Thompson's Sampling

Information Theory

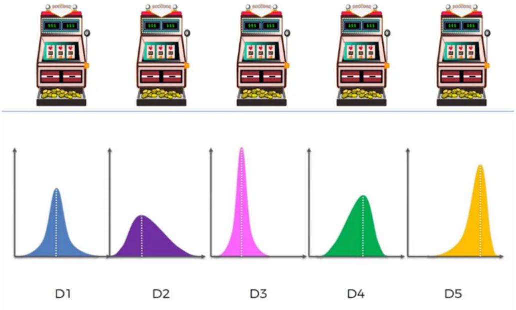
Juswanth T

May 14, 2024

Outline

- Introduction
- Simulation
- Regret Bound for 2 arms case

Introduction



Scenario: Pull machine k
$\rightarrow$ sample from **unknown** reward distribution $D_k \rightarrow$ observe reward.
Problem: Given a finite number of pulls T, how can I optimize my winnings?
How much should I explore? How much should I exploit?

A Bayesian Approach to Multi-Armed Bandits

Thompson sampling is a Bayesian approach to the multi-armed bandit problem. It addresses the exploration-exploitation dilemma by randomly selecting actions according to their estimated probabilities of being optimal.

Key Idea

TS operates within a Bayesian framework, maintaining a probability distribution (typically Beta) for each arm's potential reward. For each time step:

- **Prior Belief:** Start with a prior distribution over the unknown parameters of each arm (e.g., Beta distribution for Bernoulli rewards)..
- **Sampling:** At each time step, sample a set of parameters from the posterior distribution of each arm

Introduction

- **Selection:** The arm with the highest sampled reward is selected.
- **Observation:** The resulting reward is observed.
- **Update:** The posterior distribution of the chosen arm is updated based on the observed reward.

Microsoft's adPredictor for CTR prediction of search ads on Bing uses the idea of Thompson Sampling.

Simulation

Python Simulation -> https://colab.research.google.com/drive/1T7aD-UBWQk-pG_YvZbJG5RMOlwEuo4LN?usp=sharing

Difficultes in Bound calculations

- Thompson Sampling is a randomized algorithm which achieves exploration by choosing to play the arm with best sampled mean, among those generated from beta distributions around the respective empirical means.
- This randomized setting is unlike the algorithms in UCB family, which achieve exploration by adding a deterministic, non-negative bias inversely proportional to the number of plays, to the observed empirical means.

Result to prove :

$$\mathbb{E}[\mathcal{R}(T)] = \mathbb{E}\left[\Delta \cdot k_2(T)\right] \leq \left(\frac{40 \ln T}{\Delta} + \frac{48}{\Delta^3} + 18\Delta\right)$$

Regret Bound for 2 arms case

- j_0 denote the number of plays of the first arm until $L = 24(\ln T)/\Delta^2$ plays of the second arm.
- t_j denote the time step at which the j^{th} play of the first arm happens (we define $t_0 = 0$).
- $Y_j = t_{j+1} - t_j - 1$ measure the number of time steps between the j^{th} and $(j+1)^{th}$ plays of the first arm (not counting the steps in which the j^{th} and $(j+1)^{\text{th}}$ plays happened)
- $s(j)$ denote the number of successes in the first j plays of the first arm.

Then the expected number of plays of the second arm in time T is bounded by

$$\mathbb{E}\left[k_2(T)\right] \leq L + \mathbb{E}\left[\sum_{j=j_0}^{T-1} Y_j\right]$$

Regret Bound for 2 arms case

- Define $X(j, s, y)$ to be the number of trials before the sample taken from the Beta(s + 1,j - s + 1) distribution exceeds y. Thus, $X(j, s, y)$ takes non-negative integer values, and is a geometric random variable with parameter (success probability) $1 - F^{beta}_{s+1,j-s+1}(y)$.
- Here $F^{\text{beta}}_{\alpha,\beta}$ denotes the cdf of the beta distribution with parameters α, β.
- Also, let $F^{B}_{n,p}$ denote the cdf of the binomial distribution with parameters (n, p).
- By the well-known formula for the expectation of a geometric random variable and the definition of X we have, $\mathbb{E}[X(j, s, y)] = \frac{1}{1-F^{b+ta}_{s+1,j-s+1}(y)} - 1$

Regret Bound for 2 arms case

- Fact : $F^{\text{beta}}_{\alpha,\beta}(y) = 1 - F^{B}_{\alpha+\beta-1,y}(\alpha-1)$

  One well-known way to generate a r.v. with cdf $F^{\text{beta}}_{\alpha,\beta}$ for integer α and β is the following: generate uniform in $[0,1]$ r.v.s $X_1, X_2, ..., X_{\alpha+\beta-1}$ independently. Let the values of these r.v. in sorted increasing order be denoted $X^*_1, X^*_2, ..., X^*_{\alpha+\beta-1}$. Then X^*_α has cdf $F^{\text{beta}}_{\alpha,\beta}$. Thus $F^{\text{beta}}_{\alpha,\beta}(y)$ is the probability that $X^*_\alpha \leq y$.

