Week 9: Covariates and Context

AIM-5014-1A: Experimental Optimization

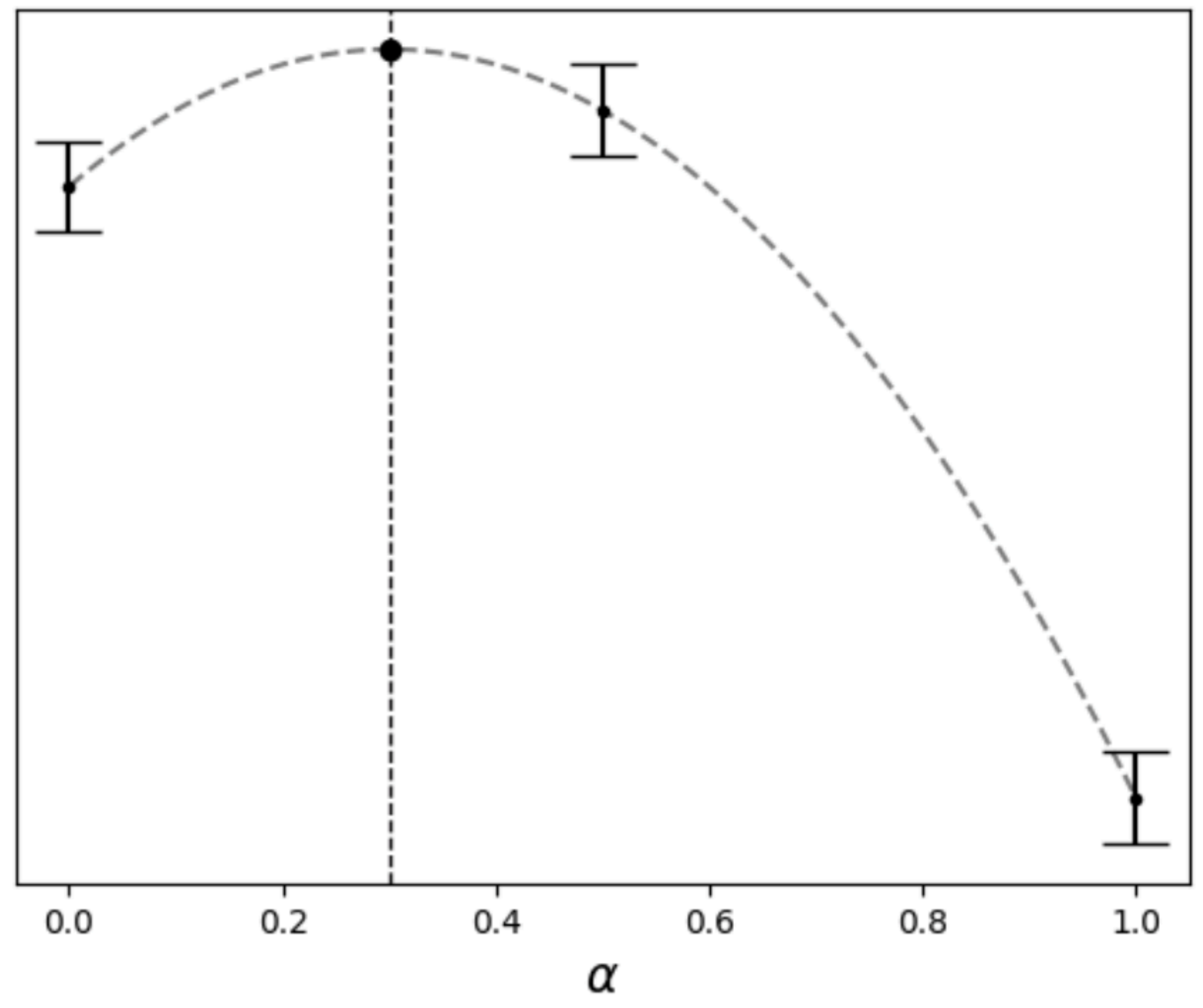
David Sweet // 20230727

Review: LLN, CLT, A/B Testing

- As $N \to \infty$, $\bar{y} \to E[BM]$ (LLN)
  - CLT: $\bar{y} \sim \mathcal{N}(E[BM], \sigma^2)$, “measured BM is gaussian”
- **Design**: $N \geq \left(\frac{2.5\hat{\sigma}_\delta}{PS}\right)^2$
- **Measure**: Randomize, $\bar{\delta} = \bar{y}_B - \bar{y}_A$, $se = \sigma_\delta/\sqrt{N}$
- **Analyze**: Accept B if $\bar{\delta} > PS$ and $\frac{\bar{\delta}}{se} \geq 1.64$ (check guardrails)
- **False Positive Traps**: Early stopping, multiple comparisons (use Bonferroni)

