Previous work: Context

- 'Best Arm Identification' / 'Pure Exploration Bandits'
  Simple Bayesian Algorithms for Best-Arm Identification, Daniel Russo
- $\theta = [.2, .3, \ldots, .5, .1]$ (Bernoulli)
- Best arm: $argmax_i\ \theta_i$
- Possible objectives:
  - Fixed #samples: high confidence
  - Fixed confidence: few #samples
- Conventional approaches:
  - Uniform sampling: Equal effort/measurements, not equal evidence.
  - Thompson sampling: Focus on estimated best, little evidence about alternatives.
- Idea: Gather equal evidence by focusing on the 'threshold'.

Previous work: Algorithm: Top-two Thompson sampling

Given hyperparameter β, prior Π_n at step n:

$B \sim Bernoulli(\beta)$
$\hat{\theta} \sim \Pi_n$
$I = argmax_i \; \hat{\theta}$
if $B = 0$ **then**
  repeat
    $\hat{\theta} \sim \Pi_n$
    $J = argmax_i \; \hat{\theta}$
  until $J \neq I$
  $I = J$
end if

Play I, observe reward and update priors.

Previous work: Theory

- Analysis of the convergence's exponent, i.e. $\lim_{n\to\infty} \frac{1}{n} \log \Pi_n(\Theta^C_{I*}) = \Gamma$
- $\psi_{n,I*} = \Pr[I_n = I^*]$
- Under all adaptive allocation rules that have $\lim_{n\to\infty} \psi_{n,I^*} = \beta$, TTTS with parameter β has best exponent Γ^*_β.
  (Constrained optimality without tuning)
- TTTS can be tuned to $\beta^* = argmax_\beta \Gamma^*_\beta$ and hence have the best overall possible exponent.
  (Optimality with tuning)
- $\Gamma^* \leq 2\Gamma^*_{\frac{1}{2}}$, $\frac{\Gamma^*}{\Gamma^*_\beta} \leq max\{\frac{\beta^*}{\beta}, \frac{1-\beta^*}{1-\beta}\}$
  (Robustness)

Project motivation

- Best arm identification in practice often comes in different, more general shapes.
  - **Top-m arms** (set)
  - Batch updates
  - Contextual arms
  - Correlated arms

Project goal

- Define top-m arms algorithm.
- Determine and prove best possible convergence rate for top-m, Γ^*.
- Determine and prove best possible convergence rate for 'constrained' class of algorithms, Γ^*_γ.
- Prove that TTTS belongs to one such class.
- Prove that TTTS is optimal in said class.
- Prove 'robustness', i.e. relationship between constrained and non-constrained optimum.

Project: So far

- Algorithm definition (not definite).
- Attempt to understand both Russo's proofs and the interaction between them.
    - Bottom-up
    - Top-down
        - Breadth-first
        - Depth-first
- Simulation.

Current algorithm

Given a prior Π_n at step n:

$\hat{\theta} \sim \Pi_n$
$I_m = \text{top-m}(\theta)$
repeat
  $\hat{\theta} \sim \Pi_n$
  $J_m = \text{top-m}(\hat{\theta})$
until $I_m \neq J_m$
$I \sim \mathcal{U}(I_m \oplus J_m)$

Play I, observe reward and update priors.

Simulation: Setting

- Uniform Beta priors.
- Confidence in option $I = [1, 3, 4]$ at step n:

$$\Theta_I = \{\theta \in \Theta, \text{top-m}(\theta) = I\} \tag{1}$$

$$\alpha_{I,n} = \int_{\Theta_I} \pi_n(\theta) d\theta \tag{2}$$

- Approximate via Monte Carlo method, 10^7 samples $\mathcal{U}([0, 1]^{\#arms})$
- For fixed #samples, compute confidence $(\alpha_{I,n})$.
- Average over 5-40 seeds.
- $\theta = [.1, .2, .3, .4, .5]$
- m $= 3$.

Simulation: Observations

- $\sim$ 'goes in right direction'
- Hypothesis: Edge of TTTS over uniform distribution should be greater when $\#arms$ and m are 'further apart'.