Week 8 - Lecture 37

Lemma: Chernoff's bound

Let $\{X_i\}_{i\geq 1}$ be iid sequence each distributed as $\mathrm{Ber}(\mu)$.

Let $\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} X_T$

Then $\forall\ \varepsilon \in [0, 1-\mu]$

a) $\Pr(\hat{\mu} \geq \mu+\varepsilon) \leq \exp(-T\,d(\mu+\varepsilon, \mu))$ —— 1

and $\varepsilon \in [0, \mu]$

b) $\Pr(\hat{\mu} \leq \mu-\varepsilon) \leq \exp(-T\,d(\mu-\varepsilon, \mu))$ —— 2

Applying Pinsker's inequality to 1

$\Pr(\hat{\mu} \geq \mu+\varepsilon) \leq \exp(-2T\varepsilon^2)$

This is tighter than what is obtained with Hoeffding's Inequality

---

$\Pr(\hat{\mu} \geq \mu + \varepsilon) = \Pr\left(\sum_{t=1}^{T} X_t - \mu \geq T\varepsilon\right)$

$$= \Pr\left(\exp\left(\lambda \sum_{t=1}^{T}(X_t - \mu)\right) \geq \exp(\lambda T \varepsilon)\right)$$

$[\lambda > 0]$

Applying Markov inequality

$$\leq \frac{\mathbb{E}\left[\exp\left(\lambda \sum_{t=1}^{T}(X_t - \mu)\right)\right]}{\exp(\lambda T \varepsilon)}$$

$$= \prod_{t=1}^{T} \mathbb{E}\left[\exp(\lambda(X_t - \mu))\right] \times \exp(-\lambda T \varepsilon)$$

[since X_t is iid]

[Also $X_t = 1$ w.p. μ
$X_t = 0$ w.p. $(1-\mu)$]

$$= \prod_{t=1}^{T}\left[\mu \exp(\lambda(1-\mu)) + (1-\mu)\exp(-\lambda\mu)\right] \exp(-\lambda T \varepsilon)$$

$$= \left(\mu \exp((1-\mu)\lambda - \lambda\varepsilon) + (1-\mu)\exp(-\mu\lambda - \lambda\varepsilon)\right)^T \quad (3)$$

[Since is same for all t and $-\lambda T \varepsilon$ can be considered as $-\lambda\varepsilon$ added T times]

This is thus a function of λ $\therefore$ differentiating & minimizing with λ, we get :

$$\lambda = \log \frac{(\mu+\varepsilon)(1-\mu)}{\mu(1-\mu-\varepsilon)}$$

Substituting this in ③ we get :

$$\Pr(\hat{\mu} \geq \mu+\varepsilon) \leq \Bigg[\underbrace{\frac{\mu}{\mu+\varepsilon}\left(\frac{(\mu+\varepsilon)(1-\mu)}{\mu(1-\mu-\varepsilon)}\right)^{1-\mu-\varepsilon}}_{\text{divergence}(\mu+\varepsilon,\,\mu)} \Bigg]^{T}$$

$$= \exp(-T d(\mu+\varepsilon, \mu))$$

Thus Chernoff's bound proved for (a)

Similarly can be shown for (b)

Thus KL-UCB has tighter bounds than UCB for Bernoulli distributions.

Index :

$$UCB_i = \hat{\mu} + \sqrt{\frac{\alpha \log T}{N_i(t-1)}}$$

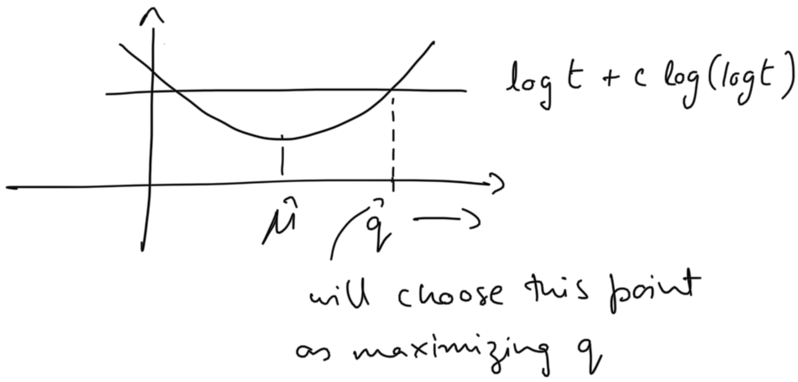
$$KL\text{-}UCB_i = \max\left\{ q \in \theta,\ N_i(t-1)\, d\left(\frac{S_i(t-1)}{N_i(t-1)}, q\right) \leq \log t + c \log(\log t) \right\}$$

For $q \geq \frac{S_i(t-1)}{N_i(t-1)}$, $d\left(\frac{S_i(t-1)}{N_i(t-1)}, q\right)$ is increasing, and as it keeps increasing at some point it will violate the inequality and that will be the index.

Thus while writing code utilize the monotonicity of the function & not solve as a convex optimization problem. Or do a search, rapidly increase q, if it crosses, bring down the step size.

Or:

Let $\frac{S_i(t-1)}{N_i(t-1)} = \hat{\mu}$, then we would have:



Thompson Sampling

[Beta distribution with parameter α, β has a distribution which is given by:

$$f(x : \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\, x^{\alpha-1}(1-x)^{\beta-1}$$
]

$\Gamma(\alpha+\beta)$ or $\Gamma(\alpha)$ is the

For each $i = 1, 2, \ldots K$, set $\underset{\text{success}}{S_i} = 0$, $\underset{\text{faliure}}{F_i} = 0$

For each $i = 1, 2, \ldots$ do.

For each $i \in [K]$ Sample $\theta_i(t)$ from $\text{Beta}(S_i+1, F_i+1)$

gamma function

assign Beta distribution to each arm with parameters (S_i+1, F_i+1)

play $I_t = \arg\max \theta_i(t)$

Observe r_t^i

if $r_t = 1$, then $S_{I_t} = S_{I_t} + 1$

else $F_{I_t} = F_{I_t} + 1$

Thompson Sampling for simple case of Bernoulli reward distribution.