  We now reinterpret this probability using the binomial distribution: The event $X^*_\alpha \leq y$ happens iff for at least α of the $X_1, ..., X_{\alpha+\beta-1}$ we have $X_i \leq y$. For each X_i we have $Pr[X_i \leq y] = y$; thus the probability that for at most $\alpha-1$ of the X_i's we have $X_i \leq y$ is $F^{B}_{\alpha+\beta-1,y}(\alpha-1)$. And so the probability that for at least α of the X_i's we have $X_i \leq y$ is $1 - F^{B}_{\alpha+\beta-1,y}(\alpha-1)$.

- **Lemma 1** For all non-negative integers $j, s \leq j$, and for all $y \in [0,1]$,

$$\mathbb{E}[X(j,s,y)] = \frac{1}{F^{B}_{j+1,y}(s)} - 1$$

Regret Bound for 2 arms case

$$\mathbb{E}\left[Y_j\right] \leq \mathbb{E}\left[\min\left\{X\left(j, s(j), \mu_2 + \frac{\Delta}{2}\right), T\right\}\right] + \mathbb{E}\left[\sum_{t=t_j+1}^{t_{j+1}} T \cdot I\left(\theta_2(t) > \mu_2 + \frac{\Delta}{2}\right)\right]$$

$$\mathbb{E}\left[Y_j \cdot I(j \geq j_0)\right] \leq \mathbb{E}\left[\min\left\{X\left(j, s(j), \mu_2 + \frac{\Delta}{2}\right), T\right\}\right] + \mathbb{E}\left[\sum_{t=t_j+1}^{t_{j+1}-1} T \cdot I\left(\theta_2(t) > \mu_2 + \frac{\Delta}{2}\right) \cdot I(j \geq j_0)\right]$$

This gives,

$$\begin{aligned}
\mathbb{E}\left[\sum_{j=j_0}^{T-1} Y_j\right] &\leq \sum_{j=0}^{T-1} \mathbb{E}\left[\min\left\{X\left(j, s(j), \mu_2 + \frac{\Delta}{2}\right), T\right\}\right] + T \cdot \sum_{j=0}^{T-1} \mathbb{E}\left[\sum_{t=t_j+1}^{t_{j+1}-1} I\left(\theta_2(t) > \mu_2 + \frac{\Delta}{2}, j \geq j_0\right)\right] \\
&\leq \sum_{j=0}^{T-1} \mathbb{E}\left[\min\left\{X\left(j, s(j), \mu_2 + \frac{\Delta}{2}\right), T\right\}\right] + T \cdot \sum_{t=1}^{T} \Pr\left(\theta_2(t) > \mu_2 + \frac{\Delta}{2}, k_2(t) \geq L\right)
\end{aligned}$$

Regret Bound for 2 arms case

- The last inequality holds because for any $t \in [t_j + 1, t_{j+1} - 1]$, $j \geq j_0$, by definition $k_2(t) \geq L$. $\overline{E_2}(t)$ is the event $\{\theta_2(t) > \mu_2 + \frac{\Delta}{2}$ and $k_2(t) \geq L\}$ used in the above equation. Next, we bound $\Pr\left(E_2(t)\right)$ and $\mathbb{E}\left[\min\left\{X\left(j, s(j), \mu_2 + \frac{\Delta}{2}\right), T\right\}\right]$.
- **Lemma 2**
$$\forall t, \quad \Pr\left(E_2(t)\right) \geq 1 - \frac{2}{T^2}$$

Regret Bound for 2 arms case

- **Lemma 3** Consider any positive $y < \mu_1$, and let $\Delta' = \mu_1 - y$. Also, let $R = \frac{\mu_1(1-y)}{y(1-\mu_1)} > 1$, and let D denote the KL-divergence between μ_1 and y, i.e. $D = y \ln \frac{y}{\mu_1} + (1-y) \ln \frac{1-y}{1-\mu_1}$.

$$\mathbb{E}[\mathbb{E}[\min\{X(j, s(j), y), T\} \mid s(j)]] \leq \begin{cases} 1 + \frac{2}{1-y} + \frac{\mu_1}{\Delta'} e^{-Dj} & j < \frac{y}{D} \ln R \\ 1 + \frac{R^y}{1-y} e^{-Dj} + \frac{\mu_1}{\Delta'} e^{-Dj} & \frac{y}{D} \ln R \leq j < \frac{4 \ln T}{\Delta'^2} \\ \frac{16}{T} & j \geq \frac{4 \ln T}{\Delta'^2} \end{cases}$$

where the outer expectation is taken over $s(j)$ distributed as Binomial (j, μ_1).

