بسم الله الرحمن الرحیم

یادگیری بندیت

جلسه ۴:
بندیت تصادفی با تعداد دسته متناهی (گردش و تجربه)

ترم بهار ۱۳۹۹ - ۱۴۰۰

الگوریتم «گردش سپس تعهد»

Explore Then Commit (ETC)

فصل ۶

برای σـ زیرگوسی‌ها

$$\mathbb{P}\left(\hat{\mu} \geq \mu + \varepsilon\right) \leq \exp\left(-\frac{n\varepsilon^2}{2\sigma^2}\right)$$

الگوریتم «گردش سپس تعهد»

انتخاب مرحله i

$$A_t = \begin{cases} (t \bmod k) + 1, & \text{if } t \leq mk; \\ \operatorname{argmax}_i \hat{\mu}_i(mk), & t > mk. \end{cases}$$

m مرحله بگرد

سپس تعهد

الگوریتم «گردش سپس تعهد»

انتخاب مرحله i

$$A_t = \begin{cases} (t \bmod k) + 1, & \text{if } t \leq mk; \\ \operatorname{argmax}_i \hat{\mu}_i(mk), & t > mk. \end{cases}$$

m مرحله بگرد

سپس تعهد

$$R_n \leq m \sum_{i=1}^{k} \Delta_i + (n - mk) \sum_{i=1}^{k} \Delta_i \exp\left(-\frac{m\Delta_i^2}{4}\right)$$

$$R_n \leq \min\left\{n\Delta,\ \Delta + \frac{4}{\Delta}\left(1 + \max\left\{0,\ \log\left(\frac{n\Delta^2}{4}\right)\right\}\right)\right\}$$

الگوریتم «گردش سپس تعهد»

انتخاب مرحله i

$$A_t = \begin{cases} (t \bmod k) + 1, & \text{if } t \leq mk; \\ \operatorname{argmax}_i \hat{\mu}_i(mk), & t > mk. \end{cases}$$

