IE613: Online Learning - Assignment 2

Subhadeep Chaudhuri | 19i190010

Question 1

Suppose that X is σ-subgaussian and X_1 and X_2 are independent and σ_1 and σ_2-subgaussian respectively, then:

1. $\mathbb{E}[X] = 0$ and $\mathrm{Var}(X) \leq \sigma^2$
2. cX is $|c|\sigma$-subgaussian for all $c \in \mathbb{R}$
3. $X_1 + X_2$ is $\sqrt{\sigma_1^2 + \sigma_2^2}$-subgaussian

Answer: (1) Given that X is σ-subgaussian, we have from the definition of a σ-subgaussian random variable

$$
\begin{aligned}
&\mathbb{E}[e^{\lambda X}] \leq e^{\frac{\lambda^2\sigma^2}{2}} \\
\implies &\mathbb{E}\Big[\sum_{i=0}^{\infty}(\frac{\lambda X)^i}{i!}\Big] \leq \Big[\sum_{i=0}^{\infty}\Big(\frac{\lambda^2\sigma^2}{2}\Big)^i\frac{1}{i!}\Big] \quad \text{[Using the Taylor series expansion of } e^x\text{]} \\
\implies &\lambda\mathbb{E}[X] + \frac{\lambda^2}{2}\mathbb{E}[X^2] \leq \frac{\lambda^2\sigma^2}{2} + o(\lambda^2) \quad \text{as } \lambda \to 0
\end{aligned}
$$

Let us now consider the following 2 cases:

- Case 1: $\lambda > 0$ We then have,

$$
\begin{aligned}
\mathbb{E}[X] + \frac{\lambda}{2}\mathbb{E}[X^2] &\leq \frac{\lambda\sigma^2}{2} + o(\lambda) \\
\text{or, } \mathbb{E}[X] &\leq 0 \text{ as } \lambda \to 0
\end{aligned}
$$

- Case 2: $\lambda < 0$ We then have,

$$
\begin{aligned}
\mathbb{E}[X] + \frac{\lambda}{2}\mathbb{E}[X^2] &\geq \frac{\lambda\sigma^2}{2} + o(\lambda) \\
\text{or, } \mathbb{E}[X] &\geq 0 \text{ as } \lambda \to 0
\end{aligned}
$$

Combining the above 2 inequalities, we have $\mathbb{E}[X] = 0$ **(Proved)**

Next, we show that $\mathrm{Var}(X) \leq \sigma^2$.
We have, $\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \mathbb{E}[X^2]$ [Since $\mathbb{E}[X] = 0$]
We had already derived the following:

$$\lambda\mathbb{E}[X] + \frac{\lambda^2}{2}\mathbb{E}[X^2] \leq \frac{\lambda^2\sigma^2}{2} + o(\lambda^2) \quad \text{as } \lambda \to 0$$

In the last part, we showed that $\mathbb{E}[X] = 0$. Thus the above limiting inequality becomes:

$$\frac{\lambda^2}{2}\mathbb{E}[X^2] \leq \frac{\lambda^2\sigma^2}{2} + o(\lambda^2) \quad \text{as } \lambda \to 0$$

Dividing throughout by $\frac{\lambda^2}{2}$ and letting $\lambda \to 0$, we then have,

$$
\begin{aligned}
&\mathbb{E}[X^2] \leq \sigma^2 \\
\implies &\mathrm{Var}(X) \leq \sigma^2 \qquad \textbf{(Proved)}
\end{aligned}
$$

---

(2) We need to show that cX is $|c|\sigma$-subgaussian for all $c \in \mathbb{R}$.
Let $Y = cX$. We then have,

$$\begin{aligned}\mathbb{E}[e^{\lambda Y}] = \mathbb{E}[e^{\lambda cX}] &\leq e^{\frac{(\lambda c)^2\sigma^2}{2}} \\ &= e^{\frac{\lambda^2(|c|\sigma)^2}{2}}\end{aligned}$$

We also have $\mathbb{E}[cX] = c\mathbb{E}[X] = 0$ and $\mathrm{Var}(cX) = c^2\mathrm{Var}(X) \leq c^2\sigma^2 = (|c|\sigma)^2$.
Thus, $cX \sim |c|\sigma$-subgaussian $\forall c \in \mathbb{R}$. **(Proved)**

---

(3) We need to show that $X_1 + X_2$ is $\sqrt{\sigma_1^2 + \sigma_2^2}$-subgaussian.
Let $Y = X_1 + X_2$. Then,

$$\begin{aligned}\mathbb{E}[e^{\lambda Y}] &= \mathbb{E}[e^{\lambda(X_1+X_2)}] \\ &= \mathbb{E}[e^{\lambda X_1}]\ \mathbb{E}[e^{\lambda X_2}] \quad \text{[Since } X_1 \text{ and } X_2 \text{ are independent]} \\ &\leq e^{\frac{\lambda^2\sigma_1^2}{2}} e^{\frac{\lambda^2\sigma_2^2}{2}} \qquad \text{[Since } X_1 \text{ and } X_2 \text{ are subgaussian random variables]} \\ &= e^{\frac{\lambda^2}{2}(\sigma_1^2+\sigma_2^2)}\end{aligned}$$

We also have,

$$\begin{aligned}\mathbb{E}[Y] &= \mathbb{E}[X_1] + \mathbb{E}[X_2] = 0 \\ \text{and } Var(Y) &= Var(X_1) + Var(X_2) + 2Cov(X_1, X_2) \\ &\leq \sigma_1^2 + \sigma_2^2 \quad \text{[Since } X_1 \text{ and } X_2 \text{ are independent } \implies \mathrm{Cov}(X_1, X_2) = 0]\end{aligned}$$

$\implies Y = X_1 + X_2 \sim \sqrt{\sigma_1^2 + \sigma_2^2}$ - subgaussian **(Proved)**

Question 2

Suppose that X is zero-mean and $X \in [a, b]$ almost surely for constants $a < b$.