Regret Bound for 2 arms case

Fact For all $n, p \in [0, 1], \delta \geq 0$,

$$F^B_{n,p}(np - n\delta) \leq e^{-2n\delta^2},$$

,

$$1 - F^B_{n,p}(np + n\delta) \leq e^{-2n\delta^2},$$

$$1 - F^B_{n+1,p}(np + n\delta) \leq \frac{e^{4\delta}}{e^{2n\delta^2}}.$$

Regret Bound for 2 arms case

Case of large j : First, we consider the case of large j, i.e. when $j \geq 4(\ln T)/\Delta'^2$. Then, by simple application of Chernoff-Hoeffding bounds, we can derive that for any $s \geq \left(y + \frac{\Delta'}{2}\right) j$,

$$F^B_{j+1,y}(s) \geq F^B_{j+1,y}\left(yj + \frac{\Delta' j}{2}\right) \geq 1 - \frac{e^{4\Delta'/2}}{e^{2j\Delta'^2/4}} \geq 1 - \frac{e^{2\Delta'}}{T^2} \geq 1 - \frac{8}{T^2}$$

giving that for $s \geq y\left(j + \frac{\Delta'}{2}\right), \mathbb{E}[X(j+1, s, y)] \leq \frac{1}{\left(1 - \frac{8}{T^2}\right)} - 1$.

Again using Chernoff-Hoeffding bounds, the probability that $s(j)$ takes values smaller than $\left(y + \frac{\Delta'}{2}\right) j$ can be bounded as,

$$F^B_{j,\mu_1}\left(yj + \frac{\Delta' j}{2}\right) = F^B_{j,\mu_1}\left(\mu_1 j - \frac{\Delta' j}{2}\right) \leq e^{-2j\frac{\Delta'^2}{4}} \leq \frac{1}{T^2} < \frac{8}{T^2}$$

Regret Bound for 2 arms case

For these values of $s(j)$, we will use the upper bound of T. Thus,

$$\mathbb{E}[\min\{\mathbb{E}[X(j, s(j), y) \mid s(j)], T\}] \leq \left(1 - 8/T^2\right) \cdot \left(\frac{1}{(1 - 8/T^2)} - 1\right) + \frac{8}{T^2} \cdot T \leq \frac{16}{T}$$

Regret Bound for 2 arms case

Clubbing all, we get

$$
\begin{aligned}
\mathbb{E}\left[k_2(T)\right] &= L + \mathbb{E}\left[\sum_{j=j_0}^{T-1} Y_j\right] \\
&\leq L + \sum_{j=0}^{T-1} \mathbb{E}\left[\mathbb{E}\left[\min\left\{X\left(j, s(j), \mu_2 + \frac{\Delta}{2}\right), T\right\} \middle| \, s(j)\right]\right] + \sum_{t=1}^{T} T \cdot \Pr\left(\overline{E_2(t)}\right) \\
&\leq L + \frac{4 \ln T}{\Delta'^2} + \sum_{j=0}^{4(\ln T)/\Delta'^2 - 1} \frac{\mu_1}{\Delta'} e^{-Dj} + \left(\frac{y}{D} \ln R\right) \frac{2}{1-y} + \sum_{j=\frac{y}{D}\ln R}^{4(\ln T)/\Delta'^2 - 1} \frac{R^y e^{-D_j}}{1-y} + \frac{16}{T} \cdot T + 2 \\
&\leq \frac{40 \ln T}{\Delta^2} + \frac{48}{\Delta^4} + 18 \qquad (1)
\end{aligned}
$$

Regret Bound for 2 arms case

This gives a regret bound of

$$\mathbb{E}[\mathcal{R}(T)] = \mathbb{E}\left[\Delta \cdot k_2(T)\right] \leq \left(\frac{40 \ln T}{\Delta} + \frac{48}{\Delta^3} + 18\Delta\right)$$

Conclusion

Thank You