Solutions

Ruipeng Li, School of Physics

June 28, 2017

Abstract

Problem1 Multi-armed Bandit Problem
Problem2 Bike sharing economy

1 Problem1 UCB Alogrithm with prior knowledge

1.1 Problem Description

Assume there are K gambling machines marked by random variables $X_{i,n}$ for $1 \leq i \leq K$ and $n \geq 1$. Successive plays of machine i yield rewards $X_{i,1}, X_{i,2}, \cdots$ which are iid according to an unknown law wtih unknown expectation μ_i. Independence also holds for rewards across machines.

Every round K, we can take a gambling machine to play and record its rewards. After several rounds, we can explore the machine with highest average rewards. We want to optimize the regret, $i.e.$

$$\mu^* n - \mu_j \sum_{j=1}^{K} E[T_j(n)] \quad where \quad \mu^* = \max_{1 \leq i \leq K} \mu_i$$

Now we are given some prior knowledge about the arms(may be inaccurate), in the form of their estimated mean values $\tilde{\mu} = (\tilde{\mu_1}, \cdots, \tilde{\mu_N})$. Based on these, we build new upper confidence bound.

1.2 Analysis

Due to prior knowledge, we consider the multi-armed bandit model from a Bayesian point of view and a prior distribution. Therefore, we focus on the most important distribution which is one-parameter exponential family, one assumes that the parameter $\theta = (\theta_1, ..., \theta_k)$ is drawn from a prior distribution.

[Distribution]: $P = \{\nu_\theta, \theta \in \Theta : \nu_\theta$ has a density $f_\theta(x) = exp(\theta x - b(\theta))$ w.r.t $\xi\}$ where $\Theta = (\theta^-, \theta^+) \subset R$ is an open interval, b a twice-differentiable and convex function(called the log-partition function) and ξ a reference measure. As we all know, if $X \sim \nu_\theta$, it can be shown that E[X]=$\dot{b}(\theta)$ and Var[X]=$\ddot{b}(\theta) > 0$, where $\dot{b}(resp.\ddot{b})$ is the derivative(resp. second derivative) of b with respect to the natural parameter θ.

At first, we will focus on Bernoulli distribution because it maximizes deviations among bounded variables with given expectation. As we all know, $H = -\sum_M P_M ln P_M$, we have maximum entropy distribution $P_M = \frac{1}{Z(\lambda)} exp\{\lambda r(M)\}$, where $Z(\lambda) = \sum exp\{\lambda r(M)\}$. From this formula we know that the maximum entropy distribution is $P_M = p^r(1-p)^{n-r}$. Then we will focus on one-parameter exponential family.

[Prior & Posterior Distributions]: For different conditional distribution, we usually take different conjugate priors. For example, if the distribution is Bernoulli distribution $B(\mu)$, we will take Beta distribution as our prior distribution Beta(a,b) and the posterior distribution is Beta(a+nx,b+n(1-x)), where n is the number of observations and x is empirical mean.

[Confidence Interval]: We need to build up an upper confidence bound, so I use the notation that $Q(\alpha, \pi)$ is the quantile of order α of the distribution π. And we need to find the order α. For traditional UCB1 algorithm, the order is $1 - \frac{1}{t}$, and we can make some change to improve the performance. We also have $KL(\nu_\theta, \nu_\lambda) = \dot{b}(\theta)(\theta - \lambda) - b(\theta) + b(\lambda)$, to simplify our notation, we introduce the KL-divergence between the distributions of mean μ and μ' :

$$d(\mu, \mu') := KL(\nu^\mu, \nu^{\mu'}) = KL(\nu_{\dot{b}^{-1}(\mu)}, \nu_{\dot{b}^{-1}(\mu')})$$

1.3 Assumption

- The joint distribution of $X_{i,t}$ is stationary, which means the distribution will not vary due to the choice of players and time. (Due to this assumption, I can use regret as our performance standard)
- $X_{i,1}, X_{i,2}, \cdots$ are i.i.d. and independence also holds for rewards across machines. (To simplify the problem, I want to see the simplest situation)
- We assume the distribution of all k machings subject to one-parameter canonical exponential family. (Because the general condition is too hard for me to analyze, I want to design an algorithm for the most important case)

