Paper Review

Asymptotically Efficient Adaptive Allocation Rules

T. L. LAI AND **HERBERT ROBBINS**

Department of Statistics, Columbia University, New York, New York 10027

Content

1. Introduction

1.1. Parameter

$\prod_j$ $(j = 1, \dots, k)$: unknown probability distribution of the reward from arm j specified respectively by univariate probability density functions $f(x; \theta_j)$

x : sample , θ_j : parameter of arm j

What is the multi-armed bandit problem?

How should we sample $x_1, x_2, \dots$ sequentially from the k populations in order to achieve the greatest possible expected value of the sum $S_n = x_1 + \cdots + x_n$ as $n \to \infty$?

Rule : The player wants to choose at each stage one of the k arms, the choice depending in some way on the record of previous trials.

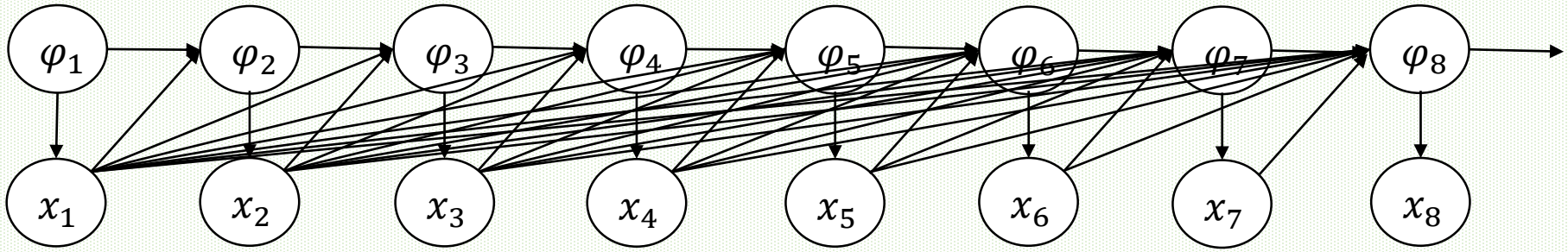
Goal : to maximize the long-run total expected reward

1. Introduction

1.1. Parameter

φ : adaptive allocation rule
φ_n : selected arm at stage n
(ex) event $\{\varphi_n = j\} \rightleftharpoons$ sample from $\prod_j$ $(j = 1, \dots, k)$ at stage n

Graphical Model - example



1. Introduction

1.1. Parameter

$\mu(\theta_j)$: mean of the reward from the arm j whose parameter is θ_j

$T_n(j) = \sum_{i=1}^{n} I\{\varphi_i = \text{j}\}$: the number of times that φ samples from $\prod_j$ up to stage n

$$ES_n = \sum_{j=1}^{k} \mu(\theta_j) ET_n(j)$$

$\mu^* = \max\{\mu(\theta_1), \dots, \mu(\theta_k)\} = \mu(\theta^*)$ for some $\theta^* \in \{\theta_1, \dots, \theta_k\}$

$$\mu(\theta) = \int_{-\infty}^{\infty} x f(x;\theta) dv(x) < \infty$$

1. Introduction

1.1. Parameter

Kullback – Leibler number

$$I(\theta,\lambda) = \int_{-\infty}^{\infty}[\log(\frac{f(x;\theta)}{f(x;\lambda)})]f(x;\theta)dv(x) \qquad 0 \le I(\theta,\lambda) \le \infty$$

$\forall\varepsilon > 0$ and $\forall\theta,\lambda$ such that $\mu(\lambda) > \mu(\theta)$, $\exists\delta = \delta(\epsilon,\theta,\lambda) > 0$ for which $|I(\theta,\lambda) - I(\theta,\lambda')| < \varepsilon$
Whenever $\mu(\lambda) \le \mu(\lambda') \le \mu(\lambda) + \delta$

To explain it is possible to find the new parameter λ' whose mean is bigger than existing parameter's.
Almost every probability distribution function which we know satisfies above property.

