Outline

• Offline Evaluation of Online Metrics
• Counterfactual Estimation
• Advanced Estimators
  1. Self-Normalized Estimator
  2. Doubly Robust Estimator
  3. Slates Estimator
• Case Studies & Demo
• Summary
\[ \widehat{U}_{\text{ips}}(\pi) = \frac{1}{n} \sum_i \frac{\pi(y_i|x_i)}{\pi_0(y_i|x_i)} \delta_i \]

\[ \hat{E}[\text{Constant}] \neq \text{Constant} \]

Two variables in IPS
- \( \delta \)
- \( \pi(y|x)/\pi_0(y|x) \)
Fix: Control Variates

Use correlated quantities to control $\pi(y|x)/\pi_0(y|x)$ variability

$$E[s_i] = \theta \text{ known}$$

**Multiplicative**

$$\frac{\theta}{E[s]} E \left[ \frac{\pi(y|x)}{\pi_0(y|x)} \delta \right]$$

e.g. Self-Normalization

**Additive**

$$E \left[ \frac{\pi(y|x)}{\pi_0(y|x)} \delta - s \right] + \theta$$

e.g. Doubly Robust Estimator
Self-Normalized Estimator

Use expected sample size as multiplicative control variate

\[
\hat{S}(\pi) = \frac{1}{n} \sum_i \frac{\pi(y_i|x_i)}{\pi_0(y_i|x_i)}
\]

\[
E[\hat{S}(\pi)] = 1
\]

\[
\hat{U}_{SNips}(\pi) = \frac{\sum_i \frac{\pi(y_i|x_i)}{\pi_0(y_i|x_i)} \delta_i}{\sum_i \frac{\pi(y_i|x_i)}{\pi_0(y_i|x_i)}}
\]

[Trotter & Tukey, 1952] [Hesterberg, 1983] [Swaminathan & Joachims, 2016]
Self-Normalization: Properties

\[ \hat{E}[\text{Constant}] = \text{Constant} \]

Equivariant

Asymptotically consistent

\[
\Pr \left( \lim_{n \to \infty} \hat{U}_{\text{SNips}}(\pi) = U(\pi) \right) = 1
\]

Small bias which decays \( O\left(\frac{1}{n}\right) \) while variance decays \( O\left(\frac{1}{\sqrt{n}}\right) \)
**News Recommender: Results**

snIPS often achieves a better bias-variance trade-off.
Doubly Robust Estimator

**Reward Prediction**

\[
\hat{U}_{\text{rp}}(\pi) = \frac{1}{n} \sum_i E_{y \sim \pi | x_i} [\hat{\delta}(x_i, y)]
\]

Low variance, High bias

**IPS**

\[
\hat{U}_{\text{ips}}(\pi) = \frac{1}{n} \sum_i \frac{\pi(y_i|x_i)}{\pi_0(y_i|x_i)} \delta_i
\]

High variance, No bias

\[
\hat{U}_{\text{dr}}(\pi) = \frac{1}{n} \sum_i \frac{\pi(y_i|x_i)}{\pi_0(y_i|x_i)} \left( \delta_i - \hat{\delta}(x_i, y_i) \right) + E_{y \sim \pi | x_i} [\hat{\delta}(x_i, y)]
\]

[Langford, Li, Dudik, 2011] [Jiang & Li, 2016]
Doubly Robust: Properties

Useful when using estimated propensities

\[ \hat{p}_i \approx \pi_0(y_i|x_i) \]

Unbiased if, either

\[ \hat{\delta}(x, y) = \delta(x, y) \]

Or,

\[ \hat{p}_i = \pi_0(y_i|x_i) \]

Default in Vowpal Wabbit

http://hunch.net/~vw/
DR dominates IPS even with a noisy $\hat{\delta}(x, y)$
Evaluating rankings (slates)

Exact match of composite actions in logs unlikely

Idea: Count per-slot matches
Slates Estimator

If $\pi_0$ samples $l$ documents from a multinomial $\mu(d|x)$, with replacement

$$\hat{U}_{\text{slates}}(\pi) = \frac{1}{n} \sum_i \left( 1 - l + \sum_{j=1}^{l} \frac{\mathbb{I}\{y_i[j] = \pi(x_i)[j]\}}{\mu(y_i[j]|x_i)} \right) \delta_i$$

For general $\pi_0$, need to record $\pi_0(y[j] = d, y[k] = d'|x)$
Define:

\[ \Gamma_{\pi_0(x)}[d, j ; d', k] = \pi_0(y[j] = d, y[k] = d' | x) \]
\[ 1_y[d, j] = \mathbb{I}\{y[j] = d\} \]

\[ \hat{U}_{\text{slates}}(\pi) = \frac{1}{n} \sum_i E_{y \sim \pi|x_i} \left[ 1^T_y \Gamma_{\pi_0(x_i)}^+ 1_y \delta_i \right] \]

