Modern Sampling Methods

Class 9: Statistical Inference with Adaptively Generated Data

January 11, 2022

Outline

- Sources of Bias
  - Adaptive sampling
  - Adaptive stopping
  - Adaptive arm choice
- Challenges to Inference
  - Fixed Arm Hypotheses
  - Best Arm Hypotheses
  - Sharp Null Hypotheses

Bandit Algorithm

Round j

(a) Start with history - arms/treatments, outcomes, (and possibly covariates including round j covariates)

(b) Based on history, form arm/treatment probabilities

(c) Based on these probabilities, randomly pick arm/treatment, T_j

(d) Based on arm/treatment, realize corresponding outcome, Y_j

(e) Check stopping rule: stop or advance to the next round

Assignment Mechanism in step (b)
As seen in Classes 7 & 8, this mechanism can be chosen for a variety of objectives aside from statistical inference.

Bandit produces data:

$$
\begin{array}{c}
(T_1, Y_1) \\
\vdots \\
(T_n, Y_n)
\end{array}
$$

where n could be random.

Can we use this data for estimation and inference as usual?
In general, no.

- Sample average of outcomes can be a *biased* estimator
- Usual t-tests from regression may be *invalid*
- Fisher's exact test still valid (for the sharp null)

Example: Explore-Then-Commit

Two Treatment Arms: $\mathcal{T} = \{0, 1\}$

$$Y_i(0) \sim N(\mu_0, 1) \qquad (\text{fix } \ \mu_0 = 0)$$

$$Y_i(1) \sim N(\mu_1, 1)$$

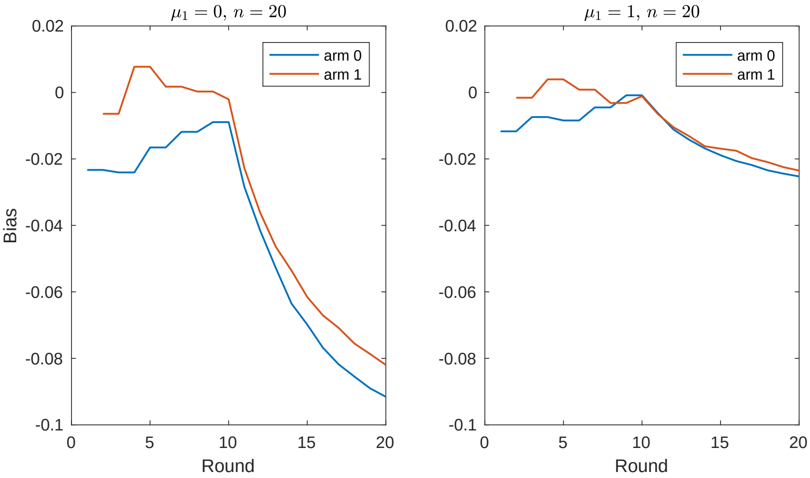
Recall ETC Algorithm with *two* arms and tuning parameter m :

1. Explore: in first $2m$ rounds, alternate between the arms. $\left(\text{set } 2m = \frac{n}{2}\right)$
2. Commit: after round $2m$, always choose the arm with highest value of

$$\hat{\mu}_t(j) = \frac{1}{N_t(j)} \sum_{i=1}^{n} 1(T_i = t) Y_i$$

Example. Bias Simulation in ETC Example

$$\text{Bias}_{t,j} = E[\hat{\mu}_t(j)] - \mu_t \text{ for } t = 0, 1.$$

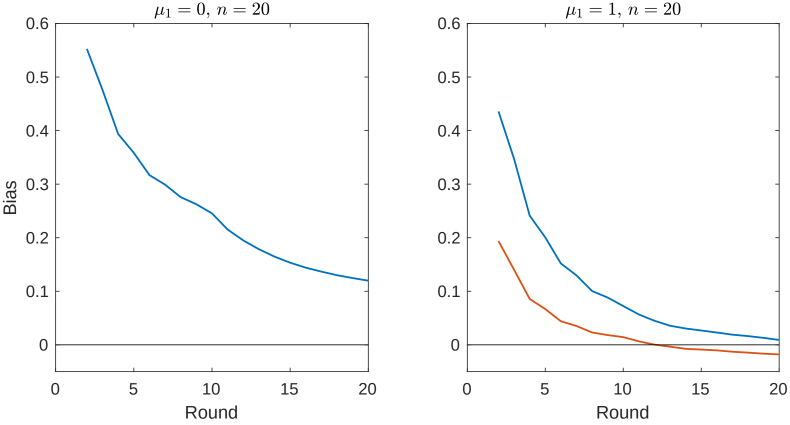


"Best arm" at round j: $T_j^* = \arg\max_{t \in \mathcal{T}} \hat{\mu}_t(j)$

(Best arm) $\text{Bias}_j^* = E[\hat{\mu}_{T_j^*}(j)] - (\max_{t \in \mathcal{T}} \mu_t)$

vs.

