Ber}(r)$ denotes the Bernoulli distribution with parameter r, then $D(\mathrm{Ber}(\frac{1}{2}) \,\|\, \mathrm{Ber}(\frac{1+\varepsilon}{2})) < \varepsilon^2$ for $\varepsilon < 8/9$.
- **(High-probability Pinsker inequality)** If $\Omega = A \cup B$ then

$$p(B) + q(A) \geq \tfrac{1}{2} \exp(-D(p\|q)).$$

- **(Divergence decomposition)**

$$D(p^\pi \| q^\pi) = \textstyle\sum_{i \in \mathfrak{I}} \mathbb{E}_p[S_i(\pi)] \cdot D(p_i \| q_i).$$

Typical use case:

Low error probability + Pinsker $\Rightarrow$ Lower bound on $D(p^\pi \| q^\pi)$
Divergence decomposition $\Rightarrow$ Lower bound on $\mathbb{E}_p[S_i(\pi)]$.

Approximate optimality of fixed design

Let p, q denote the outcome distributions under Hypothesis A (fair coin) and Hypothesis B (bias $\frac{1+\varepsilon}{2}$) respectively.

Let A, B denote the events "guess A", "guess B".

If π is a procedure satisfying $p^\pi(B) < \delta$ and $q^\pi(A) < \delta$ then

$$2\delta > p^\pi(B) + q^\pi(A) \geq \frac{1}{2}\exp(-D(p^\pi \| q^\pi))$$

so $D(p^\pi \| q^\pi) \geq \log(1/4\delta)$.

Approximate optimality of fixed design

Let p, q denote the outcome distributions under Hypothesis A (fair coin) and Hypothesis B (bias $\frac{1+\varepsilon}{2}$) respectively.

Let A, B denote the events "guess A", "guess B".

If $p^\pi(B) < \delta$ and $q^\pi(A) < \delta$ then $D(p^\pi \| q^\pi) \geq \log(1/4\delta)$.

Approximate optimality of fixed design

Let p, q denote the outcome distributions under Hypothesis A (fair coin) and Hypothesis B (bias $\frac{1+\varepsilon}{2}$) respectively.

Let A, B denote the events "guess A", "guess B".

If $p^\pi(B) < \delta$ and $q^\pi(A) < \delta$ then $D(p^\pi \| q^\pi) \geq \log(1/4\delta)$.

If $S(\pi)$ denotes the number of coin tosses, then

$$\begin{aligned}\log(1/4\delta) \leq D(p^\pi \| q^\pi) &= \mathbb{E}[S(\pi)] \cdot D(\mathrm{Ber}(\tfrac{1}{2}) \,\|\, \mathrm{Ber}(\tfrac{1+\varepsilon}{2})) \\ &< \mathbb{E}[S(\pi)] \cdot \varepsilon^2.\end{aligned}$$

Hence $\log(1/4\delta)/\varepsilon^2$ samples are required, in expectation.

Biased coin detection: executive summary

To distinguishing a fair coin from an ε-biased coin with error probability δ, $O(\log(1/\delta)/\varepsilon^2)$ samples are necessary and sufficient.

Upper bound: concentration of measure (Chernoff-Hoeffding) shows that for large sample size, the sample average is probably close enough to the true expectation.

Lower bound: sample size must be large enough to push the KL-divergence between null hypothesis and experimental hypothesis above a confidence threshold.

Example #2: Best arm identification

Design a procedure to select one out of n coins.

One step = pick any coin and toss it.

Ensure that bias of selected coin is ε-close to maximum bias, with probability at least $1 - \delta$. *("procedure is (ε, δ)-PAC")*

Goal: minimize expected number of steps.

Fixed design for best arm identification

Fixed design procedure: Flip each coin s times, select the one with highest empirical frequency.

To make an incorrect selection, either the best coin or the selected coin must deviate from its expected frequency by at least $\varepsilon/2$.

Probability of any coin deviating by $\varepsilon/2$ is less than $\exp(-\frac{1}{2}\varepsilon^2 s)$.

So $s = 2\log(2/\delta)/\varepsilon^2$ suffices? (2 bad events, each of prob $< \delta/2$)

Fixed design for best arm identification

Fixed design procedure: Flip each coin s times, select the one with highest empirical frequency.

To make an incorrect selection, either the best coin or the selected coin must deviate from its expected frequency by at least $\varepsilon/2$.

Probability of any coin deviating by $\varepsilon/2$ is less than $\exp(-\frac{1}{2}\varepsilon^2 s)$.

So $s = 2\log(2/\delta)/\varepsilon^2$ suffices? (2 bad events, each of prob $< \delta/2$)

No! Watch out for selection bias.