Algorithm 1 Bayes-UCB

Input: Π^0 (The initial prior on θ); c (parameters of the quantile); n (horizon);

for $t = 1$ to n do
  for each arm j=1,...,K do
    compute
      $q_j(t) = Q(1 - \frac{1}{t(\log t)^c}, \pi_j^{t-1})$
  end for
  draw arm $I_t = arg\max_{j=1\ldots K} q_j(t)$
  get reward $Y_t = X_{I_t,t}$ and update Π^t according to $\pi_j^t(\theta_j) \propto \nu_{\theta_j}(Y_t)\pi_j^{t-1}(\theta_j)$
end for

1.4 Performance

We assume that the rewards have a Bernoulli distribution, and when the prior is the Beta(1, 1), or uniform, law. We have the following theorem. We show that the Bayes- UCB algorithm is optimal, in the sense that it reaches the lower-bound of Lai and Robbins.

Theorem 1.1 *(Bernoulli Case) For any $\epsilon > 0$, choosing the parameter $c \geq 5$ in the Bayes-UCB algorithm, the number of draws of any sub-optimal arm j is upper-bounded by*

$$E[T_j(n)] \leq \frac{1+\epsilon}{d(\mu_j, \mu^*)} log(n) + o_{\epsilon,c}(log(n))$$

Proof: Proof is considered in [**?**].

Then we consider that the rewards have a one-parameter canonical exponential family. This index policy is also asympotically optimal which reaches the lower-bound of Lai and Robbins.

Lemma 1.2 *For exponential family function, we have such inequality:*
1. There exists two positive constants A and B such that for all x,v that satisfy $\mu_0- < x < v < \mu_0+$,for all $n \geq 1$,for all $a \in \{1, \cdots, K\}$,

$$\frac{A}{n} e^{-nd(x,v)} \leq P(\nu < X \leq \mu^+) \leq B\sqrt{n} e^{-nd(x,\nu)}$$

2. There exists C such that for all x,v that satisfy $\mu_0- < x < v < \mu_0+$,for all $n \geq 1$,for all $a \in \{1, \cdots, K\}$,

$$P(\nu < X \leq \mu^+) \geq \frac{C}{\sqrt{n}}$$

Proof: Proof is considered in [**?**].

Theorem 1.3 *(One-parameter canonical exponential family Case) Let ν^μ be an exponential bandit model. Assume that for all a, π_a^0 has a density f_a w.r.t the Lebesgue measure such that $f_a(u) > 0$ for all $u \in J = \dot{b}(\Theta)$. Let $\epsilon > 0$, the algorithm that draws each arm once and for $t \geq K$ selects at time t+1*

$$A_{t+1} = arg \max_a q_a(t)$$

, which satisfies

$$\forall a \neq a^*, \quad E[T_a(n)] \leq \frac{1+\epsilon}{d(\mu_a, \mu^*)} log(n) + o_\epsilon(log(n))$$

Proof: Without losing generality, we assume arm 1 to be optimal and arm a to be a suboptimal arm.

$$E[T_a(n)] = E[\sum_{t=0}^{n-1} 1_{(A_{t+1}=a)}] \quad \hat{\mu}_a(t) = (X_{a,1} + X_{a,2} + \cdots + X_{a,s})/s = \hat{\mu}_{a,T_a(t)}$$

When we play a at round t+1, the condition must be $q_a(t) \geq q_1(t)$, we have

$$E(T_a(n)) \leq \sum_{t=0}^{n-1} P(\mu_1 - g_t \geq q_1(t)) + \sum_{t=0}^{n-1} P(\mu_1 - g_t \leq q_a(t), A_{t+1} = a)$$

, where g_t is decreasing sequence and the composition is used by the one used for KL-UCB. [**?**]

For the first term, due to the lower bound in lemma

$$\{\mu_1 - g_t \geq q_1(t)\} = \{F_{\pi_1}(\mu_1 - g_t) \leq 1 - \frac{1}{tlog^c t}\} \subseteq \{\frac{Ae^{-T_1(t)}d(\hat{\mu}_1, \mu_1 - g_t)}{T_1(t)} \leq \frac{1}{tlog^c(t)}\}$$

For the second term, due to the upper bound in lemma

$$\{\mu_1 - g_t \leq q_a(t), A_{t+1} = a\} = \{P(\mu_1 - g_t < \hat{\mu}_a \leq \mu^+) \geq \frac{1}{tlog^c t}, A_{t+1} = a\}$$

$$\subseteq \{B\sqrt{T_a(t)}e^{-T_a(t)d(\hat{\mu}_a, \mu_1 - g_t)} \geq \frac{1}{tlog^c t}, A_{t+1} = a\}$$

Let's sum such two terms, from which we can see the first term is the order of o(log(n)) and the second term is the order of log(n).