(ex) Bernoulli, Poisson, exponential

1. Introduction

1.2. Hypothesis or Condition

$\int_{-\infty}^{\infty} |x| f(x;\theta) dv(x) < \infty$ for all $\theta \in \Theta$

Probability density function should have only one parameter. (univariate density function)

Definition

1. **Consistent** := the rules that satisfy $\lim_{n\to\infty} \boldsymbol{n^{-1} ES_n} = \boldsymbol{\mu^*}$ whenever the $\mu(\theta_j)$ are not all equal

2. **Asymptotically Efficient** := the rules that satisfy

$$R_n(\theta_1, \ldots, \theta_k) \sim \left\{ \sum_{j:\, \mu(\theta_j) < \mu^*} \frac{\mu^* - \mu(\theta_j)}{I(\theta_j, \theta^*)} \right\} \log n \qquad \text{as } n \to \infty$$

1. Introduction

1.3. Objective

Maximizing ES_n ⟺ Minimizing the "regret"

Regret

$$R_n(\theta_1, \dots, \theta_k) = n\mu^* - ES_n = \sum_{j:\, \mu(\theta_j) < \mu^*} (\mu^* - \mu(\theta_j)) ET_n(j)$$

Total Reward

$$ES_n = \sum_{j=1}^{k} \mu(\theta_j) ET_n(j)$$

Content

2. Lower Bound
For The Expected Sample Size
From an Inferior Population

2. Lower Bound For The Expected Sample Size From an Inferior Population

Before proof, below conditions are important

1. $\forall \varepsilon > 0$ and $\forall \theta, \lambda$ such that $\mu(\lambda) > \mu(\theta)$, $\exists \delta = \delta(\epsilon, \theta, \lambda) > 0$ for which $|I(\theta, \lambda) - I(\theta, \lambda')| < \varepsilon$
Whenever $\mu(\lambda) \leq \mu(\lambda') \leq \mu(\lambda) + \delta$

2. Let φ be a rule whose regret satisfies, for each fixed $\theta = (\theta_1, \ldots, \theta_k)$, the condition that as $n \to \infty$
$R_n(\theta) = o(n^a)$ for every a > 0
Calculating the regret except for the best arm.

What does this condition mean?

2. Lower Bound For The Expected Sample Size From an Inferior Population

Condition 2

I think paper wants to say below things.

Let φ be a rule whose regret satisfies, for each fixed $\theta = (\theta_1, \dots, \theta_k)$, the condition that as $n \to \infty$

$R_n(\theta) = o(f(x))$, Function f(x) satisfies below.

$$\lim_{n\to\infty} \frac{f(x)}{\log n} = \infty$$

Not only for n^a, but also for f(x) which satisfies above condition.

And for every function f(x) which satisfies $\lim_{n\to\infty} \frac{f(x)}{n} = \infty$, then

$$\lim_{n\to\infty} n^{-1} ES_n = \mu^*$$

2. Lower Bound For The Expected Sample Size From an Inferior Population

The rules satisfies those

Consistent := the rules that satisfy $\lim_{n\to\infty} n^{-1}ES_n = \mu^*$ whenever the $\mu(\theta_j)$ are not all equal

Asymptotically Efficient := the rules that satisfy $R_n(\theta_1, \dots, \theta_k) \sim \left\{\sum_{j:\,\mu(\theta_j)<\mu^*} \frac{\mu^*-\mu(\theta_j)}{I(\theta_j,\theta^*)}\right\}\log n$ as $n \to \infty$

2. Lower Bound For The Expected Sample Size From an Inferior Population

Theorem 1. Assume above conditions are satisfied, then for every θ such that the $\mu(\theta_j)$ are not equal.

$$\lim_{n\to\infty} inf \frac{R_n(\theta)}{\log n} \geq \sum_{j:\,\mu(\theta_j)<\mu^*} \frac{\mu^* - \mu(\theta_j)}{I(\theta_j, \theta^*)}$$

2. Lower Bound For The Expected Sample Size From an Inferior Population

$\theta = (\theta_1, \dots, \theta_k)$

let P_θ denote the probability measure under which θ_j is the parameter corresponding to population Π_j, j=1,…,k.