(Best arm) $\text{Bias}_j^{**} = E[\hat{\mu}_{T_j^*}(j)] - \mu_{T_j^*}$



Sample Average Bias

- Sampling Rule: How does the frequency an arm is chosen $(N_t(j))$ depend on realized values of that arm's outcomes?

  e.g. ECT, UCB, Thompson
  If the count of an arm is *increasing* in the realized values of that arm's outcomes, then the sample average for that arm is more *negatively* biased.

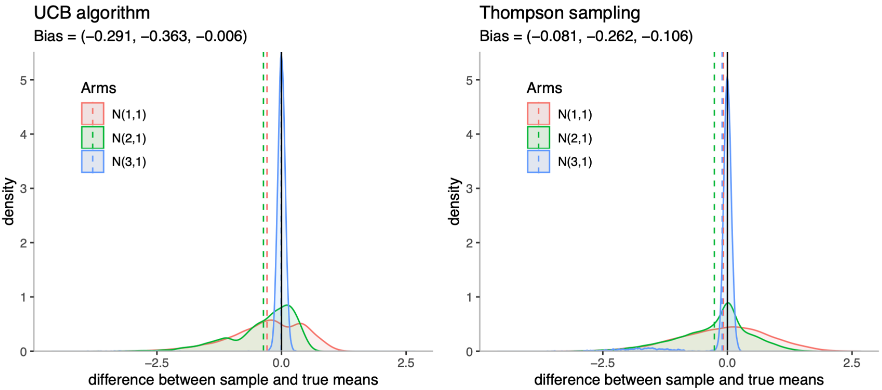
- Stopping Time: How does the stopping time depend on the realized values of outcomes?

  If the algorithm stops *earlier* as realized outcomes of a fixed arm *increase*, then the sample average for the fixed arm is more *positively* biased.

- Choice of Arm: How does the arm of interest depend on the realized values of outcomes?

  e.g. "Best arm":
  (a) If a given arm is *more likely* to be chosen as its realized outcome values *increase*, then the sample average for the chosen arm is more *positively* biased (Bias^{**})
  (b) On the other hand, the mean of the in-sample chosen arm is always *negatively* biased for the true best arm mean.

$$\begin{aligned}\text{Bias}_j^* &= E[\hat{\mu}_{T_j^*}(j)] - \left(\max_{t\in\mathcal{T}} \mu_t\right) \\ &= \underbrace{Bias_j^{**}}_{(a)} + \underbrace{\left(\mu_{T_j^*} - \left(\max_{t\in\mathcal{T}} \mu_t\right)\right)}_{(b)\ \leq 0}\end{aligned}$$



From: Shin, Ramdas, Rinaldo (2019)

Inference

There are many potential hypothesis of interest.

- *Fixed Arm*
  - $H_0: \ \mu_t = 0$
  - $H_0: \ \mu_t - \mu_0 = 0$

    $\vdots$
- *Best Arm*
  - $H_0: \ \max_{t \in \mathcal{T}} \mu_t = 0$
  - $H_0: \ \max_{t \in \mathcal{T}} \mu_t - \mu_0 = 0$
  - $H_0: \ \max_{t \in \mathcal{T}} \mu_t - \min_{t \in \mathcal{T}} \mu_t = 0$

    $\vdots$
- *Sharp Null*
  - $H_0: \ Y_i(1) = Y_i(0) \ \ \forall i$

Let's *explore* ... and then commit.