1. Show that X is $\frac{(b-a)}{2}$ - subgaussian.
2. Using Cramer-Chernoff method show that if $X_1, X_2, \ldots, X_n$ are independent and $X_t \in [a_t, b_t]$ almost surely with $a_t < b_t$ for all t, then

$$\mathbb{P}\Big(\sum_{t=1}^{n}(X_t - \mathbb{E}[X_t]) \geq \epsilon\Big) \leq exp\Big(-\frac{2\epsilon^2}{\sum_{t=1}^{n}(b_t - a_t)^2}\Big)$$

Answer: (1) We have, $\mathbb{E}[X] = 0$ and $X \in [a, b]$ almost surely for constants $a < b$.
To show that X is $\frac{(b-a)}{2}$ - subgaussian, we will show that $\mathbb{E}[e^{tX}] \leq exp\Big(\frac{t^2}{2}(\frac{b-a}{2})^2\Big)$
Let us define the cumulant-generating function as follows: $\psi(t) = log(\mathbb{E}(e^{tX}))$
Then, we have

$$\begin{aligned}\psi'(t) &= \frac{\mathbb{E}(Xe^{tX})}{\mathbb{E}(e^{tX})} \\ \text{and } \psi''(t) &= \frac{\mathbb{E}(X^2e^{tX})}{\mathbb{E}(e^{tX})} - \Big[\frac{\mathbb{E}(Xe^{tX})}{\mathbb{E}(e^{tX})}\Big]^2\end{aligned}$$

We see from the expression of $\psi(t)$ that it can be interpreted as the variance of the random variable X under the probability measure $dQ = \frac{e^{tX}}{\mathbb{E}(e^{tX})}dP$. Since we know $X \in [a, b]$ a.s., thus we have under any probability measure

$$\mathrm{Var}(X) = \mathrm{Var}\Big(X - \frac{a+b}{2}\Big) \leq \mathbb{E}\Big[\Big(X - \frac{a+b}{2}\Big)^2\Big] \leq \frac{(b-a)^2}{4} \qquad \text{[By Popoviciu's inequality on variances]}$$

Using the fundamental theorem of calculus and integrating $\psi''(t)$, we get

$$\psi(t) = \int_0^t \int_0^\mu \psi''(p)dp\ d\mu \leq \int_0^t \int_0^\mu \Big(\frac{b-a}{2}\Big)^2 dp\ d\mu \leq \frac{t^2}{2}\Big(\frac{b-a}{2}\Big)^2$$
$$\implies \mathbb{E}(e^{tX}) \leq exp\Big(\frac{t^2}{2}\Big(\frac{b-a}{2}\Big)^2\Big)$$
$$\implies X \sim \frac{(b-a)}{2} \text{ - subgaussian} \qquad \textbf{(Proved)}$$

---

(2) We are given that $X_t \in [a_t, b_t]$ almost surely with $a_t < b_t$. We can then say that $\Big(X_t - \mathbb{E}(X_t)\Big) \in \big[a_t - \mathbb{E}(X_t), b_t - \mathbb{E}(X_t)\big]$ almost surely. We also see that $\mathbb{E}\big[X_t - \mathbb{E}(X_t)\big] = 0$. Thus, from the result proved in the last part, we can say that $\big[X_t - \mathbb{E}(X_t)\big]$ is $\frac{\big(b_t - \mathbb{E}(X_t)) - (a_t - \mathbb{E}(X_t)\big)}{2} \equiv \frac{(b_t - a_t)}{2}$ - subgaussian. We use the Cramer-Chernoff method to get the following:

$$\mathbb{P}\Big(\sum_{t=1}^{n}(X_t - \mathbb{E}[X_t]) \geq \epsilon\Big) \leq \mathbb{E}\Big[exp\Big(\lambda \sum_{t=1}^{n}(X_t - \mathbb{E}[X_t])\Big)\Big]e^{-\lambda\epsilon} \qquad \text{for } \lambda > 0$$
$$= \Big(\prod_{t=1}^{n} \mathbb{E}[e^{\lambda(X_t - \mathbb{E}[X_t])}]\Big)e^{-\lambda\epsilon}$$
$$\leq \Big(\prod_{t=1}^{n} e^{\frac{\lambda^2(b_t-a_t)^2}{8}}\Big)e^{-\lambda\epsilon} \qquad \text{[By Hoeffding's lemma proved in the last part]}$$
$$= \Big(\prod_{t=1}^{n} exp\Big[\frac{\lambda^2(b_t-a_t)^2}{8} - \frac{\lambda\epsilon}{n}\Big]\Big)$$

We now try to minimize the term in the power of the exponential term written above w.r.t. λ (since e^x is a convex function in x). Differentiating w.r.t. λ and equating to 0, we get:

$$\frac{\lambda(b_t-a_t)^2}{4} - \frac{\epsilon}{n} = 0$$
$$\implies \lambda = \frac{4\epsilon}{n(b_t-a_t)^2}$$

Plugging this value of λ to the last inequality obtained, we get

$$\Big(\prod_{t=1}^{n} exp\Big[\frac{\lambda^2(b_t-a_t)^2}{8} - \frac{\lambda\epsilon}{n}\Big]\Big) \leq \prod_{t=1}^{n} exp\Big(-\frac{2\epsilon^2}{n^2(b_t-a_t)^2}\Big)$$
$$= exp\Big(-2\epsilon^2 \sum_{t=1}^{n}\frac{1}{n^2(b_t-a_t)^2}\Big)$$
$$\leq exp\Big(-2\epsilon^2 \frac{1}{\sum_{t=1}^{n}(b_t-a_t)^2}\Big)$$

The last inequality is obtained by applying the inequality AM > HM on the terms $(b_t - a_t)\ \forall t = 1(1)n$. We got the inequality as follows:

$$\frac{\sum_{t=1}^{n}(b_t-a_t)^2}{n} \geq \frac{n}{\sum_{t=1}^{n}\frac{1}{(b_t-a_t)^2}}$$
$$\implies \frac{1}{\sum_{t=1}^{n}(b_t-a_t)^2} \leq \sum_{t=1}^{n}\frac{1}{n^2(b_t-a_t)^2}$$
$$\implies \frac{-2\epsilon^2}{\sum_{t=1}^{n}(b_t-a_t)^2} \geq \sum_{t=1}^{n}\frac{-2\epsilon^2}{n^2(b_t-a_t)^2}$$

Thus, we have,

$$\mathbb{P}\Big(\sum_{t=1}^{n}(X_t - \mathbb{E}[X_t]) \geq \epsilon\Big) \leq exp\Big(-\frac{2\epsilon^2}{\sum_{t=1}^{n}(b_t-a_t)^2}\Big) \qquad \textbf{(Proved)}$$

Question 3 [Expectation of maximum]

Let $X_1, X_2, \ldots, X_n$ be a sequence of σ-subgaussian random variables (possibly dependent) and $Z = max_{t\in[n]} X_t$. Prove that

1. $\mathbb{E}[Z] \leq \sqrt{2\sigma^2 log(n)}$
2. $\mathbb{P}\big(Z \geq \sqrt{2\sigma^2 log(n/\delta)}\big) \leq \delta$ for any $\delta \in (0,1)$

Answer: (1) Let $\lambda > 0$. Then we have,

$$\begin{aligned} exp(\lambda\mathbb{E}(Z)) &\leq \mathbb{E}(exp(\lambda Z)) \quad \text{[By Jensen's inequality]} \\ &\leq \sum_{t=1}^{n} \mathbb{E}(exp(\lambda X_t)) \\ &\leq nexp(\frac{\lambda^2\sigma^2}{2}) \quad \text{[From the definition of subgaussian r.v.]} \\ \implies \lambda\mathbb{E}(Z) &\leq log(n) + \frac{\lambda^2\sigma^2}{2} \\ \implies \mathbb{E}(Z) &\leq \frac{log(n)}{\lambda} + \frac{\lambda\sigma^2}{2} \end{aligned}$$