m مرحله بگرد

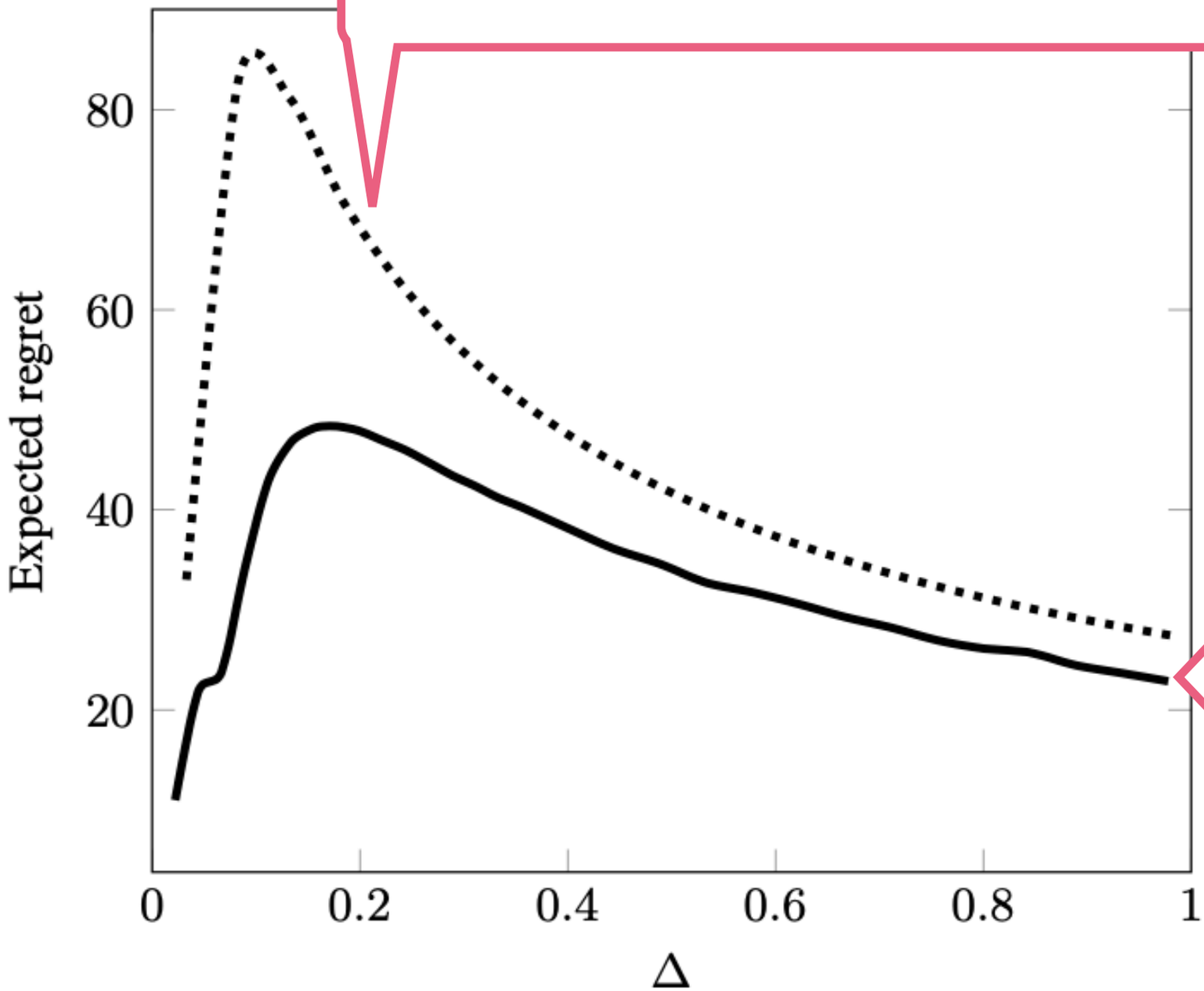
سپس تعهد

$$R_n \leq m \sum_{i=1}^{k} \Delta_i + (n - mk) \sum_{i=1}^{k} \Delta_i \exp\left(-\frac{m\Delta_i^2}{4}\right)$$

$$R_n \leq \min\left\{n\Delta,\ \Delta + \frac{4}{\Delta}\left(1 + \max\left\{0,\ \log\left(\frac{n\Delta^2}{4}\right)\right\}\right)\right\}$$

$$\limsup_{n\to\infty} \frac{R_n}{\log(n)} \leq \sum_{i:\Delta_i>0} \frac{4}{\Delta_i}.$$

$$R_n \leq \min\left\{n\Delta,\ \Delta + \frac{4}{\Delta}\left(1 + \max\left\{0,\ \log\left(\frac{n\Delta^2}{4}\right)\right\}\right)\right\}$$



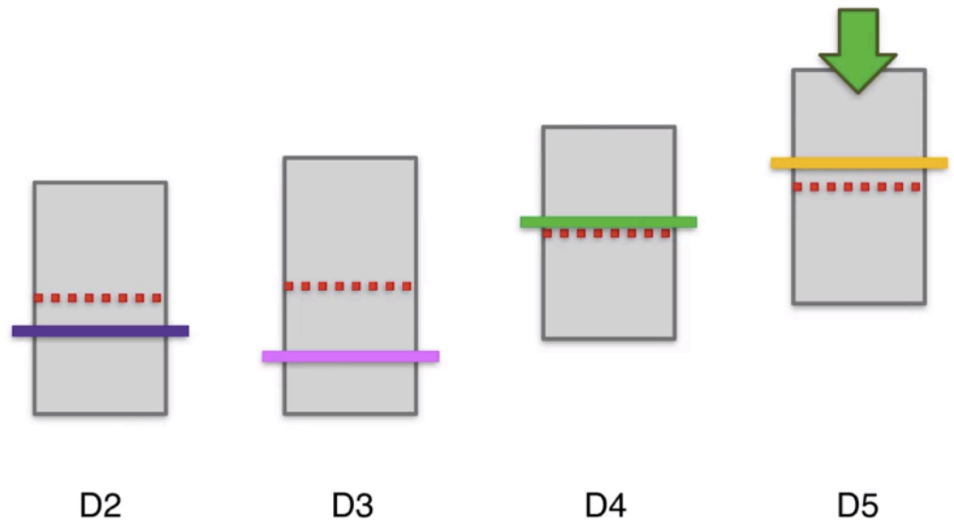
الگوریتم «گردش سپس تعهد» با

$$m = \max\left\{1,\ \left\lceil \frac{4}{\Delta^2}\log\left(\frac{n\Delta^2}{4}\right)\right\rceil\right\}$$