There are $n-1$ suboptimal coins. If one of them deviates by $\varepsilon/2$ we are more likely to select it. So

$$\Pr(\text{selected coin deviates}) \gg \Pr(\text{coin } i \text{ deviates}).$$

Fixed design for best arm identification

Fixed design procedure: Flip each coin s times, select the one with highest empirical frequency.

To make an incorrect selection, at least one of the n coins must deviate from its expected frequency by at least $\varepsilon/2$.

Probability of any coin deviating by $\varepsilon/2$ is less than $\exp(-\frac{1}{2}\varepsilon^2 s)$.

So $s = 2\log(n/\delta)/\varepsilon^2$ suffices. (n bad events, each of prob $< \delta/n$)

Total sample complexity is $2n\log(n/\delta)/\varepsilon^2 = O_{\varepsilon,\delta}(n \log n)$.

Lower bound for best arm identification

Define null model p: coin 1 has bias $\frac{1}{2} + \varepsilon$, all others have bias $\frac{1}{2}$.

For $i = 2, \ldots, n$, define alternative model q_i:
same as p except coin i has bias $\frac{1}{2} + 2\varepsilon$.

Lower bound for best arm identification

Define null model p: coin 1 has bias $\frac{1}{2} + \varepsilon$, all others have bias $\frac{1}{2}$.

For $i = 2, \ldots, n$, define alternative model q_i:
same as p except coin i has bias $\frac{1}{2} + 2\varepsilon$.

Let B_i = "select i", A_i = "don't select i".

$$2\delta > p^\pi(B_i) + q_i^\pi(A_i) \geq \frac{1}{2} \exp(-D(p^\pi \| q_i^\pi))$$

Lower bound for best arm identification

Define null model p: coin 1 has bias $\frac{1}{2}+\varepsilon$, all others have bias $\frac{1}{2}$.

For $i = 2, \ldots, n$, define alternative model q_i:
same as p except coin i has bias $\frac{1}{2}+2\varepsilon$.

Let B_i = "select i", A_i = "don't select i".

$$\begin{aligned}
2\delta > p^\pi(B_i) + q_i^\pi(A_i) &\geq \frac{1}{2}\exp(-D(p^\pi \| q_i^\pi)) \\
\log(1/4\delta) &\leq D(p^\pi \| q_i^\pi) \\
&= \frac{1}{2}\mathbb{E}_p[S_i(\pi)] \cdot D(\mathrm{Ber}(\tfrac{1}{2}) \,\|\, \mathrm{Ber}(\tfrac{1}{2}+2\varepsilon)) \\
&< 8\varepsilon^2 \mathbb{E}_p[S_i(\pi)].
\end{aligned}$$

Lower bound for best arm identification

Define null model p: coin 1 has bias $\frac{1}{2} + \varepsilon$, all others have bias $\frac{1}{2}$.

For $i = 2, \ldots, n$, define alternative model q_i:
same as p except coin i has bias $\frac{1}{2} + 2\varepsilon$.

Let $B_i =$ "select i", $A_i =$ "don't select i".

$$
\begin{aligned}
2\delta > p^\pi(B_i) + q_i^\pi(A_i) &\geq \frac{1}{2}\exp(-D(p^\pi \| q_i^\pi)) \\
\log(1/4\delta) &\leq D(p^\pi \| q_i^\pi) \\
&= \frac{1}{2}\mathbb{E}_p[S_i(\pi)] \cdot D(\mathrm{Ber}(\tfrac{1}{2}) \,\|\, \mathrm{Ber}(\tfrac{1}{2} + 2\varepsilon)) \\
&< 8\varepsilon^2 \mathbb{E}_p[S_i(\pi)]. \\
\sum_{i=2}^{n} \mathbb{E}_p[S_i(\pi)] &> \log\left(\frac{\ln(1/4\delta)}{8\varepsilon^2}\right)(n-1).
\end{aligned}
$$

Sequential design vs. fixed design

For fixed design, $O(n \log(n/\delta)/\varepsilon^2)$ samples suffice.

For any sequential protocol, at least $O(n \log(1/\delta)/\varepsilon^2)$ are necessary.

These almost match, but what about the $\log(n)$?

Sequential design vs. fixed design

For fixed design, $O(n \log(n/\delta)/\varepsilon^2)$ samples suffice.

For any sequential protocol, at least $O(n \log(1/\delta)/\varepsilon^2)$ are necessary.

These almost match, but what about the $\log(n)$?

For fixed design procedures, the $\log(n)$ factor turns out to be unavoidable.

But there is a sequential procedure that is (ε, δ)-PAC and avoids the $\log(n)$ factor: **median elimination**. (Even-Dar et al., 2006)

Median elimination

Idea: run in phases, with each phase eliminating half of the remaining arms, until only 1 remains.