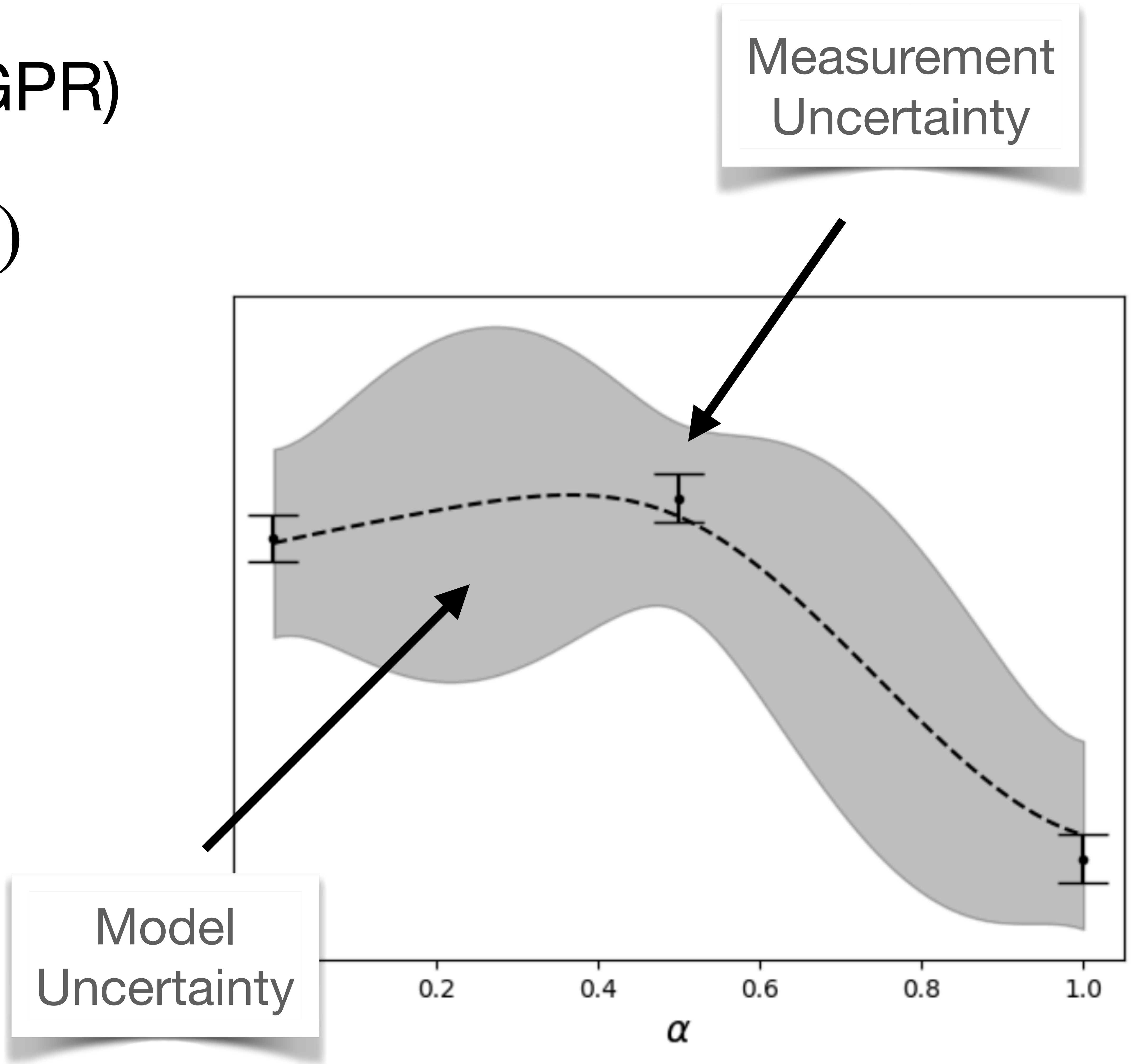
Review: Response Surface Methodology

- Parameters:
  - categorial: discrete unordered, strings; ex: A/B
  - ordinal: discrete ordered, integers; ex: 1, 2, 3, …
  - continuous: double; ex., [0,1]   <== RSM
- Surrogate, $y(x)$, models response surface, $E[y(x)]$
- Find optimum, $x^* = \arg\max_x y(x)$, and validate by A/B test