الگوریتم «کران بالای اطمینان»

Upper Confidence Bound (UCB)

فصل ۷



الگوریتم UCB(δ)

$$= \begin{cases} \infty & \text{if } T_i(t-1) = 0 \\ \hat{\mu}_i(t-1) + \sqrt{\frac{2\log(1/\delta)}{T_i(t-1)}} & \text{otherwise.} \end{cases}$$

Input k and δ

for $t \in 1, \ldots, n$ **do**

Choose action $A_t = \text{argmax}_i \, \text{UCB}_i(t-1, \delta)$

Observe reward X_t and update upper confidence bounds

end for

الگوریتم UCB(δ)

$$= \begin{cases} \infty & \text{if } T_i(t-1) = 0 \\ \hat{\mu}_i(t-1) + \sqrt{\frac{2\log(1/\delta)}{T_i(t-1)}} & \text{otherwise.} \end{cases}$$

Input k and δ

for $t \in 1, \ldots, n$ **do**

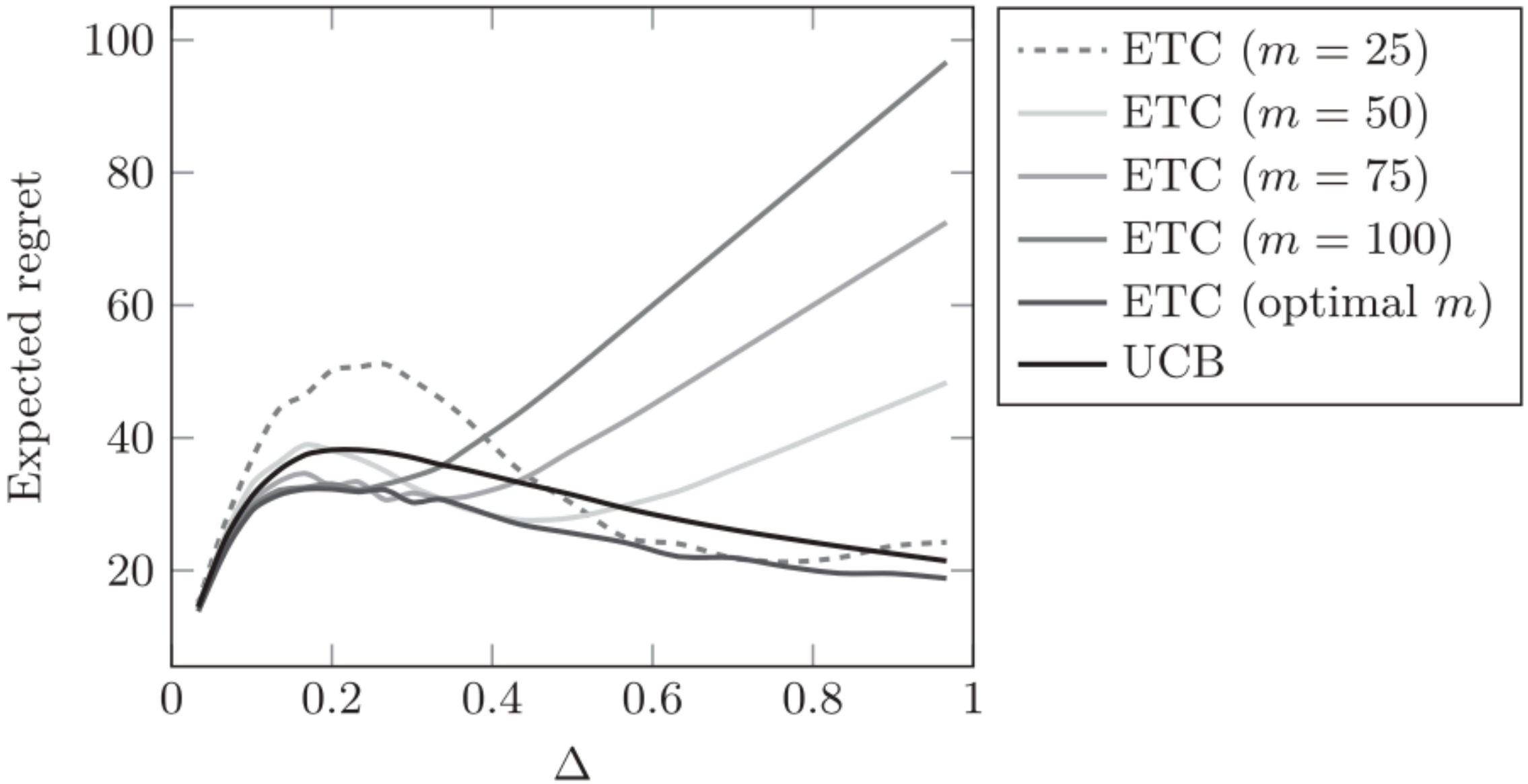
Choose action $A_t = \text{argmax}_i \, \text{UCB}_i(t-1, \delta)$

