Simple Bayesian Algorithms for Best-Arm Identification

Guiyu Hong

The Chinese University of Hong Kong, Shenzhen

GuiyuHong@link.cuhk.edu.cn

August 5, 2021

Overview

1. First Section
   - Settings
   - Algorithms
2. Numerical Experiments
3. Theoretical Results

Introduction

Multi-Arm bandits has been researched for a long time. Typically, people use *Regret* as performance measure. However, here are still many cases that are not sensitive to regret. For example

1. A/B tests
2. Simulation Optimization (tuning)
3. Design of Clinical Trials

In these cases, we want to identify the best arm within shorter trails. Maybe the best algorithms for regret is not the best for identification.

Problem Formulation

We consider the problem in *frequentist* setting, but the algorithm that we consider is *Bayesian* algorithm.
Suppose here is k arm with mean $(\theta_1^*, \cdots, \theta_k^*)$. At each time $n \in \mathbb{N}$, one choose design $I_n \in \{1, 2, \cdots, k\}$ and observe Y_{n,I_n} as response.
$Y_n \triangleq (Y_{n,1}, \cdots, Y_{n,k})$ is independently across time. We focus on one dimensional exponential response, i.e.

$$p(y|\theta) = b(y) \exp(\theta T(y) - A(\theta)). \tag{1}$$

For convenience, $T(\cdot)$ is strictly increasing $\Rightarrow E[Y|\theta]$ is increasing of θ.
Let $I^* = \arg\max_i \theta_i^*$ and suppose $\theta_i \neq \theta_j \quad \forall\ i \neq j$.

Problem Formulation

Let Π_1 be the prior distribution on parameter region Θ ($\theta^* \in \Theta$). Based on observation sequence $(I_1, Y_{1,I_1}, \cdots, I_{n-1}, Y_{n-1,I_{n-1}})$, we have posterior measure Pi_n with density

$$\pi_n(\theta) = \frac{\pi(\theta)L_{n-1}(\theta)}{\int_\Theta \pi(\theta)L_{n-1}(\theta)d\theta}, \quad n \geq 2, \tag{2}$$

where

$$L_{n-1}(\theta) = \prod_{l=1}^{n-1} p(Y_{l,I_l}|\theta_{I_l})$$

is the likelihood.

Some Notations

To describe the algorithm and related results, we need following additional notations.

1. Advantage region $\Theta_i \triangleq \left\{\theta \in \Theta \middle| \theta_i > \max_{j \neq i} \theta_j \right\}$.
2. Posterior Probability of i-th arm $\alpha_{n,i} \triangleq \Pi_n(\Theta_i) = \int_{\Theta_i} \pi_n(\theta) d\theta$.
3. Assigned Probability $\psi_{n,i} \triangleq \mathbb{P}(I_n = i | \mathcal{F}_{n-1})$
4. Accumulated Effort $\Psi_{n,i} \triangleq \sum_{l=1}^{n} \psi_{n,i}$
5. Average Effort $\bar{\psi}_{n,i} = n^{-1} \Psi_{n,i}$

Top-Two Probability Sampling

At each cycle, we do the following things:

1. Calculate α_i, let $\hat{I}^* = \arg\max_i \alpha_i$ and $\hat{J}^* = \arg\max_{j \neq \hat{I}^*} \alpha_j$
2. Toss a coin B $bin(p)$, p is a hyper-parameter
3. If $B = 1$ use $\hat{I}^*$ otherwise use $\hat{J}^*$
4. Update posterior distribution

Top-Two Value Sampling

Define utility function $u : \theta \rightarrowtail \mathbb{R}$ as continuous and strictly increasing function. Then we can define value function $v_i(\vec{\theta} = \max_j u(j) - \max_{j \neq i} u(j))$. Then define $V_{n,i} = \mathbb{E}^{\Pi_n}[v_i]$.

