EE6106 LECTURE17 (Dated : 27 MARCH 2024)

Lowerbound :
$$\mathbb{E}[\tau] \geq C^*(\nu)\log\left(\frac{1}{4\delta}\right)$$
$$C^*(\nu)^{-1} = \sup_{\alpha \in P_k} \inf_{\nu' \in \mathcal{E}_{alt}(\nu)} \sum \alpha_i D(\nu_i, \nu_i')$$

Insights :

$$\alpha^* \approx \frac{\mathbb{E}[N_i(\tau)]}{\mathbb{E}[\tau]} \approx \text{ratio of expected number of pulls of arm } i \text{ upto stoppage to expected stoppage} \quad (1)$$

$$\inf_{\nu' \in \mathcal{E}_{alt}(\nu)} \sum \mathbb{E}_\nu[N_i(\tau)] D(\nu_i, \nu_i') \approx \log\left(\frac{1}{\delta}\right) \quad (2)$$

The condition (1) guides "sampling"
The condition (2) guides "stopping"

TRACK & STOP

Sampling: Pull $\arg\max_i \left(\alpha_i^*(\hat{\nu})t - N_i(t)\right)$

1) Here $\hat{\nu}$ is an estimate of the instance ν. Further, this is generally possible when the family is expressed as a single-parameter family or so
2) $\alpha_i^*(\hat{\nu})t$ is like a target pull fraction
   $N_i(t)$ is the actual value
   This is the "Track" part of the algorithm
3) α^* is the value of α which solves the optimization problem
$$\sup_{\alpha \in P_k} \inf_{\nu' \in \mathcal{E}_{alt}(\nu)} \sum \alpha_i D(\nu_i, \nu_i')$$
   $\alpha^*(\hat{\nu})$ means that we feed the estimate to the optimization problem, instead of the true value.

Stopping: $\inf_{\nu' \in \mathcal{E}_{alt}(\hat{\nu})} \sum_{i=1}^{k} N_i(t) D(\hat{\nu}_i, \nu_i') \gtrsim \log\left(\frac{1}{\delta}\right)$

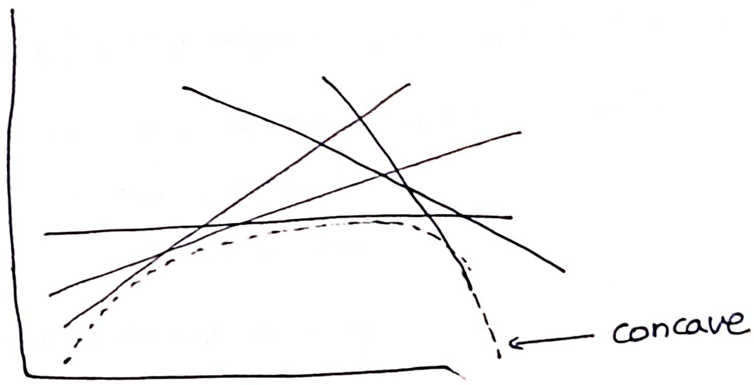
1) The space over which we are infimizing is the space of alternate solutions of the instance $\hat{\nu}$.

When can this go wrong?
1) This will work only when $\hat{\nu} \to \nu$. This is possible only when all arms get sampling enough so that convergence is possible asymptotically over time.
   Hence, this algorithm requires FORCED EXPLORATION

2) If $\delta \to 0$, then $\mathbb{E}[\tau] \to \infty$, $\hat{\nu} \to \nu$

Track and Stop was proposed by Garivier & Kaufmann (2016).