Design goal for phase j: with probability at least $1 - \delta_j$, the bias of the best remaining arm decreases by at most ε_j.

Setting $\delta_j = \delta/2^j$ and $\varepsilon_j = \frac{1}{3}\left(\frac{3}{4}\right)^j$ will then ensure the (ε, δ)-PAC property.

Median elimination

Idea: run in phases, with each phase eliminating half of the remaining arms, until only 1 remains.

Design goal for phase j: with probability at least $1 - \delta_j$, the bias of the best remaining arm decreases by at most ε_j.

In phase j, sample each arm s_j times.
For any arm i, $\Pr(\text{arm } i \text{ error} > \frac{1}{2}\varepsilon_j) < \exp(-\frac{1}{2}\varepsilon_j^2 s_j) \leq \frac{1}{3}\delta_j$ if $s_j = \lceil 2 \log(3/\delta_j)\varepsilon_j^{-2} \rceil$.

Eliminating *every* ε_j-good arm requires one of two bad events:

1. Error on best arm: probability $\frac{1}{3}\delta_j$.
2. Fraction of errors on other arms exceeds $1/2$: by Markov, probability $\left(\frac{1}{3}\delta_j\right) / \left(\frac{1}{2}\right) \leq \frac{2}{3}\delta_j$.

Sample complexity of median elimination

$$s_j = \left\lceil 2 \log\left(3/\delta_j\right) \varepsilon_j^{-2} \right\rceil = O\left(\frac{\log(1/\delta)}{\varepsilon^2} \cdot j \cdot (4/3)^{2j} \right)$$

Sample complexity of median elimination

$$s_j = \left\lceil 2 \log\left(3/\delta_j\right) \varepsilon_j^{-2} \right\rceil = O\left(\frac{\log(1/\delta)}{\varepsilon^2} \cdot j \cdot (4/3)^{2j} \right)$$

Sample complexity:

$$\sum_{j=1}^{\log_2 n} s_j \cdot (n/2^j) = O\left(n \cdot \frac{\log(1/\delta)}{\varepsilon^2} \cdot \sum_{j=1}^{\infty} j (4/3)^{2j} 2^{-j} \right)$$

$$= O\left(n \cdot \frac{\log(1/\delta)}{\varepsilon^2} \right)$$

matching the information-theoretic lower bound up to constant factors.

Best arm identification: executive summary

To select an arm that is ε-close to optimal with probability $1-\delta$, $O(n\log(1/\delta)/\varepsilon^2)$ samples are necessary and sufficient.

Fixed design procedures, which sample each arm a pre-specified number of times, must inflate the number of samples by a factor of $\log(n)$ in order to mitigate selection bias.

Median elimination culls unpromising arms, draws more samples from the surviving ones to improve estimation accuracy. This mitigates selection bias in a more sample-efficient manner.

Example #3: Multi-armed bandits

Given: n arms; arm i produces random payoffs $R_{i,t} \in [0, 1]$ by sampling from unknown stationary distribution F_i.

One step = pull an arm, observe its reward.

Let μ_i = expected payoff of arm i, $\mu_* = \max\{\mu_i\}$.

Regret of policy π at time T:

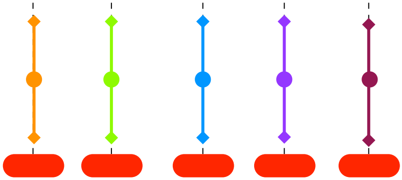
$$R(\pi, T) = \mu_* T - \mathbb{E}\left[\sum_{t=1}^{T} \mu_{\pi(t)}\right].$$

Goal: minimize regret.

"Exploration vs. exploitation" trade-off: pulling suboptimal arms has an opportunity cost, but is necessary in order to confidently identify the best arm.

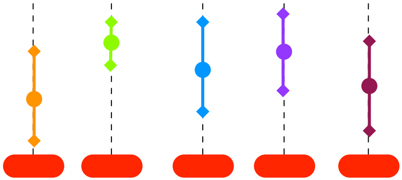
The UCB1 Algorithm (Auer et al., 2002)

1. Play each arm once.
2. Thenceforward, maintain a confidence interval for each arm, centered at its empirical average.



The UCB1 Algorithm (Auer et al., 2002)

1. Play each arm once.
2. Thenceforward, maintain a confidence interval for each arm, centered at its empirical average.
3. Always pull arm with highest upper confidence bound (UCB).



The UCB1 Algorithm (Auer et al., 2002)

1. Play each arm once.
2. Thenceforward, maintain a confidence interval for each arm, centered at its empirical average.
3. Always pull arm with highest upper confidence bound (UCB).

