UCB over arbitrary time horizons.

Source: Chap 8, Bandit Algorithms, L. & S, 2019.

So far: UCB(δ) with $\delta = \frac{1}{n^2}$, regret bounds.

→ We assumed knowledge of time horizon n, and set $\delta = 1/n^2$.

→ $\{\Delta_i\}$ are unknown.

UCB with knowledge of n:

Prev: $(1/\delta) = n^2$.

Instead: $f(t) = 1 + t \ln^2(t)$: Time varying parameter.

i.e.,

$$A_t = \underset{1 \le j \le k}{\operatorname{argmax}} \; \hat{\mu}_j(t-1) + \sqrt{\frac{2 \ln f(t)}{T_j(t-1)}}$$

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

Note: As t increases $f(t)$ grows slightly faster than linear in time.

Setting: As before with k arms, unstructured environment, 1-subGaussian noise in rewards.

Thm (8.1 in text):

$$R_n \leq \sum_{\{j:\Delta_j>0\}} R_{n,j}$$

$$R_{n,j} = \inf_{\varepsilon \in (0,\Delta_j)} \Delta_j \left(1 + \frac{5}{\varepsilon_j^2} + \frac{2\left(\ln(f(n)) + \sqrt{\pi \ln(f(n)) + 1}\right)}{(\Delta_i - \varepsilon_j)^2} \right)$$

$$\overline{\lim_{n\to\infty}} \frac{R_n}{\ln(n)} \leq \sum_{\{j:\Delta_j>0\}} \left(\frac{2}{\Delta_j}\right)$$

If $\varepsilon_j = \frac{\Delta_j}{2}$,

$$R_n \leq \sum_{\{j:\Delta_j>0\}} \left(\Delta_j + \frac{8\left(\ln f(n) + \sqrt{\pi \ln f(n)}\right) + 28}{\Delta_j} \right)$$

Recall $f(n) = 1 + n \ln^2(n)$

$$\therefore \quad R_n \leq C \sum_{\{j : \Delta_j > 0\}} \left(\Delta_j + \frac{\ln(n)}{\Delta_j} \right)$$

C: fixed constant.

Aside: We first need the following bound.

Counting the expected number of sub-optimal arm plays:

$\{X_i, i = 1, 2, \cdots, n\}$ iid, with

$E[X_i] = 0$, $X_i \sim 1$-subGaussian.

$$\hat{\mu}_t = \frac{1}{t} \sum_{s=1}^{t} X_s .$$

$\varepsilon > 0$, $a > 0$ fixed parameters.

$$C_n = \sum_{t=1}^{n} \chi\left\{\hat{\mu}_t + \sqrt{\frac{2a}{t}} \geq \varepsilon\right\}$$

interpretation of $U_j(\cdot)$ with parameter $\ln(f(t)) = a$ and at some time such that t samples has been acquired by that time, and this index function exceeds the true mean by $\varepsilon > 0$.

Also note that for $u = (2a/\varepsilon^2)$, we have:

$$C_n \leq C_n' \equiv u + \sum_{t=\lceil u \rceil}^{n} \chi\left\{\hat{\mu}_t + \sqrt{\frac{2a}{t}} \geq \varepsilon\right\}$$

Claim (Lemma 8.2): Setting as above. Then,

$$E[C_n] \leq E[C_n']$$

$$\leq 1 + \frac{2}{\varepsilon^2}\left(a + \sqrt{\pi a} + 1\right).$$

Proof:

$$E[C_n] \le E[C_n']$$

$$= u + \sum_{t=\lceil u \rceil}^{n} P\left(\hat{\mu}_t + \sqrt{\frac{2a}{t}} \ge \varepsilon\right)$$

$\frac{2a}{\varepsilon^2}$ → $1 - u \ge G$

$$\le u + \sum_{t=\lceil u \rceil}^{n} e^{-\frac{t}{2}\left(\varepsilon - \sqrt{\frac{2a}{t}}\right)^2}$$

$$\le (1+u) + \int_{u}^{\infty} e^{-\frac{t}{2}\left(\varepsilon - \sqrt{\frac{2a}{t}}\right)^2} dt$$

Aside:

$x = \varepsilon\sqrt{t} - \sqrt{2a}$

$$= \int_{0}^{\infty} e^{-x^2/2} \cdot 2\left(\frac{x + \sqrt{2a}}{\varepsilon}\right) dx$$

Use Gaussian Integral formula: (Ref: Wikipedia on Gaussian Integrals)

$$\int_{0}^{\infty} x^{2n+1} e^{-ax^2} dx = \frac{n!}{2a^{n+1}}$$

$$\le 1 + \frac{2}{\varepsilon^2}\left(a + \sqrt{\pi a} + 1\right)$$

∎

Back to Main Regret Result:

Proof of Thm:

$$R_n = \sum_{\{j:\Delta_j>0\}} \Delta_j \, E[T_j(n)]. \qquad \text{(Regret Decomp.)}$$

Pick any arm $j \in \{1,\dots,k\}$ with $\Delta_j > 0$.

