Let $(X_t)_{t=1}^n$ be a sequence of indep. RVs : $E[X_t] = \mu \ \forall t = 1, 2, \ldots, n$

and $\overline{X}_t = X_t - \mu \sim \text{subG}(1)$

Let $\hat{\mu} := \frac{1}{n} \sum_{t=1}^{n} X_t \Rightarrow E[\hat{\mu}] = \frac{1}{n} \sum_{t=1}^{n} \mu = \mu$

$\therefore \hat{\mu} - \mu \sim \text{subG}\left(\frac{\sigma^2}{n}\right)$ (or) $\text{subG}\left(\frac{1}{n}\right)$

$\Rightarrow \mathbb{P}[\mu - \hat{\mu} > t] = \mathbb{P}\left[e^{s(\mu - \hat{\mu})} > e^{st}\right]$ (Markov's inequality)

$\leq e^{-st} E\left[e^{s(\mu - \hat{\mu})}\right] \leq e^{-st} e^{\frac{s^2}{2n}} \leq e^{-nt^2/2}$ ↓ Least Lower bound

Chernoff Bound → $\mathbb{P}[\mu - \hat{\mu} > t] \leq e^{-nt^2/2}$

$\Rightarrow \mathbb{P}[\mu > \hat{\mu} + t] \leq e^{-nt^2/2}$

Let $\delta := e^{-nt^2/2} \Rightarrow t = \sqrt{\frac{2 \ln(1/\delta)}{n}}$

$\Rightarrow \mathbb{P}\left[\mu > \hat{\mu} + \sqrt{\frac{2 \ln(1/\delta)}{n}}\right] \leq \delta$

before round t the learner knows the following →

For a k-armed bandit, let $T_i(t-1)$ be the number of samples from arm i and $\hat{\mu}_i(t-1) = \frac{1}{T_i(t-1)} \sum_{u \, \& \, X_u \sim P_i} X_u$, then a 'reasonable' upper bound for the mean of the i^{th} arm during round t is given by →

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

$\sqrt{\frac{2\log(1/\delta)}{T_i(t-1)}}$ → Confidence width / Exploration bonus

δ → Confidence level

$UCB_i(t-1,\delta)$ → Also called the index of arm i during round t.

UCB(δ) Algorithm →

Input k and δ

for $t \in 1, 2, \ldots, n$ do

$$A_t = \underset{i}{\operatorname{argmax}}\ UCB_i(t-1,\delta)$$

$$X_t \sim P_{A_t}$$

$$T_i(t) = T_i(t-1) \quad \forall\ i \neq A_t$$

$$\hat{\mu}_i(t) = \hat{\mu}_i(t-1) \quad \forall\ i \neq A_t$$

$$UCB_i(t,\delta) = UCB_i(t-1,\delta) \quad \forall\ i \neq A_t$$

$$T_{A_t}(t) = T_{A_t}(t-1) + 1$$

$$\hat{\mu}_{A_t}(t) = \frac{T_{A_t}(t-1)\,\hat{\mu}_{A_t}(t-1) + X_t}{T_{A_t}(t)}$$

$$UCB_{A_t}(t,\delta) = \hat{\mu}_{A_t}(t) + \sqrt{\frac{2\log(1/\delta)}{T_{A_t}(t)}}$$

end for

Initialization

$T_i(0) = 0 \quad \forall\ i \in [k]$

$\hat{\mu}_i(0) = 0 \quad \forall\ i \in [k]$

$UCB_i(0,\delta) = \inf \quad \forall\ i \in [k]$

Regret Analysis →

Theorem → For UCB(δ) algorithm, for any horizon n, if $\delta = 1/n^2$, then the expect regret/regret is bounded as follows →

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

Proof → WLOG, let arm 1 be the optimal arm, i.e., $\mu^* = \mu_1$.

Observations →

1) All arms are sequencially chosen once during the 1st period of the algorithm.

2) After 1st period, arm i is chosen $\Leftrightarrow$

   a) The 'index' of arm i is larger than μ^*.

   (or)

   b) The 'index' of the optimal arm is smaller than μ^*.

Definitions →

1) $(X_{ti})_{t\in[n], i\in[k]}$ → Collection of independent RVs | $X_{ti} \sim P_i$

2) $\hat{\mu}_{is} := \frac{1}{s}\sum_{u=1}^{s} X_{ui}$ → Emperical mean of the 1st s samples of X_{ui}.