(Thinking easily, P_θ is the rules you want to apply.)

$$\Theta_j = \{\theta: \mu(\theta_j) < \max \theta_i \ (i \neq j)\} \ (\text{“}\theta_j \text{ is not best”})$$
$$\Theta_j^* = \{\theta: \mu(\theta_j) > \max \theta_i \ (i \neq j)\} \ (\text{“}\theta_j \text{ is the unique best”})$$

Theorem 2. Fix $j \in \{1, \dots, k\}$, and define Θ_j and Θ_j^*. Let φ be any rule such that for every $\theta \in \Theta_j^*$ $\sum_{i \neq j} E_\theta T_n(i) = o(n^a)$ for every a > 0. Then for every $\theta \in \Theta_j$ and every $\epsilon > 0$,

$$\lim_{n \to \infty} P_\theta\{T_n(j) \geq (1-\epsilon)(\log n)/\, I(\theta_j, \theta^*)\} = 1,$$

And hence

$$\lim_{n \to \infty} inf\, E_\theta T_n(j)/\log n \geq \frac{1}{I(\theta_j, \theta^*)}$$

2. Lower Bound For The Expected Sample Size From an Inferior Population

Proof of Theorem 2

[WLOG]
Fix j=1, $\theta \in \Theta_1$, and $\theta^*=\theta_2$, then $\mu(\theta_2) > \mu(\theta_1)$ and $\mu(\theta_2) \geq \mu(\theta_i)$ for other i.

Fix any $0 < \delta < 1$. we can choose $\lambda \in \Theta$ such that
$\mu(\lambda) > \mu(\theta_2)$ and $|I(\theta_1, \lambda) - I(\theta_1, \theta_2)| < \delta I(\theta_1, \theta_2)$

If we Misunderstand the 1st arm for the best arm
(Worst case)

Define the new parameter vector $\gamma = (\lambda, \theta_2, \ldots, \theta_k)$.
Then $\gamma \in \Theta_1^*$, so $\sum_{i \neq 1} E_\theta T_n(i) = o(n^a)$ for every a > 0

Expected number of selecting arms except best arm

By Markov inequality

With $0 < a < \delta$, and therefore

$$(n - O(\log n)) P_\gamma\{T_n(1) < (1-\delta)(\log n)/I(\theta_1, \lambda)\} \leq E_\gamma(n - T_n(1)) = o(n^a)$$

Proof of Theorem 2

Let Y_1, Y_2, ... denote successive samples from Π_1, and defining $L_m = \sum_{i=1}^{m} \frac{\log(f(Y_i;\theta_1))}{\log(f(Y_i;\lambda))}$, it follows that

$P_\gamma(C_n) = o(n^{a-1})$, where
$C_n = \{T_n(1) < (1-\delta)\big((\log n)/\, I(\theta_1, \lambda)\ and\ L_{T_n(1)} \le (1-a)\log n\big\}$.

Note that

$$P_\gamma\{T_n(1) = n_1, \dots, T_n(k) = n_k, L_{n_1} \le (1-a)\log n\} = \int_{\{T_n(1)=n_1,\dots,T_n(k)=n_k,\ L_{n_1}\le(1-a)\log n\}} dP_\gamma$$

$$= \int_{\{T_n(1)=n_1,\dots,T_n(k)=n_k,\ L_{n_1}\le(1-a)\log n\}} \frac{dP_\gamma}{dP_\theta} \times dP_\theta$$

$$= \int_{\{T_n(1)=n_1,\dots,T_n(k)=n_k,\ L_{n_1}\le(1-a)\log n\}} \prod_{i=1}^{n_1} \frac{f(Y_i;\lambda)}{f(Y_i;\theta_1)} \times dP_\theta$$

2. Lower Bound For The Expected Sample Size From an Inferior Population

$$\int_{\{T_n(1)=n_1,\ldots,T_n(k)=n_k,\ L_{n_1}\le(1-a)\log n\}} \prod_{i=1}^{n_1} \frac{f(Y_i;\lambda)}{f(Y_i;\theta_1)} \times dP_\theta$$

$$\geq \exp(-(1-a)\log n)\cdot P_\theta\{T_n(1)=n_1,\ldots,T_n(k)=n_k, L_{n_1}\leq(1-a)\log n\}$$