Let us choose $\lambda = \frac{1}{\sigma}\sqrt{2log(n)}$. Plugging thus choice of λ in the above inequality, we get:

$$\begin{aligned} \mathbb{E}(Z) &\leq \frac{log(n)\ \sigma}{\sqrt{2log(n)}} + \frac{\sqrt{2log(n)}\ \sigma}{2} \\ &= \sqrt{2\sigma^2 log(n)} \qquad \textbf{(Proved)} \end{aligned}$$

---

(2) We observe that

$$P\Big(Z \geq \sqrt{2\sigma^2 log(\frac{n}{\delta})}\Big) \leq P\Big(\cup_{i=1}^{n} A_i\Big) \quad \text{where } A_i = \Big\{X_i \geq \sqrt{2\sigma^2 log(\frac{n}{\delta})}\Big\}$$

Now, using Boole's inequality, we can write:

$$P\Big(Z \geq \sqrt{2\sigma^2 log(\frac{n}{\delta})}\Big) \leq P\Big(\cup_{i=1}^{n} A_i\Big) \leq \sum_{i=1}^{n} P(A_i)$$

Now using Cramer-Chernoff method, we can show that

$$P(A_i) = P\Big(X_i \geq \sqrt{2\sigma^2 log(\frac{n}{\delta})}\Big) \leq \frac{\delta}{n}$$

Using the above probability bound, we see:

$$P\Big(Z \geq \sqrt{2\sigma^2 log(\frac{n}{\delta})}\Big) \leq \sum_{i=1}^{n} P(A_i) \leq \sum_{i=1}^{n} \frac{\delta}{n} = \delta \qquad \textbf{(Proved)}$$

Question 4 [Bernstein's inequality]

Let $X_1, X_2, \ldots, X_n$ be a sequence of independent random variables with $X_t - \mathbb{E}[X_t] \leq b$ almost surely and $S = \sum_{t=1}^{n}(X_t - \mathbb{E}(X_t))$ and $v = \sum_{t=1}^{n} V[X_t]$

1. Show that $g(x) = \frac{1}{2} + \frac{x}{3!} + \frac{x^2}{4!} + \cdots = \frac{(exp(x)-1-x)}{x^2}$ is increasing.
2. Let X be a random variable with $\mathbb{E}[X] = 0$ and $X \leq b$ almost surely. Show that $\mathbb{E}[exp(X)] \leq 1 + g(b)V[X]$.
3. Prove that $(1+\alpha)log(1+\alpha) - \alpha \geq \frac{3\alpha^2}{6+2\alpha}$ for all $\alpha \geq 0$. Prove that this is the best possible approximation in the sense that the 2 in the denominator cannot be increased.
4. Let $\epsilon > 0$ and $\alpha = \frac{b\epsilon}{v}$, prove that

$$\begin{aligned}\mathbb{P}\Big(S \geq \epsilon\Big) &\leq exp\Big(-\frac{v}{b^2}((1+\alpha)log(1+\alpha) - \alpha)\Big)\\ &\leq exp\Big(-\frac{\epsilon^2}{2v(1+\frac{b\epsilon}{3v})}\Big)\end{aligned}$$

5. Use the previous result to show that

$$\mathbb{P}\Big(S \geq \sqrt{2vlog(\frac{1}{\delta})} + \frac{2b}{3}log(\frac{1}{\delta})\Big) \leq \delta$$

Answer: (1) We have $g(x) = \frac{(exp(x)-1-x)}{x^2}$. Differentiating w.r.t x, we have

$$\begin{aligned} g'(x) &= \frac{x^2(exp(x)-1) - 2x(exp(x)-1-x)}{x^4}\\ \implies x^3 g'(x) &= h(x) = xe^x - 2e^x + 2 + x\end{aligned}$$

We have $h'(x) = xe^x - e^x + 1$ and $h''(x) = xe^x$. We observe that h' is increasing on $(0,\infty)$ and decreasing on $(-\infty, 0)$. Also, $h(0) = 0$.
So we see that sign of $h(x)$ varies similarly with that of x i.e. $sign(h(x)) = sign(x) = sign(x^3)$. Since we already saw $x^3 g'(x) = h(x)$, thus we can conclude that $g'(x) > 0$ i.e $g(x)$ is increasing. **(Proved)**

---

(2) We had $g(x) = \frac{(exp(x)-1-x)}{x^2} \implies exp(x) = 1 + x + g(x)x^2$. Therefore,

$$\begin{aligned}\mathbb{E}(exp(X)) &= 1 + \mathbb{E}(X) + \mathbb{E}(g(X)X^2)\\ &\leq 1 + \mathbb{E}(g(X)X^2) \qquad [\text{Since } \mathbb{E}(X) = 0]\\ &\leq 1 + \mathbb{E}(g(b)X^2) \qquad [\text{Since } x \leq b \;\; a.s. \implies \mathbb{E}(X) \leq \mathbb{E}(b)]\\ &\leq 1 + g(b)\mathbb{E}(X^2)\\ &\leq 1 + g(b)V(X) \qquad [\text{Since } (\mathbb{E}(X))^2 = 0] \qquad \textbf{(Proved)}\end{aligned}$$

---

(3) Let us assume $f(\alpha) = (1+\alpha)\log(1+\alpha)\alpha$. Let us try to derive the required inequality.

$$\begin{aligned} f(\alpha) &= (1+\alpha)\log(1+\alpha)\alpha\\ &= (1+\alpha)\sum_{n=1}^{\infty}\frac{(-1)^{n-1}\alpha^n}{n}\alpha\\ &= \sum_{n=1}^{\infty}\frac{(-1)^{n-1}\alpha^n}{n} + \alpha\sum_{n=1}^{\infty}\frac{(-1)^{n-1}\alpha^n}{n}\alpha\\ &= \sum_{n=1}^{\infty}\frac{(-1)^{n-1}\alpha^n}{n} + \sum_{n=1}^{\infty}\frac{(-1)^{n-1}\alpha^{n+1}}{n}\alpha\\ &= \sum_{n=1}^{\infty}\frac{(-1)^{n-1}\alpha^n}{n} + \sum_{n=2}^{\infty}\frac{(-1)^{n}\alpha^n}{n-1}\alpha\end{aligned}$$

$$
\begin{aligned}
&= \alpha + \sum_{n=2}^{\infty} \frac{(-1)^{n-1}\alpha^n}{n} + \sum_{n=2}^{\infty} \frac{(-1)^n \alpha^n}{n-1}\alpha \\
&= \sum_{n=2}^{\infty} \alpha^n \left(\frac{(-1)^{n-1}}{n} + \frac{(-1)^n}{n-1} \right) \\
&= \sum_{n=2}^{\infty} (-1)^n \alpha^n \left(\frac{-1}{n} + \frac{1}{n-1} \right) \\
&= \sum_{n=2}^{\infty} (-1)^n \alpha^n \left(\frac{1}{(n-1)n} \right) \\
&= \sum_{n=2}^{\infty} \frac{(-1)^n \alpha^n}{(n-1)n} \\
&= \frac{\alpha^2}{2} - \frac{\alpha^3}{6} + \frac{\alpha^4}{12} - \frac{\alpha^5}{20} \pm ...
\end{aligned}
$$