From 1 and 2, we define 3 and 4 as follows,

3) $X_{tA_t} := X_t$

4) $\hat{\mu}_{iT_i(t)} := \hat{\mu}_i(t)$

5) G_i is called a 'good' event $\Leftrightarrow$

$$G_i = \left\{\mu_1 < \min_{t\in[n]} UCB_1(t,\delta)\right\} \cap \left\{\hat{\mu}_{iu_i} + \sqrt{\frac{2}{u_i}\log\left(\frac{1}{\delta}\right)} < \mu_1\right\}$$

where $u_i \in [n]$

claims → 1) If G_i occours, then $T_i(n) \leq u_i$

2) G_i^c occours with low probability.

(dependent on u_i)

We know that $R_n = \sum_{i=1}^{k} \Delta_i \mathbb{E}[T_i(n)]$. From the claims, we observe that,

$T_i(n) \leq n \ \forall i \in [k] \Rightarrow \mathbb{E}[T_i(n)] = \mathbb{E}[\mathbb{I}\{G_i\} T_i(n)] + \mathbb{E}[\mathbb{I}\{G_i^c\} T_i(n)]$

$\leq u_i + \mathbb{P}[G_i^c] n$ — (I)

Proof → 1) Suppose G_i occours and $T_i(n) > u_i$

$\Rightarrow \exists\, t \in [n] \mid T_i(t-1) = u_i$

$\Rightarrow UCB_i(t-1, \delta) = \hat{\mu}_{i(t-1)} + \sqrt{\frac{2\log(1/\delta)}{T_i(t-1)}}$

$= \hat{\mu}_{iu_i} + \sqrt{\frac{2\log(1/\delta)}{u_i}}$

But from G_i

$\hat{\mu}_{iu_i} + \sqrt{\frac{2\log(1/\delta)}{u_i}} < \mu_1 < UCB_1(t-1, \delta)$

$\Rightarrow UCB_i(t-1, \delta) < UCB_1(t-1, \delta)$

If $UCB_i(t'-1, \delta) < UCB_1(t'-1, \delta) \Rightarrow$

$UCB_i(t', \delta) = UCB_i(t'-1, \delta) < \mu_1 < UCB_1(t', \delta)$

$\Rightarrow UCB_i(t', \delta) < UCB_1(t', \delta)$ ($\because$ From defn. of G_i)

$\therefore$ By mathematical induction, arm i will not be chosen $\forall\, t' \in [n]$ and $t' \geq t$.

$\therefore T_i(t') = u_i \ \forall t' \in [n] \mid t' \geq t-1$ →

Hence $T_i(t) \leq u_i$ if G_1 occours.

2)

$$G_i^c = \left\{\mu_1 \geq \min_{t\in[n]} UCB_1(t,\delta)\right\} \cap \left\{\hat{\mu}_{iu_i} + \sqrt{\frac{2}{u_i}\log\left(\frac{1}{\delta}\right)} > \mu_1\right\}$$

$$\Rightarrow \mathbb{P}[G_i^c] \leq \mathbb{P}\left[\mu_1 \geq \min_{t\in[n]} UCB_1(t,\delta)\right] + \mathbb{P}\left[\hat{\mu}_{iu_i} + \sqrt{\frac{2}{u_i}\log\left(\frac{1}{\delta}\right)} > \mu_1\right]$$

(1) (2)

(1) $\left\{\mu_1 \geq \min_{t\in[n]} UCB_1(t,\delta)\right\}$

$$= \bigcup_{t\in[n]} \left\{\mu_1 \geq UCB_1(t,\delta)\right\}$$

$$\therefore \mathbb{P}\left[\mu_1 \geq \min_{t\in[n]} UCB_1(t,\delta)\right] \leq \sum_{t=1}^{n} \mathbb{P}\left[\mu_1 \geq UCB_1(t,\delta)\right] \leq \sum_{t=1}^{n} \delta$$

$$\leq n\delta$$

(∵ From defn. of UCB)

(2) $\hat{\mu}_{iu_i} + \sqrt{\frac{2}{u_i}\log\left(\frac{1}{\delta}\right)} \geq \mu_1$

$$\Rightarrow \hat{\mu}_{iu_i} - \mu_i \geq \Delta_i - \sqrt{\frac{2}{u_i}\log\left(\frac{1}{\delta}\right)}$$