1. Calculate V_i, let $\hat{I}^* = \arg\max_i V_i$ and $\hat{J}^* = \arg\max_{j \neq \hat{I}^*} V_j$
2. Toss a coin B $bin(p)$, p is a hyper-parameter
3. If $B = 1$ use $\hat{I}^*$ otherwise use $\hat{J}^*$
4. Update posterior distribution

Top-Two Thompson Sampling

Likewise, we add additional sampling to TS, we get

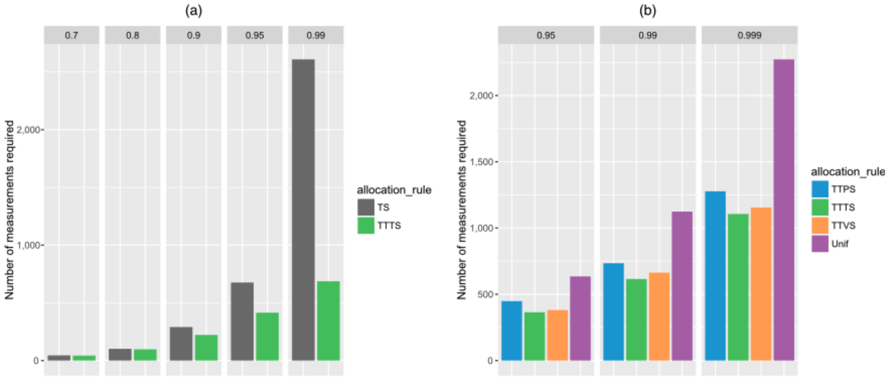
1. Calculate α_i and sample $\hat{I}$ according to α_i
2. Toss a coin B $bin(p)$
3. If $B = 1$ use $\hat{I}$ else sample $\hat{J}$ until $\hat{J} \neq \hat{I}$
4. Update posterior distribution

Numerical Illustration

Set $\theta^* = (.1, .2, .3, .4, .5)$ and $Y_{n,i}$ follows a binary distribution. We set hyper parameter $p = 0.5$ and observe how many times we need when confidence interval of optimal arm superseding a threshold, i.e. $\max_i \alpha_{n,i} > c$.

We compare Top-Two methods with TS methods and uniformly testing methods.

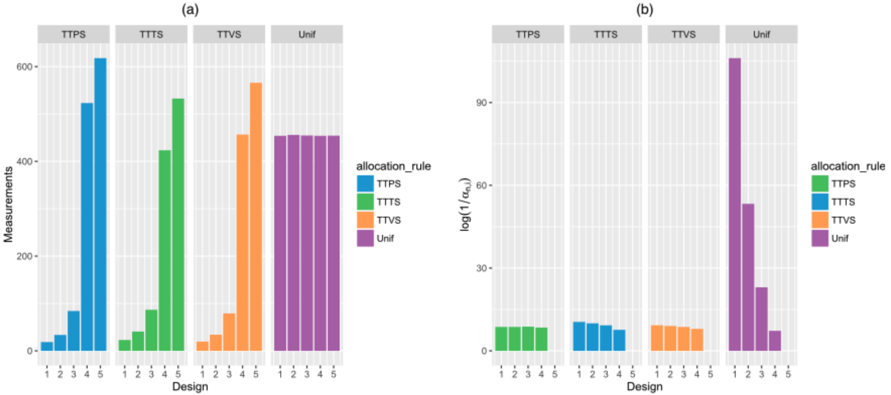
Comparing with TS



Notes. (a) TS vs. top-two TS. (b) Comparison with uniform allocation.

Allocations

Figure 2. (Color online) Distribution of Measurements and Posterior Beliefs at Termination



Notes. (a) Measurements collected of each design. (b) Value of $\log(1/\alpha_{n,i})$ for each design i.

Main Theorems

We focus on convergence rate of $\Pi_n(\Theta^c_{I^*}) = \sum_{i \neq I^*} \alpha_{n,i}$, i.e. the convergence rate of false probability.
We assume that $\Theta = (\underline{\theta}, \bar{\theta})^k$, i.e. a bounded rectangle and $0 < \inf_{\theta \in \Theta} \pi_1(\theta) < \sup_{\theta \in \Theta} \pi_1(\theta) < \infty$ (regular prior). Moreover, we assume $\sup_\theta |A'(\theta)| < \infty$.
Then, let we define following rates:

$$\Gamma^* = \max_{\psi} \min_{\theta \in \Theta^c_{I^*}} \sum_{i=1}^{k} \psi_i d(\theta^*_i \| \theta_i), \tag{3}$$

and

$$\Gamma^*_\beta = \max_{\psi : \psi_{I^*} = \beta} \min_{\theta \in \Theta^c_{I^*}} \sum_{i=1}^{k} \psi_i d(\theta^*_i \| \theta_i). \tag{4}$$