Observe reward X_t and update upper confidence bounds

end for

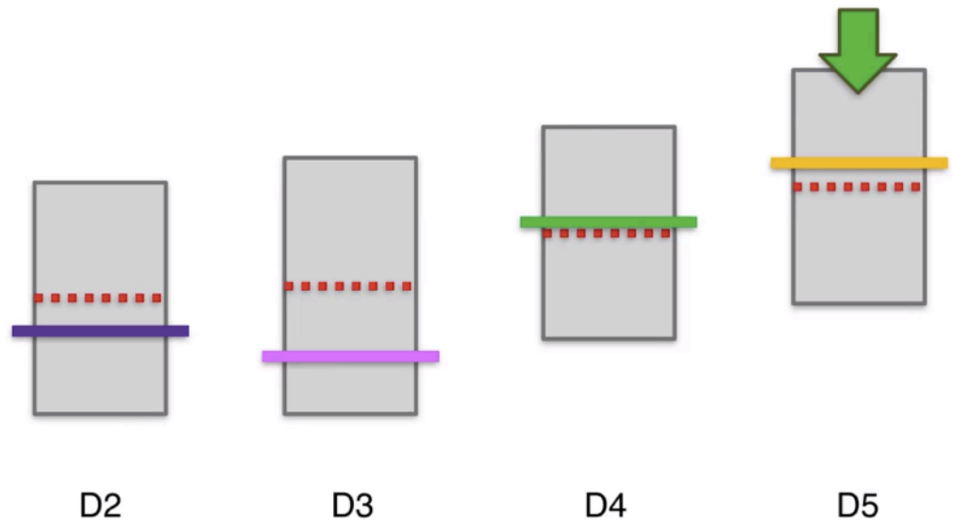
$$R_n \leq 3 \sum_{i=1}^{k} \Delta_i + \sum_{i:\Delta_i > 0} \frac{16 \log(n)}{\Delta_i}.$$

$$\leq 8\sqrt{nk \log(n)} + 3 \sum_{i=1}^{k} \Delta_i$$


Expected regret
100
80
60
40
20
0
0.2
0.4
0.6
0.8
1
Δ
ETC ($m = 25$)
ETC ($m = 50$)
ETC ($m = 75$)
ETC ($m = 100$)
ETC (optimal m)
UCB

الگوریتم «کران بالای اطمینان» ـ هر زمانی

فصل ۸



$$= \begin{cases} \infty & \text{if } T_i(t-1) = 0 \\ \hat{\mu}_i(t-1) + \sqrt{\frac{2\log(1/\delta)}{T_i(t-1)}} & \text{otherwise.} \end{cases}$$

