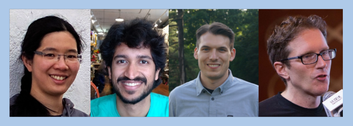


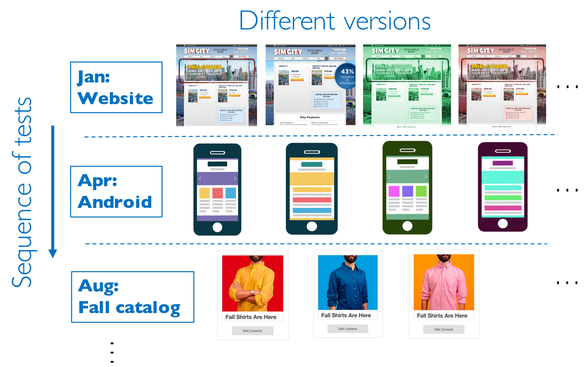
A framework for Multi-(A)rmed/(B)andit Testing with online FDR control

Fanny Yang[†], Aaditya Ramdas[⋆,†], Kevin Jamieson[◦], Martin J. Wainwright[†,⋆]

Dept. of EECS[†], Dept. of Statistics[⋆], UC Berkeley; Allen School of C.S.E.[◦], U. of Washington

What we solve

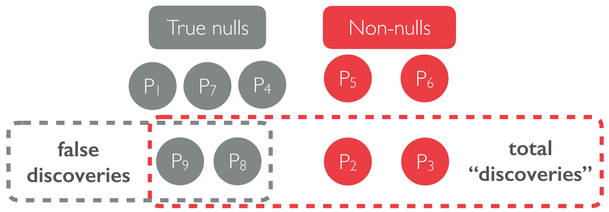
Problem setting:
Multiple experiments (*tests*) are conducted sequentially with goal to detect whether one of new versions (*arms*) is better than default by repeatedly getting feedback (*samples*)



Our three high-level goals:

- *"False alarms"*: Not make too many wrong detections across all tests
  → Control **false discovery rate** (FDR) **at every** test J (online)

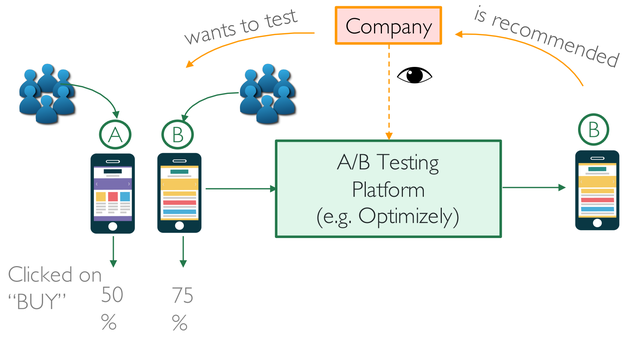
$$\mathrm{FDR}(J) = \mathbb{E}\frac{\#\text{ false discoveries}}{\#\text{ total discoveries}}$$



- *"Power"*: If default is not best, **find best arm**, with at least confidence $1-\delta$ **using few samples**
- *p-value peeking*: Allow for **continuous monitoring** of p-value

Status quo: A/B testing

One test, two arms: A user is assigned to each arm with prob. 0.5 (uniform sampling); adjusted p-value allows continuous monitoring



Hypothesis testing model where A is default (*control*):

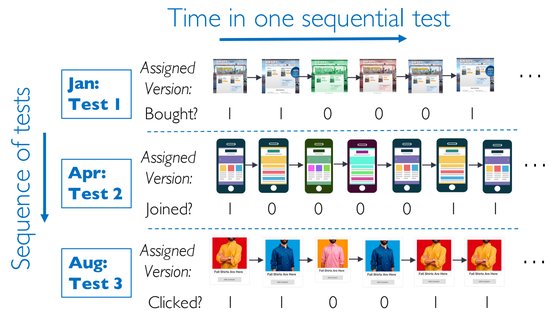
- Samples drawn from distributions $\mathbb{P}_A, \mathbb{P}_B$ with means μ_A, μ_B
- Null hypothesis is $H_0 : \mu_A > \mu_B$, alternative: $H_1 : \mu_A < \mu_B$
- Compute p-value, i.e. how extreme a test statistic T is
- Null is rejected, $R = 1$, if $p < \alpha$ (significance level)