For the first term, $\sum_{t=0}^{n-1} P(T_1(t)d(\hat{\mu}_1(t), \mu_1 - g_t) \geq log(\frac{Atlog^c t}{T_1(t)}))$ from [**?**] lemma 5, we know that it's the order $o(log(n))$.

For the second term, it's more complicate. $\sum_{t=0}^{n-1} P(B\sqrt{T_a(t)}e^{-T_a(t)d(\hat{\mu}_a, \mu_1 - g_t)} \geq \frac{1}{tlog^c t}, A_{t+1} = a) \leq \sum_{t=0}^{n-1}\sum_{s=1}^{t} P(B\sqrt{T_a(t)}e^{-T_a(t)d(\hat{\mu}_a, \mu_1 - g_t)} \geq \frac{1}{tlog^c t}, A_{t+1} = a, T_a(t) = s) \leq \sum_{s=1}^{n} P(sd(\hat{\mu}_{a,s}, \mu_1 - g_s) \leq logn + cloglogn + log(B) + \frac{1}{2}logs)$ from [**?**] lemma 6, we know that it's the order $log(n) + o(log(n)$

From these two terms, we have conclusion that $E(T_a(n)) \sim log(n)$. Similar Proof is considered in [**?**]

1.5 Conclusion

From the prior knowledge $\tilde{\mu} = (\tilde{\mu_1}, \cdots, \tilde{\mu_N})$, we can construct prior distribution Π^0 with such mean. When we use the algorithm Bayes-UCB, we don't need to test each gambling machine once so that we can reduce K rounds. Then we will execute the algorithm to find upper confidence Bound and play the gambling machine with the highest rank. After that, we can update the posterior distribution and take another round. It's a well-proved algorithm which is suitable for such problem. So the proof is similar in [**?**] with some different details.

2 Problem2 Bike sharing economy

2.1 Analysis

We apply game theory framework to such problems. Different from previous model like Cournot Model and Bertrand Model, different companies offer different platforms with network effect, which are different products. Actually sharing bicycle martket is not a two-side market, it's a rental

market with network externality instead. Its mode is called B2C(business to customer) unlike Uber which is a typical C2C company.

First of all, we will analyze the action of each company respectively according to the Nash equilibrium. Then we will analyze the benefit of sharing.

2.2 Model and Assumption

Problem Description: For Producers, We assume that there are K companies, each of which offers N_i sharing bike (i=1,...K). Each company will require deposits and go after maximum profits. For consumers, in this problem, we assume that consumers' decision can be divided into two phases. First, they would choose a platform to pay for its deposit according to the total number of bicycles per person. Then they will rent bicycle depends on its price. The reason why we can divide the decision into two phases is that due to competition, the price among different companies remains the same so that they can attract consumers equally.

2.2.1 Modeling for Producers

There are K companies, each of which offers N_i sharing bike (i=1,...K).Profits (π_j) follow such formula: $\pi_j = p \times q_j - c_j/T_j \times N_j$,where p means the income per time per bicycle, T_j is average life of the bike,c_i is the cost of each bicycle and q_j means the number of customers who choose bicycles provided by this company.
Assumption:

- We assume that each company offer sharing bicycle with the same price p. Due to people's sensitiveness to price and low cost of changing platform, any bicycle with higher price will lose its consumers.
- We assume that the cost includes fixed costs, production costs and their maintenance costs. The cost function has such formula: $c(N_j) = c_j N_j$, because I think that the margin cost must remain the same for sharing bike.
- The bicycles are all the same.