Input k and δ
for $t \in 1, \ldots, n$ **do**
    Choose action $A_t = \text{argmax}_i \, \text{UCB}_i(t-1, \delta)$
    Observe reward X_t and update upper confidence bounds
end for

الگوریتم هرزمانی:

2: Choose each arm once
3: Subsequently choose

$$A_t = \text{argmax}_i \left(\hat{\mu}_i(t-1) + \sqrt{\frac{2\log f(t)}{T_i(t-1)}} \right)$$

where $f(t) = 1 + t\log^2(t)$

الگوریتم هرزمانی:

2: Choose each arm once

3: Subsequently choose

$$A_t = \text{argmax}_i \left(\hat{\mu}_i(t-1) + \sqrt{\frac{2 \log f(t)}{T_i(t-1)}} \right)$$

$$\text{where } f(t) = 1 + t \log^2(t)$$

$$R_n \leq C \sum_{i=1}^{k} \Delta_i + 2\sqrt{Cnk \log(n)}.$$

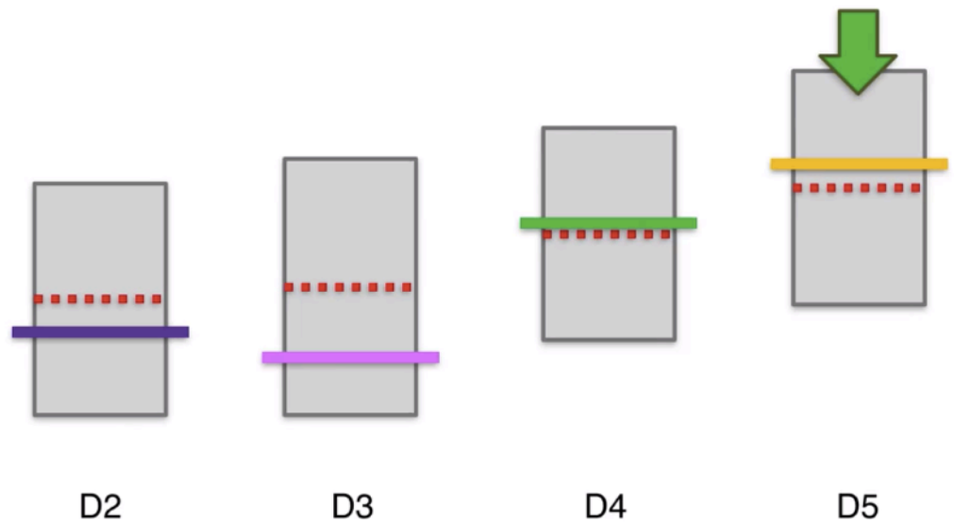
$$\limsup_{n \to \infty} \frac{R_n}{\log(n)} \leq \sum_{i:\Delta_i > 0} \frac{2}{\Delta_i}.$$

خیلی بهینه!

الگوریتم «کران بالای اطمینان» ـ الگوریتم MOSS

Minimax optimal strategy in the stochastic case

فصل ۹



الگوریتم MOSS :

2: Choose each arm once

3: Subsequently choose

$$A_t = \operatorname{argmax}_i \hat{\mu}_i(t-1) + \sqrt{\frac{4}{T_i(t-1)} \log^+ \left(\frac{n}{k T_i(t-1)} \right)}$$

where $\log^+(x) = \log \max\{1, x\}$.

الگوریتم MOSS :

2: Choose each arm once
3: Subsequently choose

$$A_t = \operatorname{argmax}_i \hat{\mu}_i(t-1) + \sqrt{\frac{4}{T_i(t-1)} \log^+ \left(\frac{n}{kT_i(t-1)} \right)}$$

where $\log^+(x) = \log \max\{1, x\}$.

$$R_n \leq 39\sqrt{kn} + \sum_{i=1}^{k} \Delta_i$$

الگوریتم MOSS :

2: Choose each arm once
3: Subsequently choose

$$A_t = \text{argmax}_i \, \hat{\mu}_i(t-1) + \sqrt{\frac{4}{T_i(t-1)} \log^+ \left(\frac{n}{k T_i(t-1)} \right)}$$