Major shortcomings of conventional A/B testing

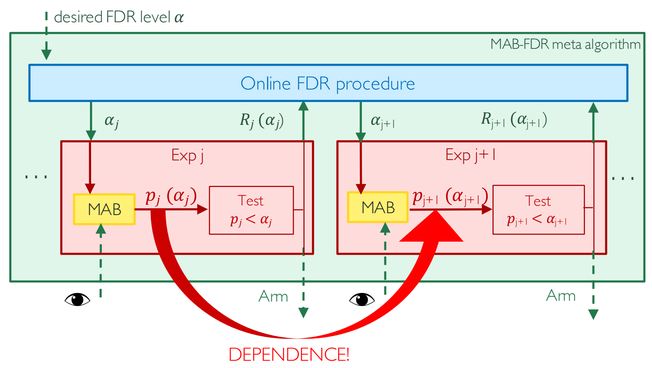
- For K arms, uniform sampling requires $O(K)$ number of samples
- Does not incorporate correction for multiple tests over time

Our approach: doubly-sequential framework

- Use best-arm MAB algorithm for one test → doubly-sequential



- Simultaneously control FDR and sample complexity by interaction between best-arm MAB and FDR



- How is α_j used in best-arm MAB to have low sample complexity?
- How do we obtain "always valid p-values" from MAB exp.?
- p-values are now dependent, FDR still controlled?

Contrib. I: MAB-FDR procedure for multiple tests

Algorithm 1 MAB-FDR

for $j = 1, 2, \ldots$ **do**
  Test j receives one control and $K(j)$ alternatives
  $\alpha_j = \textsc{OnlineFDR}(\{P^\ell\}_{\ell=1}^{j-1})$
  $(i_b, P_T^j) = \textsc{Best-arm MAB}_j(\alpha_j)$
  if i_b is *not* the control and $P_T^j \le \alpha_j$ **then**
    reject the null hypothesis of test j; return i_b to user
  Return P_T^j to $\textsc{OnlineFDR}$

with T either MAB stopping time or user-defined truncation time.

Theorem:

- $FDR(J) \le \alpha$ and $BDR(J) \ge \frac{\sum_{j\in\mathcal{H}_1(J)}(1-\alpha_j)}{|\mathcal{H}_1(J)|}$
- For each experiment i, MAB only draws as many samples as are needed to guarantee online FDR control at level α and power.

Contrib. II: Testing vs. MAB perspective for one test

Setting: K arms, $i_\star$ actual best arm, i_b bandit arm

- What's the null hypothesis? For FDR ("false alarms") control, we only care if control is actually best, i.e.

$$H_0 : \mu_0 > \mu_i \ \ \forall i = 1, \ldots, K \quad \textit{vs.} \quad H_1 : \exists i \ \text{ s.t. } \mu_0 < \mu_i$$

- What about power? Best-arm MAB goal: find best arm

$$BDR = \frac{\mathbb{E}\sum_{j\in\mathcal{H}_1} R_j \mathbb{I}_{\mu_{i_b}=\mu_{i_\star}}}{|\mathcal{H}_1(J)|}$$

Insight: Best-arm MAB is capable of both testing for null hypothesis adaptively **and** finding best arm

Key quantity: Always valid p-values using best-arm MAB

- Main tool: Non-asymptotic Law of Iterated Logarithm (LIL)
  → always valid confidence interval:

$$\mathbb{P}(\exists t : \mu_{i,t} \notin [\mathrm{LCB}_i(t,\delta), \mathrm{UCB}_i(t,\delta)]) \le \delta$$

  with $\mathrm{UCB}_i(t,\delta), \mathrm{LCB}_i(t,\delta) = \widehat{\mu}_i \pm c\sqrt{\frac{\log(\frac{1}{\delta})+\log\log(n)}{n}}$
- Compute p-value using duality w/ confidence intervals & LIL:

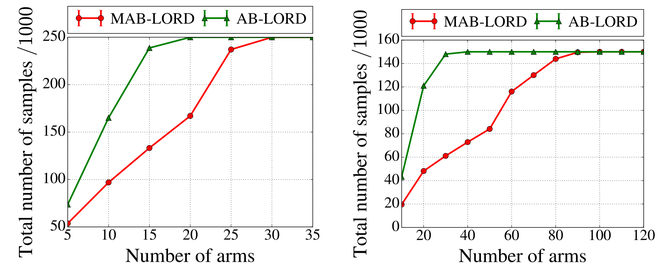
$$P_{i,t} : \sup\left\{\gamma \in [0,1] \ \middle|\ \mathrm{LCB}_i(n_i(t),\gamma) \le \mathrm{UCB}_0(n_0(t),\gamma) + \epsilon\right\}$$

  p-value at each time-step t (within one MAB exp.) is defined as $P_t := \min_{s\le t} \min_{i=1,\ldots,K} P_{i,s}$.

Proposition: p_t is an always valid p-value!

Simulations

Sample complexity vs. number of arms per experiment for: Bernoulli (left, 50 hyp.) and Gaussian (right, 500 hyp.) draws



Background: onlineFDR procedure

Key idea: The more rejections up until j, the higher the *wealth* and thus available significance level α_j for next test j