Review: Bayesian Optimization

- Surrogate: Gaussian Process Regression (GPR)
  - non-parametric, estimates both $\hat{y}(x)$, $\hat{\sigma}(x)$
- Acquisition function: $af(\hat{y}(x), \hat{\sigma}(x))$
  - determines next arms, $\{x_a\}$, to measure
  - balances exploration with exploitation



Case: Catalog Search

- Shopping site search results; ex., “toothpaste”
- Goal: More clicks earlier in list
    - Like this?: Crest, Colgate, Tom’s, Hello
    - Like this?: Tom’s, Hello, Crest, Colgate
- ML model determines ranking score; two versions: A & B
- Compare by A/B test

Case: Catalog Search

- Metric: Rank of item the user first clicks on
  - Negate, so metric gets maximized: -rank
- Ex: Crest, **Colgate**, Tom's, Hello;
  - User clicks Colgate, metric is $y_i = -2$
- Confounder: User's avg. purchasing rate (APR)
  - Users who buy more (less) often will do so with model A or B

Confounder

- BAD: Show A to high APR, B to low APR
  - Will measure $\bar{y}_a > \bar{y}_b$, but would just measure APR difference
- GOOD: Randomize

Aside: Linear Regression for A/B Testing

- You can analyze an A/B test with linear regression
- Collect all observations of metric, y_i
  - If observation from model B, then $\chi_{b,i} = 1$
- Linear model

$$y_i = \alpha + \beta x + \varepsilon_i$$

$$y_i = \bar{y}_a + \bar{\delta}\chi_{b,i} + \varepsilon_i$$

- Two parameters to fit $\bar{y}_a$, $\bar{\delta}$, just like α and β

Aside: Linear Regression for A/B Testing

- $y_i = \bar{y}_a + \bar{\delta}\chi_{b,i} + \varepsilon_i$
- Set $X = \begin{bmatrix} 1 & \chi_b \end{bmatrix}$, then write $y = X\beta + \varepsilon$ and regress:

$$\begin{bmatrix} \bar{y}_a \\ \bar{\delta} \end{bmatrix} = \beta = (X^\top X)^{-1}(X^\top y), \begin{bmatrix} se^2_{\bar{y}_a} \\ se^2_{\bar{\delta}} \end{bmatrix} = se^2_\beta = VAR(\varepsilon)(X^\top X)^{-1}$$

- IOW, regression gives you $\bar{\delta}$ and $t = \dfrac{\bar{\delta}}{se_{\bar{\delta}}}$

Same result as the usual way

Confounder

- GOOD: Randomize to break $corr(APR, \chi_b)$
  - But APR still correlated with $\bar{y}$
- BETTER: Randomize, & also model $corr(APR, \bar{y})$

$$y_i = \bar{y}_a + \bar{\delta}\chi_{b,i} + \beta_{APR}APR_i + \varepsilon_i$$

- Add APR as a regressor, regress to get $\bar{\delta}$ and $t = \dfrac{\bar{\delta}}{se_{\bar{\delta}}}$

Covariate Adjustment

- APR is called a *covariate*; adding to regression is called *covariate adjustment*
- Effect:
    - $\bar{\delta}$ has lower $se_{\bar{\delta}}$
    - NB: $E[\bar{\delta}]$ doesn't change; property of the system
    - Lower $se_{\bar{\delta}}$ ==> higher $t = \dfrac{\bar{\delta}}{se_{\bar{\delta}}}$
- Higher t, lower FP

Covariate Adjustment

- Could add any other covariates
- Time of day, age of user, user's avg. rate of purchase of specific product classes, etc.

$$y_i = \bar{y}_a + \bar{\delta}\chi_{b,i} + \sum_k \beta_k x_{k,i} + \varepsilon_i$$

- Add each user's historical avg.; called CUPED
  https://exp-platform.com/Documents/2013-02-CUPED-ImprovingSensitivityOfControlledExperiments.pdf

$$y_i = \bar{y}_a + \bar{\delta}\chi_{b,i} + \sum_k \beta_k x_{k,i} + \sum_u \beta_u \bar{y}_{hist,u_i} + \varepsilon_i$$

Covariate Adjustment

- Go nuts: Build an ML regression model of all features in your feature store:

$$y_i = ML(x_i) + \varepsilon_i$$

- Ex., ML == NN or GBM
- Then replace linear covariates

$$y_i = \bar{\delta}\chi_{b,i} + \beta_{ML}(ML(x_i) - \bar{y}_a) + \varepsilon_i$$

- Called MLRATE (ML Regression Average Treatment Effect)
  https://arxiv.org/abs/2106.07263