Non-Adaptive (e.g. Explore)

- Explore Phase: Alternate between $T = 0$ and $T = 1$
  $\Rightarrow$ m i.i.d. observations on $Y(1)$ and $Y(0)$
  $\Rightarrow$ $\hat{\mu}_0(2m)$ and $\hat{\mu}_1(2m)$
- For m large,

$$\hat{\mu}_0(2m) \approx N\left(\mu_0, \frac{\sigma_0^2}{m}\right)$$
$$\hat{\mu}_1(2m) \approx N\left(\mu_1, \frac{\sigma_1^2}{m}\right)$$

- Standard inference applies based on these averages for fixed arm hypothesis testing: t-tests, two sample t-test, or equivalent regression formulations:

$$Y_j = \alpha_0 \mathbf{1}\{T_j = 0\} + \alpha_1 \mathbf{1}\{T_j = 1\} + \varepsilon_j$$

- Best Arm:

$$\hat{\mu}^{max} = \max\{\hat{\mu}_0, \hat{\mu}_1\} \qquad \text{(suppressing “}(2m)''\text{)}$$

$$\mu^{max} = \max\{\mu_0, \mu_1\}$$

- Well behaved estimators for μ_0 and μ_1.
  Suppose $\mu^{max} = \mu_0 \geq \mu_1$.

$$\hat{\mu}^{max} - \mu^{max} = \max\{\hat{\mu}_0 - \mu_0, (\hat{\mu}_1 - \mu_1) - (\mu_0 - \mu_1)\}$$

$$\approx \max\left\{N\left(0, \frac{\sigma_0^2}{m}\right), N\left(-(\mu_0 - \mu_1), \frac{\sigma_1^2}{m}\right)\right\}$$

(e.g. $\mu_0 = \mu_1$ vs $\mu_0 >> \mu_1$)

- Best arm mean is max$\{\mu_0, \mu_1\}$.
  The max function is:
  - continuous
  - not differentiable (at $\mu_0 = \mu_1$)
  - directionally differentiable
- Non-differentiability creates challenges for estimation/inference
  - Hirano and Porter (2012)
  - Fang and Santos (2019)

Adaptive (e.g. Then-Commit)

Now standard approaches to fixed arm hypotheses have challenges too.

Commit/Adapt: $$\hat{\mu}_j(n) = \frac{\sum_{j:t_j=t} Y_j}{\mathbf{N_t(n)}}$$

Under adaptation, the average outcome for arm j is a ratio of random variables, where the denominator corresponds to $(X'X)^{-1}$ in OLS. A key "stability" condition for valid normal approximation is:

$$a_n \frac{N_t(n)}{n} \xrightarrow{p} 1 \qquad \text{for some sequence of scalars } a_n$$

Important case where this condition for approximate normality can fail for common bandit algorithms:

$H_0: \; ATE = 0$

e.g. $Y(0), Y(1) \;\sim\; N(0,1)$
Can show posterior for Thompson sampling does not concentrate

$$P(\mu_1 > \mu_0 \mid T^{(n-1)}, Y^{(n-1)}) \;\xrightarrow{d}\; Unif[0,1]$$

Implications:

- Standard inference like t-test using OLS coefficients not valid
- Lack of uniformity in large sample approximations: confidence intervals not reliable (can severely undercover)

Stability

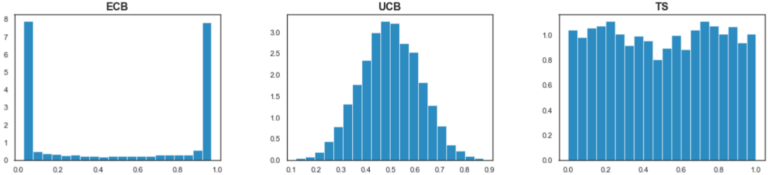


Figure 2: Histograms of the distribution of $N_1(n)/n$, the fraction of times arm 1 is picked under ε-greedy, UCB and Thompson sampling. The bandit problem has $p = 2$ arms which have i.i.d. $\mathsf{Unif}([-0.7, 1.3])$ rewards and a time horizon of $n = 1000$. The distribution is plotted over 4000 Monte Carlo iterations.

From: Deshpande, Mackey, Syrgkansis, Taddy (2020)

Problem identified: lack of approximate normality invalidates usual OLS inference

Fixes:

- Modify standard approaches to obtain approximate normality
  - **Batching** (Zhang, Janson, Murphy 2020)
  - Adaptive Re-weighting (Hadad, Hirshberg, Zhan, Wager, Athey 2019)
- Work with the non-Gaussianity
  - Confidence Sequences (Deshpande, Mackey, Syrgkansis, Taddy 2020; Howard, Ramdas, McAuliffe, Sekhon 2021)

Batching

In many settings, bandit algorithms are run in "batches" where each batch contains many units and data is collected sequentially by batch. The assignment mechanism for batch b can depend on the data from previous batches $(1, \ldots, b-1)$. The batch b mechanism is then applied to all units j within the batch.