2.2.2 Modeling for Consumers

For consumers, we need two-phase decision. For the first phase, we assume that $q_j \propto N_j$. Actual it means that everyone have the same opportunities to use sharing bicycle. So $\forall j, q_j/N_j = \lambda$, where λ means that λ persons use one bicycle per unit time. For the second phase, we assume linear demand function.
Assumption:

- we assume the first phase, $q_j = \lambda N_j$
- we assume linear demand function $q = \sum q_j = A - Bp$

2.2.3 Modeling for Market

We assume that the market is perfect information because they can estimate the demand function roughly. *Assumption:*

- there are no transaction costs and friction in such markets.
- Every company will estimate the demand function correctly and know other companies' yields of bicycles.

2.3 An naive Analysis

We list following equations:

$$max(\pi_j) = max(pq_j - \frac{c_j}{T_j}N_j) \tag{1}$$

$$s.t. \quad q_j = \lambda N_j \tag{2}$$

$$\sum q_j = A - Bp \tag{3}$$

And we will get a quadratic response function.

$$max(\pi_j) = max\{\frac{-\lambda^2 N_j^2}{B} + (\frac{A}{B}\lambda - \frac{c_j}{T_j} - \frac{\lambda^2 \sum\limits_{i \neq j} N_i}{B})N_j\} \tag{4}$$

$$N_j = \frac{A\lambda - Bc_j/T_j - \lambda^2 \sum\limits_{i \neq j} N_i}{2\lambda^2} \tag{5}$$

It reaches Nash Equilibrium. We take sum for 1 to K.

$$N = \sum_{j=1}^{k} N_j = \frac{AK\lambda - B\sum_j (c_j/T_j)}{\lambda^2 \times (K+1)} \tag{6}$$

2.4 Analysis with Platform effect

The difference for platform effect is that λ is not a constant, but a increasing function with N_j We list following equations:

$$max(\pi_j) = max(pq_j - \frac{c_j}{T_j}N_j) \tag{7}$$

$$s.t. \quad q_j = \lambda(N_j)N_j \tag{8}$$

$$\sum q_j = A - Bp \tag{9}$$

And we will get a response function.

$$\max(\pi_j) = \max\{-\frac{\lambda(N_j)^2 N_j^2}{B} + (\frac{A}{B}\lambda(N_j) - \frac{c_j}{T_j} - \frac{\lambda(N_j)^2 \sum\limits_{i\neq j} N_i}{B})N_j\} \tag{10}$$

$$take \quad \lambda(N_j) = \lambda N_j^{\alpha}, \alpha \in [0,1) \tag{11}$$

$$0 = (2+2\alpha)\lambda N_j^{2\alpha+1} + \lambda(2\alpha+1)(\sum_{i\neq j} N_i)N_j^{2\alpha} - A\lambda(1+\alpha)N_j^{\alpha} + \frac{c_j}{T_j}B \tag{12}$$

2.5 The benefit of sharing

We assume that the number of person who want to buy bicycles is T, $T = \mu(N)\sum\limits_j \lambda_j N_j$. μ means that the ratio of the persons who buy bicycles to the total persons who rent bicycles and it will decrease if N increases because people would like to rent bicycles instead of purchasing bicycles with the increasing number of total sharing bicycles.

Assumption: $N \uparrow \mu(N) \downarrow$,$N\mu(N) \uparrow$

We define benefits $G = \sum\limits_{j=1}^{k}(\mu(N)\lambda(N_j)N_j)/rate - N_j$, where rate means the average ratio of utility time to total time if anyone buy a bicycle.

In such naive models, $G = \dfrac{AK\lambda - B\sum_j(c_j/T_j)}{\lambda^2 \times (K+1)}(\mu(N)\lambda/rate - 1)$. From this model, we can get some intuitive insights. First of all, if $K \to \infty$, $G \to \frac{A}{\lambda}(\mu(N)\lambda/rate - 1)$, given appropriate parameter, the benefits from sharing bicycles will approximately be a value with K. It means that with the competition increasing, N will increase. However, the change of benefits depend on the formula of $\mu(N)$. Generally, the benefits will increase first and then decrease with K increasing. To sum up, it means that appropriate competition will increase the benefits while excessive competition will decrease the benefits.

With such platforms, we calculate (12). $G = \sum\limits_{j=1}^{k}\mu(N)\lambda N_j^{1+\alpha}/rate - N_j$. We can get numerical results by setting parameters reasonably. But we focus on the competition K. If k increase, we know that N will increase. There will be a dilemma between $\mu(N)$ and $N^{1+\alpha}$. With more data, I think I can further my research by numerical simulation.