Randomization With Covariates

- Thompson Sampling reduces number of observations needed
- Recall: Model $\bar{y}_a \sim \mathcal{N}(E[y_a], \sigma_a^2), \bar{y}_b \sim \mathcal{N}(E[y_b], \sigma_b^2)$
- When each user arrives:
    - Draw one value each of $\bar{y}_a, \bar{y}_b$ from normal dists
    - Send user to model A if $\bar{y}_a > \bar{y}_b$, and vice-versa
- Covariates called *context* here

Contextual Bandit

Contextual Bandit

- Model $\bar{y}_a \sim \mathcal{N}(E[y_a | x], \sigma_a^2)$
- IOW:
- $$y_i = \bar{y}_a + \bar{\delta}\chi_{b,i} + \sum_k \beta_k x_{k,i} + \varepsilon_i$$

$\varepsilon_i \sim \mathcal{N}(0,\sigma^2)$

- Just set
  - $\chi_{b,i} = 0$ to query $\bar{y}_a$
  - $\chi_{b,i} = 1$ to query $\bar{y}_b$

Also need to update regression parameters sequentially

Contextual Bandit

- Higher precision w/covariates (context) ==>
  - Better decisions about which arm to use
- Could take it further w/interaction terms like $\chi_b x_k$
  - I.e., In some contexts, A is better
  - In some contexts, B is better
- Beyond scope of this lecture

Design for Covariate Adjustment

- Randomization breaks correlation between effect (χ_b) and confounder (x_k)
  - *On average*
  - As $N_{\text{experiments}} \rightarrow \infty, \; < corr(x_k, \chi_b) > \rightarrow 0$
- In any single experiment, $corr(x_k, \chi_b)$ is non-zero, $se[corr(x_k, \chi_b)] > 0$
- Try to make $se[corr(x_k, \chi_b)]$ smaller with good design

Design for Covariate Adjustment

- Regression models (“regresses out”) covariate (x_k) impact on metric (y)
  - *On average*
  - As $N_{\text{experiments}} \rightarrow \infty, \ < \beta_k > \rightarrow E[\beta_k]$
- In any single experiment, β_k has $se[\beta_k] > 0$
- Try to make $se[\beta_k]$ even smaller with good design

$\beta_k \propto corr(x_k, y)$

Design for Covariate Adjustment

- Lower $se_{\bar{\delta}}$ ==> higher $t = \dfrac{\bar{\delta}}{se_{\bar{\delta}}}$ is great
    - Lowering FP for fixed experimentation cost
- Could also capitalize on lower $se_{\bar{\delta}}$ by reducing N
    - You'd be reducing the experimentation code (for fixed FP)
- Lower $se[corr(x_k, \chi_b)]$ and $se[\beta_k]$ ==> lower $se_{\bar{\delta}}$

Design for Covariate Adjustment

- Analysis time: $y_i = \bar{y}_a + \bar{\delta}\chi_{b,i} + \sum_k \beta_k x_{k,i} + \varepsilon_i$
- Design time: Estimate $se_{\bar{\delta}}$ from this regression
- Don't *know* observations, $y_i, \chi_{b,i}\, x_{k,i}$
- Instead *plan* them:
  - Use PS to design layout of $\chi_{b,i}\, x_{k,i}$ and estimate se_{δ}
  - Analogous to $N = \left(\frac{2.5\sigma_\delta}{PS}\right)^2$

Design for Covariate Adjustment

- Ask: What would $se_{\bar{\delta}}$ be if I
  - Used N observations? (usual A/B test design $se = \sigma_\delta/\sqrt{N}$)
  - Included covariate x_k?
  - Used n_a observations of $\bar{y}_a$ and of n_b of $\bar{y}_b$?
  - Exposed $n_{a,k-high/low}$ observations to a high/low level of x_k, and similarly $n_{b,k-high/low}$
- And so on…

Design for Covariate Adjustment

- Can optimize the design to minimize $se_{\bar{\delta}}$
  - Which minimizes N
- You're seeking similar numbers of observations for
  - A and B
  - (A, x_k high), (A, x_k low), (B, x_k high), (B, x_k low)
  - etc.
- Keeps se's low for all parameters in regression

Could literally run an optimizer

Good "exploration" of both arms and covariate space

Design for Covariate Adjustment

- Neat trick: Pairing / matching
- Pair off each user with a very similar user
  - Similar by features: demographics, usage habits, etc.
- Expose (randomly) one user from each pair to A and the other to B
- Carefully balances exposure to covariates, reducing $se[corr(x_k, \chi_b)]$ and providing samples appropriate to reduce $se[\beta_k]$

Summary

- Reduce experimentation cost by accounting for covariates
- Design: Include covariates, minimize se_δ
- Measurement: Contextual Bandit
- Analysis: Covariate adjustment