Can also develop self-normalized/doubly robust variants

[Swaminathan et al, 2016]
Slates Estimator: Properties

- Typically, **exponentially better** sample complexity than IPS
- Unbiased if reward decomposes per-slot

\[ \exists \Phi_x \text{ s.t. } \delta(x, y := \boxed{\text{red}} \boxed{\text{blue}} \boxed{\text{green}}) = \Phi_x(\boxed{\text{red}} \boxed{\text{blue}} \boxed{\text{green}}) + \Phi_x(\boxed{\text{blue}} \boxed{\text{green}}) + \Phi_x(\boxed{\text{red}} \boxed{\text{green}}) + \Phi_x(\boxed{\text{red}} \boxed{\text{blue}} \boxed{\text{green}}) \ldots \]

Can capture higher-order interactions with suitable \( \Gamma_{\pi_0}(x) \)
(sn)Slates better than IPS but can have asymptotic bias
Outline

• Offline Evaluation of Online Metrics
• Counterfactual Estimation
• Advanced Estimators
• Case Studies & Demo
  1. Yahoo Front Page [Li, Chu, Langford, Wang; 2011]
  2. Bing Speller [Li, Chen, Kleban, Gupta; 2014]
  3. Search Ranking [Swaminathan et al; 2016]
• Summary
\( y \sim \pi(x) \) Pick \( \text{STORY} \in \{F1, \ldots, F20\} \) to highlight for different users

Metric \( \delta \): CTR

Logging \( \pi_0 \): \textbf{Uniform} random

Setup: Deploy policies, check online \( \leftrightarrow \) IPS correlation
IPS is quite accurate for several (spatio-temporal) policies
IPS indeed gives unbiased CTR estimates for different articles
Pick (possibly many) reformulation candidates

Logging $\pi_0$: **Independent Bernoulli** per rank

$$\Pr(\text{Pick } d_j) = \frac{1}{1 + \alpha \exp[\beta(\text{score}(d_j) - \text{score}(d_1))]}$$

Setup: Deploy policies, check online $\leftrightarrow$ IPS correlation
IPS reliably estimates CTR, etc. despite non-uniform logging
Re-rank 5 out of 8 candidates

Metric $\delta$: Time-to-success
Utility rate

Logging $\pi_0$: Bootstrap from
Uniform/Plackett-Luce

Setup: Report RMSE vs. bootstrap sample size
Plackett-Luce for Slates

SoftMax/Multinomial without replacement

(contextual bandit 0.40
container store 0.20
contrave 0.20
continental airlines 0.06
continuous delivery 0.06
continental tires 0.06
content management system 0.06
container homes 0.02)

(renormalize probabilities after each draw)
Search Ranking: Case Study

Slates estimator dominates IPS
Outline

• Offline Evaluation of Online Metrics
• Counterfactual Estimation
• Advanced Estimators
• Case Studies & Demo
• Summary
**Evaluation: Key Questions**

What system $\pi_0$ should we deploy?

What should we record in our logs?

How can we estimate $U(\pi)$ using logged data from $\pi_0$?
What system $\pi_0$ to deploy?

To enable reliable counterfactual estimation
  If possible, stochastic with logged randomization
  If not, log enough to estimate propensities $\hat{p}_i$

How to explore?

Uniform? typically, bad
“Around current system”? “less risk” & “better targeted”
...
What should we record?

Log EVERYTHING!

To reliably “replay” $\pi$ on logged data,

- Candidate set of actions $\{Y_i\}$
- Features for each candidate $\{f(x_i, y)\}$
- Action at the point of randomization $y_i$

$\langle x_i, y_i, \delta_i, p_i \rangle$

[Bottou et al, 2013] [Li, Chen, Kleban, Gupta, 2015]
How can we evaluate $U(\pi)$?

- Offline Evaluation of Online Metrics
  - Related: Test collections for offline metrics
    [Carterette et al, 2010] [Aslam et al, 2009] [Schnabel et al, 2016b]…

- “Model the world”
  - Related: Click models; Collaborative filtering
    [Chuklin et al, 2015] [Schnabel et al, 2016a]

- “Model the bias in data”
  - Randomization is essential

Reward Prediction

Off-policy estimator
Summary: Off-policy Estimators

- IPS Estimator
  Simple, effective fix for non-uniform (biased) data

- Self-Normalized Estimator

- Doubly Robust Estimator

- Slates Estimator
Demo: Code Samples


Download Code_Data.zip

(Recommend) Install Anaconda-Python3; joblib

- Run experiment:
  ```
  python EvalExperiment.py
  ```

- Plot results:
  ```
  python plot_sigir.py --mode [estimate/error] --path ../../Logs/ssynth...
  ```
QUESTIONS?