And $C_n = \bigcup_{n_1+\cdots+n_k=n \text{ and } n_1<(1-\delta)\ (\log n)/I(\theta_1,\lambda)} \{T_n(1)=n_1,\ldots,T_n(k)=n_k, L_{n_1}\leq(1-a)\log n\}$
This is a disjoint union of events.

$$P_\theta(C_n) \leq n^{1-a} P_\gamma(C_n) \ \to 0 \ as\ n\to\infty$$

By the strong law of large numbers, $L_m/m \to I(\theta_1,\lambda) > 0$, and
therefore $max_{i\le m} L_i/m \to I(\theta_1,\lambda)$, a.s. $[P_\theta]$

Since $1-a > 1-\delta$, it follows that
$P_\theta\{L_i > (1-a)\log n \ for\ some\ i < (1-\delta)(\log n)/\ I(\theta_1,\lambda)\} \to 0\ as\ n\to\infty$

2. Lower Bound For The Expected Sample Size From an Inferior Population

$$\lim_{n\to\infty} P_\theta\{T_n(1) < (1-\delta)(\log n)/\, I(\theta_1, \lambda\,)\}=0$$

This implies that $\lim_{n\to\infty} P_\theta\{T_n(1) < (1-\delta)(\log n)/[(1+\delta)\, I(\theta_1, \theta_2)]\}=0$

Theorem 1 : asymptotic lower bound for the regret
Theorem 2 : if φ is an asymptotically efficient rule, then the number of observations that φ takes from any inferior population Π_j up to stage n is about $(\log n)/I(\theta_j, \theta^*)$

As n increases we should be increasingly confident that we are not sampling from an inferior population

2. Lower Bound For The Expected Sample Size From an Inferior Population

From now, we need to make sure that the other populations have been sampled enough for us to be reasonably confident that they are indeed inferior.

One way of doing this is to compare certain **upper confidence bounds** for the mean of an apparently inferior population with the estimated mean of the leader.

Content

3. Construction of Asymptotically Efficient Allocation Rules
– Upper Confidence Bound-

3. Construction of Asymptotically Efficient Allocation Rules – Upper Confidence Bound-

[Sample-from-the-leader]

Define the leader at stage n as the population with the largest estimated mean among all populations that have been sampled δn times, for some predetermined positive number δ<1/k.

And adapt some rules!!

3. Construction of Asymptotically Efficient Allocation Rules – Upper Confidence Bound-

[φ^* : rule with upper confidence bound]

To begin with, at stage j =1,2,…,k, the rule takes one observation Π_j

Choose j_n such that $\hat{\mu}_n(j_n) = \max\{\hat{\mu}_n(j) : T_n(j) \geq \delta n\}$. $0 < \delta < \frac{1}{k}$

At stage n+1, n +1 = km + j (unique j)

For this j, we decide whether we take an observation from Π_j or Π_{j_n}.

$\hat{\mu}_n(j_n) \leq U_n(j)$, then observation from Π_{j_n}

Otherwise, observation from Π_j

3. Construction of Asymptotically Efficient Allocation Rules – Upper Confidence Bound-

φ^* be the rule by choosing the next arm through the upper confidence bound.

Theorem 3.

(i) For every $\theta = (\theta_1, \dots, \theta_k)$ and every j such that $\mu(\theta_j) < \mu(\theta^*)$,

$$E_\theta T_n(j) \leq \left(\frac{1}{I(\theta_j, \theta^*)} + o(1) \right) \log n \, .$$

(ii) $E_\theta T_n(j) \sim \frac{\log n}{I(\theta_j, \theta^*)}$ for every j such that $\mu(\theta_j) < \mu(\theta^*)$

3. Construction of Asymptotically Efficient Allocation Rules – Upper Confidence Bound-

3. Construction of Asymptotically Efficient Allocation Rules – Upper Confidence Bound-