If we just look at the first two terms, $\frac{\alpha^2}{2} - \frac{\alpha^3}{6} = \frac{\alpha^2}{2}(1 - \frac{\alpha}{3})$, and since $\frac{1}{(1+z)} = 1 - z + z^2 \pm ...$, we get $1 - \frac{\alpha}{3} \geq \frac{1}{1+\frac{\alpha}{3}}$. Therefore, $f(\alpha) \geq \frac{\alpha^2}{2}(\frac{1}{1+\frac{\alpha}{3}}) = \frac{3\alpha^2}{6+2\alpha}$ **(Proved)**

Expanding the approximation, we get

$$
\begin{aligned}
f^*(\alpha) &\sim \frac{\alpha^2}{2} \frac{1}{1+\alpha/3} \\
&= \frac{\alpha^2}{2}(1 - \frac{\alpha}{3} + \frac{\alpha^2}{9} - \frac{\alpha^3}{27} \pm ...) \\
&= \frac{\alpha^2}{2} - \frac{\alpha^3}{6} + \frac{\alpha^4}{18} - \frac{\alpha^5}{54} \pm ...
\end{aligned}
$$

Thus,
$$
\begin{aligned}
f(\alpha) - f^*(\alpha) &= (\frac{\alpha^4}{12} - \frac{\alpha^5}{20} \pm ...) - (\frac{\alpha^4}{18} - \frac{\alpha^5}{54} \pm ...) \\
&= \alpha^4(\frac{1}{12} - \frac{1}{18}) - \alpha^5(\frac{1}{20} - \frac{1}{54}) \pm ... \\
&= \frac{\alpha^4}{36} - \frac{17\alpha^5}{540} \pm ...
\end{aligned}
$$

It certainly looks like $f(\alpha) > f^*(\alpha)$. A proof can be developed by looking at the successive terms in the difference and showing that the terms are decreasing and alternating in sign.

---

(4) Given that $S = \sum_{t=1}^{n}(X_t - \mathbb{E}(X_t)), v = \sum_{t=1}^{n} V(X_t), \epsilon > 0$ and $\alpha = \frac{b\epsilon}{v}$
We know,

$$\mathbb{P}(S \geq \epsilon) \leq e^{-\lambda\epsilon} \prod_{t=1}^{n} \mathbb{E}(e^{Z_t}), \quad \text{with } Z_t = X_t - \mathbb{E}(X_t)$$

Using the result obtained in Question 4.2, we have,

$$\mathbb{E}(exp(\lambda Z_t)) \leq 1 + g(\lambda b)\lambda^2 V(Z_t) \quad \leq exp\Big(g(\lambda b)\lambda^2 V(Z_t)\Big)$$

So, we have

$$\mathbb{P}(S \geq \epsilon) = \mathbb{P}(\sum_{t=1}^{n} Z_t > \epsilon) \leq exp\Big(-\lambda\epsilon + g(\lambda b)\lambda^2 v\Big) \tag{1}$$

Since $exp()$ is a convex function, we now minimize $f(\lambda) = -\lambda\epsilon + g(\lambda b)\lambda^2 v$

We have, $f(\lambda) = -\lambda\epsilon + \lambda^2 v\Big(\frac{e^{\lambda b}-1-\lambda b}{\lambda^2 b^2}\Big) = -\lambda\epsilon + \frac{v}{b^2}[e^{\lambda b} - 1 - \lambda b]$

$$\begin{aligned} f'(\lambda) &= -\epsilon + \frac{v}{b^2}[be^{\lambda b} - b] = 0 \\ \implies -\epsilon &+ \frac{v}{b}[e^{\lambda b} - 1] = 0 \\ \implies \lambda &= \frac{log(1+\alpha)}{b} \end{aligned}$$

Thus, we have

$$\begin{aligned} f(\lambda) &= -\epsilon\frac{log(1+\alpha)}{b} + g(log(1+\alpha))\lambda^2 v \\ &= -\frac{v}{b^2}\Big[(1+\alpha)log(1+\alpha) - \alpha\Big] \end{aligned}$$

Then from (1), we have

$$\mathbb{P}(S \geq \epsilon) \leq exp\Big[-\frac{v}{b^2}[(1+\alpha)log(1+\alpha) - \alpha]\Big] \qquad \textbf{(Proved)}$$

Now, we had seen in Question 4.3 that $(1+\alpha)log(1+\alpha) - \alpha \geq \frac{3\alpha^2}{6+2\alpha}$. Putting $\alpha = \frac{b\epsilon}{v}$, we get

$$\frac{3\alpha^2}{6+2\alpha} = \frac{\frac{3b^2\epsilon^2}{v^2}}{6+2\frac{b\epsilon}{v}} = \frac{\epsilon^2}{\frac{2v^2}{b^2}[1+\frac{\epsilon b}{3v}]}$$

Therefore, from the inequality proved just above, we get:

$$\mathbb{P}(S \geq \epsilon) \leq exp\Big[-\frac{\epsilon^2}{2v[1+\frac{\epsilon b}{3v}]}\Big] \qquad \textbf{(Proved)}$$

---

(5) To get the required inequality, we need to solve $\delta = exp\big(-\frac{\epsilon^2}{2v(1+\frac{b\epsilon}{3v})}\big)$ for $\epsilon > 0$. Rearranging the terms on both sides, we get the following quadratic equation in ϵ:

$$\begin{aligned} log(\frac{1}{\delta}) &= \frac{\epsilon^2}{2v + \frac{2}{3}b\epsilon} \\ \implies \epsilon^2 - 2v\ log(\frac{1}{\delta}) &- \frac{2}{3}\ b\epsilon\ log(\frac{1}{\delta}) = 0 \end{aligned}$$

The positive root of the above quadratic equation is $\epsilon = \frac{1}{2}\left(\frac{2}{3}b\ log(\frac{1}{\delta}) + \sqrt{\big(\frac{2}{3}b\ log(\frac{1}{\delta})\big)^2 + 8v\ log(\frac{1}{\delta})}\right)$

