A framework for Multi-(A)rm(ed)/(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)

<table>
<thead>
<tr>
<th>Our three high-level goals:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Control false discovery rate (FDR) at every test J (online)</td>
</tr>
<tr>
<td>2. Simultaneously control FDR and sample complexity by interaction between best-arm MAB and FDR</td>
</tr>
<tr>
<td>3. How αj is used in best-arm MAB to have low sample complexity?</td>
</tr>
<tr>
<td>4. How do we obtain “always valid p-values” from MAB exp.?</td>
</tr>
<tr>
<td>5. p-values are now dependent, FDR still controlled?</td>
</tr>
</tbody>
</table>

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

Algorithm 1 MAB-FDR
for j = 1, 2, ... do
  Test j receives one control and K(j) alternatives
  $\alpha_j = \text{OnlineFDR}(\{P_i^j\}_{i=1}^{K})$
  $(b_j, P_j^f) = \text{Best-arm MAB}(\alpha_j)$
  if $b_j$ is not the control and $P_j^f \leq \alpha_j$ then
    reject the null hypothesis of test j; return $b_j$ to user
  Return $P_j^f$ to OnlineFDR
with T either MAB stopping time or user-defined truncation time.

Theorem:
- $\text{FDR}(J) \leq \alpha$ and $\text{BDR}(J) \geq \sum \frac{\alpha_{\text{null}}(1-\alpha_j)}{\text{E}[T](J)}$
- For each experiment j, MAB only draws as many samples as are needed to guarantee online FDR control at level α and power.

Contrib. I: MAB-FDR procedure for multiple tests
Setting: K arms, i, actual best arm, $b_i$ bandit arm
- What’s the null hypothesis? For FDR (“false alarms”) control, we only care if control is actually best, i.e. $H_0 : \mu_0 > \mu_i \forall i$. $H_1 : \exists i \text{ s.t. } \mu_0 < \mu_i$
- What about power? Best-arm MAB goal: find best arm $BDR = \frac{\sum \mathbb{E}[R_i^f]}{\text{E}[T](J)}$.

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:
  $\text{FDR control level } \alpha$

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 $\alpha_j$ for next test j