Note
1) The infimum of a family of function is conve



2) Smoothness & Differentiability is not guaranteed

3) Computation of α^* is a problem in convex optimization

Assume all arms to be Bernoulli

For arms a or b ($a \neq b$)

$$Z_{a,b}(t) = \log\left(\frac{\max_{\mu_a' \geq \mu_b'} L_{\mu_a'}(X_{t,a}) L_{\mu_b'}(X_{t,b})}{\max_{\mu_b' \geq \mu_a'} L_{\mu_a'}(X_{t,a}) L_{\mu_b'}(X_{t,b})}\right)$$

μ_a' is $\mathbb{P}$(1 on instance μ' for arm a)

$X_{t,a}$ is observation of arm a

$L_{\mu_a'}$ is the likelihood under observations of arm a

Numerator : a is better than b

Denominator : b is better than a

The statistic $Z_{a,b}(t)$ assigns the ratio of the hypotheses "(arm a is superior to arm b)" and compares it to "arm b is better than arm a"

$$L_{\mu_a'}(X_{t,a}) = (\mu_a')^{N_a^1}(1-\mu_a')^{N_a^0}$$

$Z_{a,b}(b)$ is termed a "GENERALIZED LOG-LIKELIHOOD STATISTIC"

Claim 1: For $\hat{\mu_a} \geq \hat{\mu_b}$

$$Z_{a,b}(t) = N_a(t) d(\hat{\mu_a}, \hat{\mu_{a,b}}) + N_b(t) d(\hat{\mu_b}, \hat{\mu_{a,b}})$$

where $$\hat{\mu_{a,b}} = \frac{N_a(t)}{N_a(t)+N_b(t)} \hat{\mu_a} + \frac{N_b(t)}{N_a(t)+N_b(t)} \hat{\mu_b}$$

$d(\hat{\mu_a}, \hat{\mu_{a,b}})$ is the relative entropy between the two Bernoulli dist.

$\hat{\mu_{a,b}}$ is the average number (fraction) of heads

$$= \frac{\text{total \# heads}}{\text{total \# pulls}}$$

* $Z_{a,b}(t) = -Z_{b,a}(t)$, which is trivial to prove

$$Z(t) = \max_{a} \min_{b \neq a} Z_{a,b}(t)$$

$Z(t)$ is our stoppage statistic

The outer maximization will be achieved by ONLY the empirically best arm.

$Z(t)$ hence tells us how well separated the best arm is from the closest challenging arm.

Claim 2

$$Z(t) = t \min_{\lambda \in \mathcal{E}_{alt}(\hat{\mu})} \sum \frac{N_a(t)}{t} d(\hat{\mu_a}, \lambda_a)$$

The claim is that stopping via $Z(t)$ is the same as achieving the Information-theoretic bound which we showed earlier.

It is absolutely identical to $\inf_{\nu' \in \mathcal{E}_{alt}(\hat{\nu})} \sum_{i=1}^{k} N_i(t) D(\hat{\nu_i}, \nu'_i) \gtrsim \log\left(\frac{1}{\delta}\right)$

We STOP when $Z(t) \geq \beta(t,\delta)$, a certain threshold

FORMAL DEFINITION OF TRACK & STOP

- Pull each arm once
- While $Z(t) \leq \beta(t,\delta) = \log\left(\frac{2t(k-1)}{\delta}\right)$,

  - if ~~argmax~~ $\min_i N_i(t) < \sqrt{t}$, pull $\arg\min_i N_i(t)$ ①
  - else pull $\arg\max_i \left(t d_i^*(\hat{\mu}) - N_i(t)\right)$ ②

① is FORCED EXPLORATION ② is TRACKING

- Output $i^*(\hat{\mu})$

Thm 1 : Track and Stop is δ-sound

Thm 2 : $$\limsup_{\delta \downarrow 0} \frac{\mathbb{E}[\tau]}{\log(1/\delta)} \leq C^*(\nu) \leftarrow \text{Asymptotic Optimality}$$

Note that we cannot guarantee anything about a fixed δ.

Eventually, we can also argue equality to hold in the Thm 2.

Thm 2 talks about the expected stopping time, we can also have an almost sure inequality

Thm 3 : $$\mathbb{P}_\mu\left(\left(\limsup_{\delta \downarrow 0} \frac{\tau}{\log(1/\delta)}\right) \leq C^*(\nu)\right) = 1$$