Thus, for this choice of ϵ, we have $\mathbb{P}(S \geq \epsilon) \leq \delta$.
Further, upper bounding ϵ using the relation $\sqrt{|a|+|b|} \leq \sqrt{|a|} + \sqrt{|b|}$, we get

$$\begin{aligned} \sqrt{\big(\frac{2}{3}b\ log(\frac{1}{\delta})\big)^2 + 8v\ log(\frac{1}{\delta})} &\leq \sqrt{\big(\frac{2}{3}b\ log(\frac{1}{\delta})\big)^2} + \sqrt{8v\ log(\frac{1}{\delta})} \\ \implies \frac{1}{2}\left(\frac{2}{3}b\ log(\frac{1}{\delta}) + \sqrt{\big(\frac{2}{3}b\ log(\frac{1}{\delta})\big)^2 + 8v\ log(\frac{1}{\delta})}\right) &\leq \frac{1}{2}\left(\frac{2}{3}b\ log(\frac{1}{\delta}) + \frac{2}{3}b\ log(\frac{1}{\delta}) + 2\sqrt{2v\ log(\frac{1}{\delta})}\right) \\ &= \frac{2}{3}b\ log(\frac{1}{\delta}) + \sqrt{2v\ log(\frac{1}{\delta})} \end{aligned}$$

Hence, we have

$$\mathbb{P}\Big(S \geq \frac{2}{3}b\ log(\frac{1}{\delta}) + \sqrt{2vlog(\frac{1}{\delta})}\Big) \leq \delta \qquad \textbf{(Proved)}$$

Question 5

Show that

$$R_T \leq min\Big\{T\Delta, \Delta + \frac{4}{\Delta}\Big(1 + max\big\{0, log(\frac{T\Delta^2}{4})\big\}\Big)\Big\}$$

implies the regret of an optimally tuned Explore-then-Commit (ETC) algorithm for subgaussian 2-armed bandits with means $\mu_1, \mu_2 \in \mathbb{R}$ and $\Delta = |\mu_1 - \mu_2|$, satisfies $R_T \leq \Delta + C\sqrt{T}$ where $C > 0$ is a universal constant.

Answer: The setup is for a 2-armed bandit. Let us assume without loss of generality that arm 1 is optimal, implying suboptimality gap for arm 1 is 0, and that for arm 2 is $\Delta_2 = |\mu_1 - \mu_2| = \Delta$. We had derived in class the regret bound of ETC algorithm for k-armed bandit setting as follows:

$$R_T \leq m\sum_{i=1}^{k}\Delta_i + (T - mk)\sum_{i=1}^{k}\Delta_i\ exp(-\frac{m\Delta_i^2}{4})$$

Thus, for the 2-armed setting at hand, the above regret bound simplifies to

$$\begin{aligned}
R_T &\leq m\Delta + (T - 2m)\Delta\ exp(-\frac{m\Delta^2}{4}) \\
&\leq m\Delta + T\Delta\ exp(-\frac{m\Delta^2}{4}) \\
&= g(m)
\end{aligned}$$

We try to find m such that R_T is minimized. Hence, we find m by solving

$$\begin{aligned}
&g'(m) = 0 \\
\implies &\Delta + T\Delta\ exp(-\frac{m\Delta^2}{4})(-\frac{m\Delta^2}{4}) = 0 \\
\implies &\frac{T\Delta^3}{4}exp(-\frac{m\Delta^2}{4}) = \Delta \\
\implies &m = \frac{4}{\Delta^2}log(\frac{T\Delta^2}{4})
\end{aligned}$$

Since m should be an integer, we should have $m = \lceil\frac{4}{\Delta^2}log(\frac{T\Delta^2}{4})\rceil$. Now, it may happen that $0 \leq \frac{T\Delta^2}{4} \leq 1$, in which case we will have $log(\frac{T\Delta^2}{4}) < 0$. But by the definition of m, it has to be a non-negative integer. Thus, we modify the definition of m as follows:

$$m = max\Big\{1, \Big\lceil\frac{4}{\Delta^2}log(\frac{T\Delta^2}{4})\Big\rceil\Big\}$$

We can then rewrite the regret bound R_T as follows:

$$R_T \leq \Big\lceil\frac{4}{\Delta^2}log(\frac{T\Delta^2}{4})\Big\rceil\Delta + T\Delta\ exp\Big(-\Big\lceil\frac{4}{\Delta^2}log(\frac{T\Delta^2}{4})\Big\rceil\frac{\Delta^2}{4}\Big)$$

Again,we have, $\quad exp\Big(-\Big\lceil\frac{4}{\Delta^2}log(\frac{T\Delta^2}{4})\Big\rceil\Big) \leq exp\Big(-\frac{4}{\Delta^2}log(\frac{T\Delta^2}{4})\Big)$

Thus,

$$\begin{aligned}
R_T &\leq \Big\lceil\frac{4}{\Delta^2}log(\frac{T\Delta^2}{4})\Big\rceil\Delta + T\Delta\ exp\Big(-\frac{4}{\Delta^2}log(\frac{T\Delta^2}{4})\frac{\Delta^2}{4}\Big) \\
&= \Delta\Big\lceil\frac{4}{\Delta^2}log(\frac{T\Delta^2}{4})\Big\rceil + T\Delta\frac{4}{T\Delta^2} \\
&= \Delta\Big\lceil\frac{4}{\Delta^2}log(\frac{T\Delta^2}{4})\Big\rceil + \frac{4}{\Delta}
\end{aligned}$$

From property of ceiling functions, we know

$$\Big\lceil\frac{4}{\Delta^2}log(\frac{T\Delta^2}{4})\Big\rceil \leq \frac{4}{\Delta^2}log(\frac{T\Delta^2}{4}) + 1$$

Hence,

$$\begin{aligned}R_T &\le \Delta + \frac{4}{\Delta^2}log(\frac{T\Delta^2}{4}) + \frac{4}{\Delta}\\ &= \Delta + \frac{4}{\Delta}\Big[1 + log(\frac{T\Delta^2}{4})\Big]\\ &\le \Delta + \frac{4}{\Delta}\Big[1 + max\big\{0, log(\frac{T\Delta^2}{4})\big\}\Big]\end{aligned}$$

Now, suppose we choose $T = mk$. Then, for the current 2-armed setting i.e. for $k = 2$ with $\Delta_1 = 0$ and $\Delta_2 = \Delta$, we have

$$R_T \le m\sum_{i=1}^{2}\Delta_i = m\Delta \le 2m\Delta = T\Delta$$

$\therefore$ For $k = 2$, we can write the upper regret bound of ETC algorithm as follows:

$$R_T \le min\Bigg\{T\Delta, \Delta + \frac{4}{\Delta}\Big[1 + max\big\{0, log(\frac{T\Delta^2}{4})\big\}\Big]\Bigg\} \qquad \textbf{(Proved)}$$

We will now show that $R_T \le \Delta + C\sqrt{T}$,where C is a universal constant.