Let $(T_{j,b}, Y_{j,b})$ denote the arm and outcome for unit j in batch b, where $j = 1, \ldots, n_b$ and $b = 1, \ldots, B$. And, for simplicity, take $n_b = n$ for all b. The approximations below will rely on n large, but the number of batches B can be small.

Batch Assignment Mechanism:

$$\begin{aligned} p_{b,n} &= \Pr(T_{j,b} = 1 \mid T^{(b-1)}, Y^{(B)}(0), Y^{(B)}(1)) \\ &= \Pr(T_{j,b} = 1 \mid T^{(b-1)}, Y^{(b-1)}) \end{aligned}$$

where the superscript $^{(b)}$ is used to denote the history of observations through batch b. And,

$$T_{1,b}, \dots, T_{n,b} \mid T^{(b-1)}, Y^{(b-1)} \overset{iid}{\sim} Bernoulli(p_{b,n})$$

Now run OLS batch by batch. For batch b,

$$Y_{j,b} = \gamma_b + T_{j,b}\beta_b + \varepsilon_{j,b}$$

For large n, the OLS coefficient $\hat{\beta}_b$ has approximate normality (even allowing for failure of stability). Moreover, OLS estimates from different batches are approximately independent.

Batch Inference

- Individual batch hypothesis.
  Now batch b, t-test, $t_b = \dfrac{\hat{\beta}_b}{std\ err(\hat{\beta}_b)}$ can be used to test for zero ATE in batch b, $H_0 : \beta_b = 0$.

- Combined batch hypothesis.
  Suppose batches are stationary, $\beta_1 = \cdots = \beta_B (= \beta)$
  Let
  $$t = \frac{1}{\sqrt{B}} \sum_{b=1}^{B} t_b$$
  This combined t-statistic can be used to test $H_0 : \beta = 0$

Additional Batch Wrinkles

- Multiple Arms
- Contextual Bandits
- Non-stationarity
- Connections to re-weighting approaches

Fisher's Exact Test

Sharp Null Hypothesis:

$$H_0: \; Y_i(0) = Y_i(1) \;\; \forall i$$

Potential Outcomes $+$ Assignment Mechanism $\rightarrow$ Data

Under H_0, $(Y_i(0), Y_i(1)) = (Y_i, Y_i), \forall i$. Assuming the researcher knows the assignment mechanism, she knows how the data is generated if the null is true. So the researcher can simulate the null distribution of any statistic.

Consider the OLS coefficient, $\hat{\beta}$ from the regression $Y_i = \gamma + T_i\beta + \varepsilon_i$. We could use $|\hat{\beta}|$ as the test statistic.

Round j

(a) Start with history - arms/treatments, outcomes, (and possibly covariates including round j covariates)

(b) **Based on history, form arm/treatment probabilities**

(c) **Based on these probabilities, randomly pick arm/treatment,** T_j

(d) Based on arm/treatment, realize corresponding outcome, Y_j

(e) Check stopping rule: stop or advance to the next round

(b)-(c) are the source of randomness to generate the null distribution.
e.g. Thompson Sampling, posterior $P(\mu_1 > \mu_0 \mid T^{(n-1)}, Y^{(n-1)})$ determines probability of assignment to arm 1 (and arm 0).

Keeping outcomes fixed at observed values, simulate a new Thompson sampling sample. Compute $|\hat{\beta}|$ for this sample and denote it by $|\hat{\beta}^{1*}|$. Repeat to obtain $|\hat{\beta}^{*1}|, \ldots, |\hat{\beta}^{*S}|$ for large S. Find c^* such that 5% of simulated values $|\hat{\beta}^{*s}|$ are greater.

Computed c^* is size 5% critical value for the test.
$|\hat{\beta}| \geq c^* \;\Rightarrow\;$ Reject H_0.

Power: can use the same simulation approach for any alternative values $(Y_i(0), Y_i(1))$ to obtain $|\hat{\beta}^{**s}|$ and find the fraction of simulated values larger than c^*.

Deterministic Bandits

- ETC, UCB: Treatment assignment is a deterministic function of history. Then steps (b) and (c) degenerate (probabilities 0 and 1). Without any randomness, all Fisher simulations return same treatment and outcomes as the original sample, which is not useful for testing.