An arm sampled s times in first t steps has conf radius $\sqrt{\frac{\log t}{s}}$.

Hoeffding: $\Pr(\text{true mean outside conf interval}) = O(t^{-2})$.

Call a time step "weird" if at least one arm violates its confidence interval. $\mathbb{E}[\#\text{ weird steps}] < n \sum_{t=1}^{\infty} t^{-2} = \frac{\pi^2}{6} n$.

The UCB1 Algorithm (Auer et al., 2002)

1. Play each arm once.
2. Thenceforward, maintain a confidence interval for each arm, centered at its empirical average.
3. Always pull arm with highest upper confidence bound (UCB).

An arm sampled s times in first t steps has conf radius $\sqrt{\frac{\log t}{s}}$.

Excluding weird time steps, arm i with $\mu_* - \mu_i = \Delta_i > 0$ is only pulled if its confidence interval is wide enough to bridge the gap to the best arm: $2\sqrt{\frac{\log t}{s_i}} \geq \Delta_i$.

Among first T time steps, at most $\frac{4\log(T)}{\Delta_i^2}$ non-weird pulls of arm i.

$$R(UCB1, T) < \frac{\pi^2}{6} n + 4\log(T) \sum_{i:\Delta_i > 0} \frac{1}{\Delta_i}$$

The UCB1 Algorithm (Auer et al., 2002)

1. Play each arm once.
2. Thenceforward, maintain a confidence interval for each arm, centered at its empirical average.
3. Always pull arm with highest upper confidence bound (UCB).

An arm sampled s times in first t steps has conf radius $\sqrt{\frac{\log t}{s}}$.

$$R(UCB1, T) < \frac{\pi^2}{6} n + 4 \log(T) \sum_{i : \Delta_i > 0} \frac{1}{\Delta_i}$$

Remark: A KL-divergence argument (omitted) establishes that this is optimal, up to constant factors.

Epilogue: Price experimentation

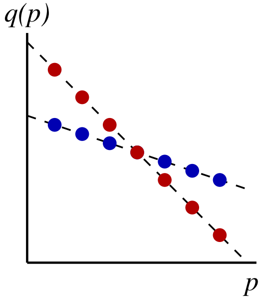
Given: sequence of prices $0 < p_1 < p_2 < \cdots < p_n \leq 1$.

Stream of consumers with values $v_t \sim F$. (F unknown.)

One step = offer price p_i, which is accepted iff $v_t > p_i$.

Hypothesis A:
$\Pr(v_t > p) = q_A(p) = 1 - p$.

Hypothesis B:
$\Pr(v_t > p) = q_B(p) = \frac{2-p}{3}$.



Epilogue: Price experimentation

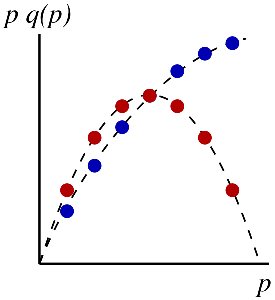
Given: sequence of prices $0 < p_1 < p_2 < \cdots < p_n \leq 1$.

Stream of consumers with values $v_t \sim F$. (F unknown.)

One step = offer price p_i, which is accepted iff $v_t > p_i$.

Hypothesis A:
$\Pr(v_t > p) = q_A(p) = 1 - p$.

Hypothesis B:
$\Pr(v_t > p) = q_B(p) = \frac{2-p}{3}$.



Epilogue: Price experimentation

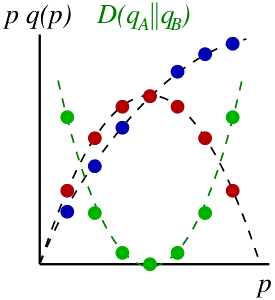
Given: sequence of prices $0 < p_1 < p_2 < \cdots < p_n \leq 1$.

Stream of consumers with values $v_t \sim F$. (F unknown.)

One step = offer price p_i, which is accepted iff $v_t > p_i$.

Hypothesis A:
$\Pr(v_t > p) = q_A(p) = 1 - p$.

Hypothesis B:
$\Pr(v_t > p) = q_B(p) = \frac{2-p}{3}$.



Some contemporary research topics

- **Reinforcement learning:** The environment has an internal state (which may or may not be observable) that evolves over time and governs the distribution of outcomes for each arm.
- **Contextual bandits:** Before choosing arm $\pi(t)$, you can observe "side information" x_t.
- **Strategic aspects:** Time steps or arms (or both) can be strategic agents. For example, "pulling an arm" is making a recommendation to a user, but users will only follow recommendations they believe to be beneficial.