Suppose at time t, $\{A_t = j\}$. There are four possibilities:

(a) $U_1(t-1) \le \mu_1 - \varepsilon$, $U_j(t-1) < \mu_1 - \varepsilon$

(b) $U_1(t-1) \le \mu_1 - \varepsilon$, $U_j(t-1) \ge \mu_1 - \varepsilon$

~~(c)~~ $U_1(t-1) > \mu_1 - \varepsilon$, $U_j(t-1) < \mu_1 - \varepsilon$ — cannot occur

(d) $U_1(t-1) > \mu_1 - \varepsilon$, $U_j(t-1) \ge \mu_1 - \varepsilon$

$$\therefore \{A_t = j\} \subseteq \{A_t = j,\ U_1(t-1) \le \mu_1 - \varepsilon\} \cup \{A_t = j,\ U_j(t-1) \ge \mu_1 - \varepsilon\}$$

(covers (a), (b)) (covers (d))

$$\subseteq \left\{ U_1(t-1) \le \mu_1 - \varepsilon \right\} \cup \left\{ U_j(t-1) \ge \mu_1 - \varepsilon, A_t = j \right\}$$

$$\therefore T_j(n) = \sum_{t=1}^{n} \chi_{\{A_t = j\}}$$

$$\le \sum_{t=1}^{n} \chi_{\left\{ U_1(t-1) \le \mu_1 - \varepsilon \right\}}$$

arm 1 is atypically small

(I)

$$+ \sum_{t=1}^{n} \chi_{\left\{ U_j(t-1) \ge \mu_1 - \varepsilon, A_t = j \right\}}$$

arm j is atypically large whenever played.

(II)

$$\text{(I)} : \quad E\left[\sum_{t=1}^{n} \chi_{\left\{ U_1(t-1) \le \mu_1 - \varepsilon \right\}} \right]$$

$$= E\left[\sum_{t=1}^{n} \chi_{\left\{ \hat{\mu}_1(t-1) + \sqrt{\frac{2 \ln f(t)}{T_1(t-1)}} \le \mu_1 - \varepsilon \right\}} \right]$$

$T_1(t-1)$: # of samples $\in \{1, 2, \dots, n\}$ (say r)

union bound (over r) + linearity of expectation (over t)

$$\leq \sum_{t=1}^{n} \sum_{r=1}^{n} P\left(\hat{\mu}_{1r} + \sqrt{\frac{2 \ln f(t)}{r}} \leq \mu_1 - \varepsilon \right)$$

t: time, r: samples

1-subG conc.

$$\leq \sum_{t=1}^{n} \sum_{r=1}^{n} e^{-\left(r \left(\sqrt{\frac{2 \ln f(t)}{r}} + \varepsilon \right)^2 / 2 \right)}$$

check these yourself

$$\leq \sum_{t=1}^{n} \frac{1}{f(t)} \sum_{r=1}^{n} e^{-r\varepsilon^2/2}$$

$$\leq \frac{5}{\varepsilon^2}$$

(II): $$E\left[\sum_{t=1}^{n} \chi_{\{ U_j(t-1) \geq \mu_1 - \varepsilon \, , \, A_t = j \}} \right]$$

$$= E\left[\sum_{t=1}^{n} \chi_{\{ \hat{\mu}_j(t-1) + \sqrt{\frac{2 \ln f(t)}{T_j(t-1)}} \geq \mu_1 - \varepsilon, \, A_t = j \}} \right]$$

$$\leq E\left[\sum_{t=1}^{n} \chi_{\{ \hat{\mu}_j(t-1) + \sqrt{\frac{2 \ln f(n)}{T_j(t-1)}} \geq \mu_1 - \varepsilon, \, A_t = j \}} \right]$$

over time

Prev. in part (I), we used a union bound to deal with $T_j(t-1) \in \{1, 2, \cdots, n\}$. Here, we have more information. We observe that the indicator function $\chi_{\{\cdots, A_t = j\}}$ equals 1 only when $A_t = j$. But this INCREMENTS $T_j(t) \leftarrow T_j(t-1) + 1$.

Thus, $T_j(t-1)$ can equal $r \in \{1, 2, \cdots, n\}$ at most once over $t \in \{1, 2, \cdots, n\}$.

$$\leq E\left[\sum_{r=1}^{n} \chi_{\left\{\hat{\mu}_{jr} + \sqrt{\frac{2 \ln f(n)}{r}} \geq \mu_1 - \varepsilon\right\}}\right]$$

note that this is over samples

$$= \sum_{r=1}^{n} P\left(\hat{\mu}_{jr} - \mu_j + \sqrt{\frac{2 \ln f(n)}{r}} \geq \Delta_j - \varepsilon\right)$$

$$\leq 1 + \frac{2}{(\Delta_j - \varepsilon)^2}\left(\ln f(n) + \sqrt{\pi \ln f(n) + 1}\right)$$

Substitute in Regret Decomp to get

the main result. Choose $\varepsilon = \frac{1}{(\ln(n))^{1/4}}$

and let $n \to \infty$ to get the asymptotic upper bound.

$\blacksquare$

Note: Please read Sec 8.2 for discussion on other choices of $f(t)$.