Let us consider $\Delta \le \frac{2}{\sqrt{T}}$. Then from $R_T \le T\Delta$, we get $R_T \le \sqrt{T}$, implying $C = 0$.
Now, if $\Delta > \frac{2}{\sqrt{T}}$, then we have

$$\begin{aligned}R_T &\le \Delta + \frac{4}{\Delta}\Big(1 + log^+(\frac{T\Delta^2}{4})\Big) \qquad \text{where } log^+(\frac{T\Delta^2}{4}) = max\{0, log(\frac{T\Delta^2}{4})\}\\ &\le \Delta + 4\sqrt{T} + max_{\Delta>0}\frac{1}{\Delta}log^+(\frac{T\Delta^2}{4})\end{aligned}$$

We try to maximize $\frac{1}{\Delta}log^+(\frac{T\Delta^2}{4})$ w.r.t Δ. Differentiating and equating to 0, we get the maximum as $\Delta = \frac{2e}{\sqrt{T}}$. Plugging this value of Δ, we get the maximum value as $4e^{-1}\sqrt{T}$.

Thus, the upper bound on regret comes out to be $R_T \le \Delta + (4 + 4e^{-1})\sqrt{T}$, implying that the required choice of $C = (4 + 4e^{-1})\ \forall\Delta$ **(Proved)**

Question 6

Fix $\delta \in (0, 1)$. Modify the ETC algorithm to depend on δ and prove a bound on the pseudo-regret $R_T = T\mu^* - \sum_{t=1}^{T} u_{A_t}$ of ETC algorithm that holds with probability $1 - \delta$ where A_t is the arm chosen in the round t.

Answer: We know the upper bound of the regret of ETC algorithm is given by

$$R_T \le m\sum_{i=1}^{k}\Delta_i + (T - mk)\sum_{i=1}^{k}\Delta_i exp(-\frac{m\Delta_i^2}{4})$$

- **Case 1:** $mk = T$

  In that case, we can see that the upper bound on regret gets modified to

$$R_T \le m\sum_{i=1}^{k}\Delta_i \le m(k-1)\Delta$$

- **Case 2:** $mk < T$

  Let $\Delta = max_i \Delta_i$, and let $N_i(T)$ be the number of pulls of arm i in T rounds. Let us also assume without loss of generality that arm 1 is optimal. Then,

$$\mathbb{P}(N_i(T) > m) = \mathbb{P}\Big(\hat{\mu_i}(mk) - \mu_i - \hat{\mu_1}(mk) + \mu_1 \geq \Delta_i\Big)$$
$$\leq exp\Big(-\frac{m\Delta_i^2}{4}\Big) \quad \leq \quad \delta$$

  We need to choose m such that the above condition is satisfied, i.e.

$$exp\Big(-\frac{m\Delta_i^2}{4}\Big) \leq \delta$$
$$or, \;\; m \;\geq\; \frac{4}{\Delta_i^2} log(\frac{1}{\delta}) \;\; \geq \;\; \frac{4}{\Delta^2} log(\frac{1}{\delta})$$

  If the above inequality holds for all i, then

$$N_i(T) \leq m \qquad \text{w.p. } 1-\delta$$
$$or, \;\; \Delta_i N_i(T) \leq m\Delta_i \qquad \text{w.p. } 1-\delta$$
$$or, \;\; \sum_{i=1}^{k} \Delta_i N_i(T) \leq m \sum_{i=1}^{k} \Delta_i \leq m(k-1)\Delta \qquad \text{w.p. } 1-\delta$$

  which implies that $R_T \leq m(k-1)\Delta$ w.p. $1-\delta$

Thus, the pseudo-regret is upper bounded with probability $(1-\delta)$ by

$$R_T \leq m(k-1)\Delta \qquad \text{where } m = min\Big\{\frac{n}{k}, \frac{4}{\Delta^2} log(\frac{1}{\delta})\Big\}$$

Question 7

Fix $\delta \in (0,1)$. Prove a bound on the random regret $R_T = T\mu^* - \sum_{t=1}^{T} X_t$ of ETC algorithm that holds with probability $1-\delta$. Compare this to the bound derived for the pseudo-regret in the question 6. What can you conclude?

Answer: We have $R_T = T\mu^* - \sum_{t=1}^{T} X_t$. Then,

$$\mathbb{E}(R_T) = T\mu^* - \sum_{t=1}^{T} \mathbb{E}(X_t) = T\mu^* - \sum_{t=1}^{T} \mu_{A_t} = \bar{R_T}$$

We now apply Markov's Inequality on R_T. We then get,

$$\mathbb{P}(R_T > \epsilon) \leq \frac{\mathbb{E}(R_T)}{\epsilon} = \frac{\bar{R_T}}{\epsilon}$$

Since we want the above probability to be δ such that we get a bound on R_T w.p. $1-\delta$, we have,

$$\frac{\bar{R_T}}{\epsilon} = \delta$$
$$or, \;\; \epsilon = \frac{\bar{R_T}}{\delta}$$

Thus, we can rewrite the probability equation as follows:

$$\mathbb{P}(R_T > \epsilon) = \mathbb{P}(R_T > \frac{\bar{R_T}}{\delta}) \;\; \leq \delta$$
$$\implies \mathbb{P}(R_T \leq \frac{\bar{R_T}}{\delta}) \geq 1-\delta$$

Now from Question 6, we have

$$\bar{R_T} \leq min\{T\Delta, \frac{4k}{\Delta}log(\frac{1}{\delta})\} \qquad , \Delta = max_i\Delta_i$$

$$\implies R_T \leq \frac{\bar{R_T}}{\delta} \leq \frac{1}{\delta}min\Big\{T\Delta, \frac{4k}{\Delta}log(\frac{1}{\delta})\Big\}$$

Therefore, the required upper bound on regret is given as follows:

$$R_T \leq min\Big\{\frac{T\Delta}{\delta}, \frac{4k}{\delta\Delta}log(\frac{1}{\delta})\Big\} \qquad \text{w.p. } 1-\delta$$

Question 8

Assume the rewards are 1-subgaussian and there are $k \geq 2$ arms. The ϵ- greedy algorithm depends on a sequence of parameters $\epsilon_1, \epsilon_2, \ldots$. First it chooses each arm once and subsequently chooses $A_t = argmax_i\hat{\mu_i}(t-1)$ with probability $1-\epsilon_t$ and otherwise chooses an arm uniformly at random.