Main Theorem

Theorem

There exist constants $\{\Gamma^*_\beta > 0 : \beta \in (0,1)\}$ *such that* $\Gamma^* = \max_\beta \Gamma^*_\beta$ *exists,* $\beta^* = \arg\max_\beta \Gamma^*_\beta$ *is unique and the following properties satisfies with probability 1:*

1. *Under Top-Two algorithms with parameter* β^**,*

$$\lim_{n\to\infty} -\frac{1}{n}\log \Pi_n(\Theta^c_{I^*}).$$

Under any adaptive allocation rule,

$$\lim\sup_{n\to\infty} -\frac{1}{n}\log \Pi_n(\Theta^c_{I^*}) \le \Gamma^*.$$

Main Theorem

Theorem

1. *Under Top Two algorithms, with parameter* $\beta \in (0,1)$*,*

$$\lim_{n\to\infty} -\frac{1}{n}\log \Pi_n(\Theta^c_{I^*}) = \Gamma^c_\beta \quad \textit{and} \quad \lim_{n\to\infty} \bar{\psi}_{n,I^*} = \beta.$$

   Under any adaptive allocation rule,

$$\limsup_{n\to\infty} -\frac{1}{n}\log \Pi_n(\Theta^c_{I^*}) \le \Gamma^*_\beta,$$

   on any sample path with $\lim_{n\to\infty} \bar{\psi}_{n,I^*} = \beta$.

2. $\Gamma^* \le 2\Gamma^*_{\frac{1}{2}}$.

These theorems show that $\Pi(\Theta^c_{I^*}) = O(e^{-n\Gamma^*_\beta})$.

Intuition and Analysis on Main Theorem

Let we denote $a_n \doteq b_n$ if $\frac{1}{n}\log(\frac{a_n}{b_n}) \to 0$. For example, $a_n + b_n \doteq \max(a_n, b_n)$ and $ca_n \doteq a_n$, etc.
Then main theorem mainly shows that $\Pi_n(\Theta_{I^*}^c) \doteq e^{-n\Gamma_\beta^*}$ and cannot be faster than $e^{-n\Gamma^*}$.
Now we show the intuition behind this theorem by KL-divergence. Define $d(\theta\|\theta') = \int \log(\frac{p(y|\theta)}{p(y|\theta')})p(y|\theta)dv(y)$, and

$$D_\Psi(\theta, \theta') = \sum_{i=1}^{k} \Psi_i d(\theta_i, \theta_i'),$$

which measures the average information gain using sampler Ψ.

Intuition of Main Theorem

We have following Proposition

Theorem

For any open set $\tilde{\Theta} \in \Theta$,

$$\Pi_n(\tilde{\Theta}) \doteq \exp\left\{-n \inf_{\theta\in\tilde{\Theta}} D_{\bar{\psi}_n}(\theta^*\|\theta)\right\}.$$

Intuition:

$$\log(\frac{\pi_n(\theta)}{\pi_n(\theta^*)}) = \log(\frac{\pi_1(\theta)}{\pi_1(\theta^*)}) + \sum_{l=1}^{n-1}\log(\frac{p(Y_{l,i}|\theta)}{p(Y_{l,i}|\theta^*)}),$$

which is a random walk with drift $\mathbb{E}\left[\log(\frac{p(Y_{l,i}|\theta)}{p(Y_{l,i}|\theta^*)})\right]$ if the policy $\psi_{n,i}$ converges to some ψ, then the drift is close to $-D_\psi(\theta^*\|\theta)$.

About the fastest rate

Since we know $\Pi_n(\Theta_{I^*}^c) \doteq \exp\left\{-n \inf_{\theta \in \Theta_{I^*}^c} D_{\bar{\psi}_n}(\theta^* \| \theta)\right\}$, to find the fastest rate, we need to find

$$\max_{\psi} \min_{\theta \in \Theta_{I^*}^c} D_{\psi}(\theta^* \| \theta),$$

which is just Γ^*.
Similarly, Γ_β^* is also defined in this way intuitively.

References

Danial Russo (2020)
Simple Bayesian Algorithms for Best-Arm Identification
Operations Research 68(6), 1625 – 1647.

The End