Assuming that u_i is large enough:

$$\Delta_i - \sqrt{\frac{2}{u_i}\log\left(\frac{1}{\delta}\right)} \geq c\Delta_i \text{ for some } c \in (0,1) \quad \text{(3)}$$

$$\Rightarrow \hat{\mu}_{iu_i} - \mu_i \geq c\Delta_i$$

$$\Rightarrow \mathbb{P}\left[\hat{\mu}_{iu_i} - \mu_i \geq c\Delta_i\right] \leq \exp\left(-\frac{u_i c^2 \Delta_i^2}{2}\right)$$

(∵ From Chernoff bound)

$$\therefore \mathbb{P}[G_i^c] \leq n\delta + \exp\left(\frac{-u_i c^2 \Delta_i^2}{2}\right)$$

We choose u_i to be the smallest integer: (3)

holds →

$$\Delta_i - \sqrt{\frac{2\log(1/\delta)}{u_i}} = c\Delta_i$$

(→ ceil)

$$\Rightarrow (1-c)^2\Delta_i^2 = \frac{2\log(1/\delta)}{u_i} \Rightarrow u_i = \left\lceil \frac{2\log(1/\delta)}{(1-c)^2\Delta_i^2} \right\rceil$$

Substituting in (T), we get →

$$E[T_i(n)] \leq u_i + n\left(n\delta + \exp\left(\frac{-u_i c^2 \Delta_i^2}{2}\right)\right)$$

$$\leq \left\lceil \frac{4\log(n)}{(1-c)^2\Delta_i^2} \right\rceil + 1 + n^{1 - 2c^2/(1-c)^2}$$

If c is close to 1, u_i blows up and c is close to 0, then we get linear regret. So we choose $c = 1/2$

$$\therefore E[T_i(n)] \leq \left\lceil \frac{16\log(n)}{\Delta_i^2} \right\rceil + 1 + \frac{1}{n}$$

as $\frac{1}{n} \leq 1$ and $\left\lceil \frac{16\log(n)}{\Delta_i^2} \right\rceil \leq \frac{16\log(n)}{\Delta_i^2} + 1$

$$\Rightarrow E[T_i(n)] \leq \frac{16\log(n)}{\Delta_i^2} + 3$$

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

Hence Proved

Theorem → If $\delta = 1/n^2$ on any, then the regret of UCB(δ) on any $v \in \mathcal{E}^k_{sg}(1)$ is bounded by →

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

(Sublinear → $O(\sqrt{nk\log(n)})$)

Proof →

$$R_n = \sum_{i=1}^{k}\Delta_i E[T_i(n)]$$

$$= \sum_{i|\Delta_i<\Delta}\Delta_i E[T_i(n)] + \sum_{i|\Delta_i\geq\Delta}\Delta_i E[T_i(n)]$$

① ②

for some Δ.

① $$\sum_{i|\Delta_i<\Delta}\Delta_i E[T_i(n)] = \sum_{i|\Delta_i<\Delta}\sum_{t=1}^{n}\mathbb{I}\{A_t=i\}\Delta_i$$

$$< \Delta\sum_{t=1}^{n}\mathbb{I}\{A_t=i\} \leq n\Delta$$

($\because \Delta_i < \Delta$ and $\sum_{t=1}^{n}\mathbb{I}\{A_t=i\} \leq n$)

② $$\sum_{i|\Delta_i\geq\Delta}\Delta_i E[T_i(n)] \leq \sum_{i|\Delta_i\geq\Delta}\Delta_i\left(3 + \frac{16\log(n)}{\Delta_i^2}\right)$$

$$\leq 3\sum_{i=1}^{k}\Delta_i + \frac{16k\log(n)}{\Delta}$$

$$\therefore R_n \leq n\Delta + \frac{16k\log(n)}{\Delta} + 3\sum_{i=1}^{k}\Delta_i$$

Choosing $\Delta = \sqrt{\frac{16k\log(n)}{n}}$,

$$R_n \leq 4\sqrt{nk\log(n)} + 4\sqrt{nk\log(n)} + 3\sum_{i=1}^{n} \Delta_i$$

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

Hence Proved.