1. Prove that if $\epsilon_t = \epsilon > 0$, then $\lim_{T\to\infty} \frac{R_T}{T} = \frac{\epsilon}{k}\sum_{i=1}^k \Delta_i$.
2. Let $\Delta_{min} = min\{\Delta_i : \Delta_i > 0\}$ where $\Delta_i = \mu^* - \mu_i$ and $\epsilon_t = min\{1, \frac{Ck}{t\Delta_{min}^2}\}$ where $C > 0$ is a sufficiently large universal constant. Prove that there exists a universal $C' > 0$ such that

$$R_T \leq C' \sum_{i=1}^k \Big(\Delta_i + \frac{\Delta_i}{\Delta_{min}^2}\log\max\Big\{e, \frac{T\Delta_{min}^2}{k}\Big\}\Big)$$

Answer: (1) The regret of an ϵ-greedy algorithm is given as follows:

$$R_T = \sum_{i=1}^k \Delta_i\mathbb{E}(N_i(T)) \qquad \text{where } \Delta_i = \mu^* - \mu_i, \mu^* \text{ being the mean of the optimal arm}$$

and $N_i(T)$ being the number of times arm i is pulled till round T.

Here,

$$\begin{aligned}
N_i(T) &= \sum_{t=1}^T \mathbb{I}\{I_t = i\} \\
&= 1 + \sum_{t=k+1}^T \mathbb{I}\{I_t = i\} \\
\implies \mathbb{E}[N_i(T)] &= 1 + \sum_{t=k+1}^T \mathbb{P}\{I_t = i\} \\
&= 1 + \sum_{t=k+1}^T \mathbb{P}\{I_t = i|\text{exploit}\}(1-\epsilon) + \sum_{t=k+1}^T \mathbb{P}\{I_t = i|\text{explore}\}(\epsilon) \\
&= 1 + \sum_{t=k+1}^T \mathbb{P}\{i = argmax_j\hat{\mu_j}(t-1)\}(1-\epsilon) + \frac{\epsilon}{k}(T-k) \\
&\leq 1 + \sum_{t=k+1}^T \mathbb{P}\{i = argmax_j\hat{\mu_j}(t-1)\}(1-\epsilon) + \frac{T\epsilon}{k} \\
&= 1 + \sum_{t=k+1}^T \mathbb{P}\{\max_{j\neq i}\hat{\mu_j}(t-1) \leq \hat{\mu_i}(t-1)\}(1-\epsilon) + \frac{T\epsilon}{k}
\end{aligned}$$

Let us define $\Delta_i = \max_i \Delta_i$ and $\Delta_{min} = \min\{\Delta_i : \Delta_i > 0\}$, where $\Delta_i = \mu^* - \mu_i$
Now,

$$\begin{aligned}
R_T &= \sum_{i=1}^{k} \Delta_i \mathbb{E}(N_i(T)) \\
&= \sum_{i=1}^{k} \Delta_i \Big\{ 1 + \sum_{t=k+1}^{T} \mathbb{P}\{\max_{j \neq i} \hat{\mu_j}(t-1) \leq \hat{\mu_i}(t-1)\}(1-\epsilon) + \frac{T\epsilon}{k} \Big\} \\
&= \sum_{i=1}^{k} \Delta_i + (1-\epsilon) \sum_{i=1}^{k} \Delta_i \sum_{t=k+1}^{T} \mathbb{P}\Big\{ \max_{j \neq i} \hat{\mu_j}(t-1) \leq \hat{\mu_i}(t-1) \Big\} + \frac{T\epsilon}{k} \sum_{i=1}^{k} \Delta_i
\end{aligned}$$

Now, $\mathbb{P}\Big\{ \max_{j \neq i} \hat{\mu_j}(t-1) \leq \hat{\mu_i}(t-1) \Big\} \leq exp\{-\frac{\Delta_i^2 t}{4k}\}$
Here, $\hat{\mu_i}(t-1) \sim \frac{1}{\sqrt{m}}$-subgaussian
The distribution of LHS is $\sqrt{\frac{2}{m}}$. We use the inequality $\mathbb{P}\{X \geq \epsilon\} \leq exp\{-\frac{\epsilon^2}{2\sigma^2}\}$. Take $\epsilon = \Delta_i, \sigma = \sqrt{\frac{2}{m}}$.
Here, $m = \frac{c}{k}$
Hence,

$$\begin{aligned}
R_T &\leq \sum_{i=1}^{k} \Delta_i + (1-\epsilon) \sum_{i=1}^{k} \Delta_i \sum_{t=k+1}^{T} exp\{-\frac{\Delta_i^2 t}{4k}\} + \frac{T\epsilon}{k} \sum_{i=1}^{k} \Delta_i \\
&\leq \sum_{i=1}^{k} \Delta_i + (1-\epsilon) \sum_{i=1}^{k} \Delta_i \sum_{t=k+1}^{T} exp\{-\frac{\Delta_{min}^2 t}{4k}\} + \frac{T\epsilon}{k} \sum_{i=1}^{k} \Delta_i \\
&\leq \sum_{i=1}^{k} \Delta_i + (1-\epsilon) \sum_{i=1}^{k} \Delta_i \sum_{t=0}^{\infty} exp\{-\frac{\Delta_{min}^2 t}{4k}\} + \frac{T\epsilon}{k} \sum_{i=1}^{k} \Delta_i \\
&= \sum_{i=1}^{k} \Delta_i + (1-\epsilon) \sum_{i=1}^{k} \Delta_i \Big[\frac{1}{1 - exp\{-\frac{\Delta_{min}^2}{4k}\}} \Big] + \frac{T\epsilon}{k} \sum_{i=1}^{k} \Delta_i
\end{aligned}$$

Hence, $\frac{R_T}{T} \leq \sum_{i=1}^{k} \frac{\Delta_i}{T} + \frac{1-\epsilon}{T} \sum_{i=1}^{k} \left[\frac{\Delta_i}{1 - exp\{-\frac{\Delta_{min}^2}{4k}\}} \right] + \frac{\epsilon}{k} \sum_{i=1}^{k} \Delta_i$
So, $lim_{T \to \infty} \frac{R_T}{T} = \frac{\epsilon}{k} \sum_{i=1}^{k} \Delta_i$ **(Proved)**

---

(2) We know, $R_t = \sum_i \Delta_i \mathbb{E}[N_i(T)]$. Now,

$$\begin{aligned}
\mathbb{E}[N_i(T)] &= 1 + \sum_{t=k+1}^{T} \Big[(1-\epsilon_t) P(i_t = i^*) + \frac{\epsilon_t}{k} \Big] \\
&\leq 1 + \sum_{t=k+1}^{T} \Big[(1-\epsilon_t) exp(-\frac{t\Delta_{min}^2}{4k}) + \frac{\epsilon_t}{k} \Big]
\end{aligned}$$

Now, letting $t > \frac{Ck}{\Delta_{min}^2} = t_0$, we have,

$$\mathbb{E}[N_i(T)] \leq 1 + \sum_{t=k+1}^{t_0} \frac{1}{k} + \sum_{t=t_0}^{T} \Big[(1 - \frac{Ck}{t\Delta_{min}^2}) exp(-\frac{t\Delta_{min}^2}{4k}) + \frac{Ck}{t\Delta_{min}^2} . \frac{1}{k} \Big] \quad (2)$$