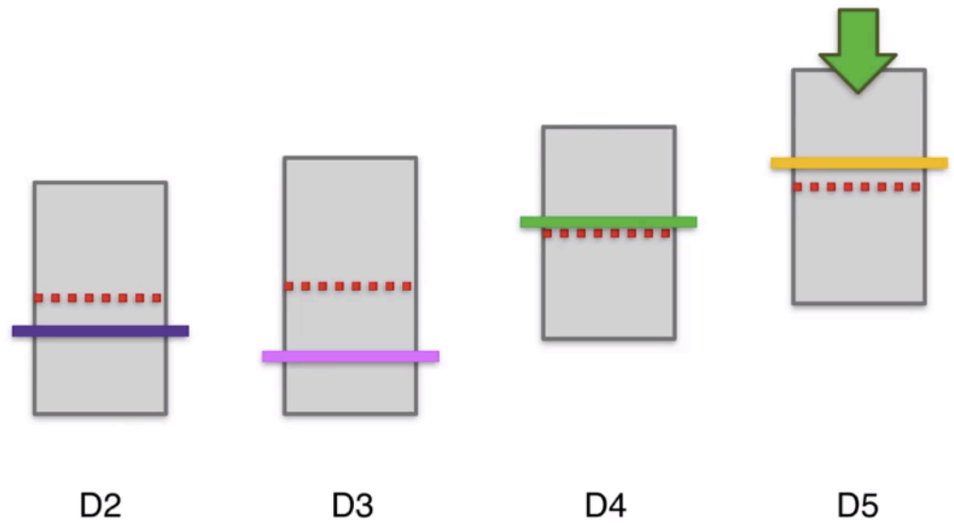
where $\log^+(x) = \log \max \{1, x\}$.

$$R_n \leq 39\sqrt{kn} + \sum_{i=1}^{k} \Delta_i$$

بهینه!

الگوریتم «کران بالای اطمینان» ـ متغیرهای برنولی

فصل ۱۰



الگوریتم KL-UCB :

2: Choose each arm once
3: Subsequently choose

$$A_t = \text{argmax}_i \max \left\{ \tilde{\mu} \in [0,1] : d(\hat{\mu}_i(t-1), \tilde{\mu}) \leq \frac{\log f(t)}{T_i(t-1)} \right\}$$

where $f(t) = 1 + t \log^2(t)$.

الگوریتم KL–UCB :

2: Choose each arm once
3: Subsequently choose

$$A_t = \text{argmax}_i \max \left\{ \tilde{\mu} \in [0,1] : d(\hat{\mu}_i(t-1), \tilde{\mu}) \leq \frac{\log f(t)}{T_i(t-1)} \right\}$$

where $f(t) = 1 + t\log^2(t)$.

$$\limsup_{n\to\infty} \frac{R_n}{\log(n)} \leq \sum_{i:\Delta_i>0} \frac{1}{2\Delta_i}$$

الگوریتم KL-UCB :

2: Choose each arm once
3: Subsequently choose

$$A_t = \text{argmax}_i \max \left\{ \tilde{\mu} \in [0,1] : d(\hat{\mu}_i(t-1), \tilde{\mu}) \leq \frac{\log f(t)}{T_i(t-1)} \right\}$$

where $f(t) = 1 + t\log^2(t)$.

$$\limsup_{n\to\infty} \frac{R_n}{\log(n)} \leq \sum_{i:\Delta_i>0} \frac{1}{2\Delta_i}$$

مقایسه با UCB(δ):

$$R_n \leq 3\sum_{i=1}^{k} \Delta_i + \sum_{i:\Delta_i>0} \frac{16\log(n)}{\Delta_i}.$$

برای برنولی‌ها:

$$\mathbb{P}\left(\hat{\mu} \geq \mu + \varepsilon\right) \leq \exp\left(-nd(\mu+\varepsilon, \mu)\right)$$

زیرگوسی‌ها:

$$\mathbb{P}\left(\hat{\mu} \geq \mu + \varepsilon\right) \leq \exp\left(-\frac{n\varepsilon^2}{2\sigma^2}\right)$$

پایان