Hence for $t \leq t_0$,

$$\mathbb{E}[N_i(T)] \leq 1 + \sum_{t=k+1}^{t_0} \frac{1}{k}$$

Simplifying Eqn. (2) gives,

$$
\begin{aligned}
\mathbb{E}[N_i(T)] &\leq 1 + \sum_{t=k+1}^{t_0} \frac{1}{k} + \sum_{t=t_0}^{T} \Big[exp(-\frac{t\Delta_{min}^2}{4k}) - \frac{Ck}{t\Delta_{min}^2} exp(-\frac{t\Delta_{min}^2}{4k}) + \frac{C}{t\Delta_{min}^2}\Big] \\
&\leq 1 + \frac{t_0 - k}{k} + \frac{T - t_0}{k} + \sum_{t=t_0}^{T} \Big(1 - \frac{Ck}{t\Delta_{min}^2} exp(-\frac{t\Delta_{min}^2}{4k})\Big) \\
&= 1 + \frac{T-k}{k} + \sum_{t+0}^{T} \frac{Ck}{t\Delta_{min}^2} exp(-\frac{t\Delta_{min}^2}{4k}) \\
&\leq 1 + \frac{1}{\Delta_{min}^2}\Big[\frac{T-k}{k}\Delta_{min}^2 + \sum_{t=t_0}^{T}(\Delta_{min}^2 - \frac{Ck}{t}) exp(-\frac{t\Delta_{min}^2}{4k})\Big] \\
&\leq C\Big(1 + \frac{1}{\Delta_{min}^2} log(\frac{T\Delta_{min}^2}{k})\Big)
\end{aligned}
$$

When $T > t_0$, the above equation holds for $\frac{T\Delta_{min}^2}{k} > e$. Therefore, we have

$$
\begin{aligned}
R_T &\leq \sum_i \Delta_i C' \Big(1 + \frac{1}{\Delta_{min}^2} logmax\{e, \frac{T\Delta_{min}^2}{k}\}\Big) \\
&= C' \sum_i \Big(\Delta_i + \frac{\Delta_i}{\Delta_{min}^2} logmax\{e, \frac{T\Delta_{min}^2}{k}\}\Big) \qquad \textbf{(Proved)}
\end{aligned}
$$

Question 9

Fix a 1-subgaussian k-armed bandit environment and a horizon T. Consider the version of UCB that works in phases of exponentially increasing length of 1, 2, 4,. . . . In each phase, the algorithm uses the action that would have been chosen by UCB at the beginning of the phase.

1. State and prove a bound on the regret for this version of UCB.
2. How would the result change if the l^{th} phase had a length of $\lceil \alpha^l \rceil$ with $\alpha > 1$?

Answer: Before proceeding with the solution, we define a few terms based on the phase number. These terms can be thought of as counterparts of the terms which have been already defined w.r.t time points t. These are as follows:

- l: Phase number, l_0: Last phase
- n: Horizon $= \sum_{l=1}^{l_0} 2^l = 2^{l_0+1} - 2 \implies l_0 = [\log_2(n+2) - 1]$
- t: Time point
- k: Number of arms/actions
- $(X_{ti}), t \in [2^{l_0+1} - 2], i \in [k]$: Collection of r.v.s with law of X_{ti} equal to P_i, 1-subgaussian variables
- $(Y_{li}), l \in [l_0], i \in k$: Defines the total reward in l^{th} phase.

$$
Y_{li} = \sum_{t=2^{l+1}-2-2^l}^{2^{l+1}-2} X_{ti}
$$

  (**Note:** Y_{li} being summation of 1-subgaussian rv.s becomes $\sqrt{2^l}$-subgaussian r.v., $l \in [l_0]$)

- $L_i(l) = \sum_{p=1}^{l} [\mathbb{I}\{A_p = i\}2^p]$: No. of times arm i has been played till phase l.

- $T_i(n) = \sum_{t=1}^{n} \mathbb{I}\{A_t = i\}$: No. of times arm i has been played till time point n.
- $\hat{\mu}_i(l) = \hat{\mu}_{iL_i(l)} = \frac{1}{L_i(l)} \sum_{p=1}^{l} Y_{pi}$: Empirical mean of i^{th} arm after phase l,
- $R_n = \sum_{i=1}^{k} \Delta_i E(T_i(n)) = \sum_{i=1}^{k} \Delta_i E(L_i(l_0))$ (Using the regret decomposition result)

Derivation of upper bound on regret

Our derivation will follow by showing that $E[L_i(l_0)]$ is not too large for suboptimal arms i. We proceed by decoupling the randomness from the behaviour of the UCB algorithm. We define an event G_i as follows:

$$G_i = \left\{ \mu_1 < \min_{l \in [l_0]} UCB_1(l, \delta) \right\} \cap \left\{ \hat{\mu}_{iL_i(u_i)} + \sqrt{\frac{2}{L_i(u_i)} log(1/\delta)} < \mu_1 \right\} \tag{3}$$

where $L_i(u_i) \in [n]$ is a constant to be chosen. G_i is a good event where μ_1 is never underestimated by the upper confidence bound of the first arm, while at the same time the upper confidence bound for the mean of arm i after u_i phases (or $L_i(u_i)$ rounds) from this arm is below the pay-off of the optimal arm.

We try to show the following:

1. If G_i occurs, then arm i will be played at most $L_i(u_i)$ times till last phase l_0, i.e. $L_i(l_0) \leq L_i(u_i)$
2. The complement event G_i^c occurs with low probability.

Since $L_i(l_0) \leq n$ no matter what, this means:

$$E[L_i(l_0)] = E[\mathbb{I}\{G_i\} L_i(l_0)] + E[\mathbb{I}\{G_i^c\} L_i(l_0)] \leq L_i(u_i) + n\mathbb{P}(G_i^c) \tag{4}$$

First, we try to prove a part of what we have stated that $L_i(l_0) \leq L_i(u_i)$, and then we would show that $P(G_i^c)$ is small.
Let us assume that the event G_i holds, and we try to show $L_i(l_0) \leq L_i(u_i)$ by contradiction. Let us suppose that $L_i(l_0) \geq L_i(u_i)$. Then, it implies that $\exists$ a phase $l' \in [l_0] \ni L_i(l' - 1) = L_i(u_i)$ and action i was chosen in phase l'. Then, using the definition of G_i,

$$\begin{aligned} \text{UCB}_i(l' - 1, \delta) &= \hat{\mu}_{iL_i(l'-1)} + \sqrt{\frac{2\log(1/\delta)}{L_i(l' - 1)}} \\ &= \hat{\mu}_{iL_i(u_i)} + \sqrt{\frac{2\log(1/\delta)}{L_i(u_i)}} \\ &< \mu_1 \\ &< \text{UCB}_1(l' - 1, \delta) \end{aligned}$$

Thus, action i could not have been chosen at the end of phase $(l' - 1)$ i.e. at the beginning of phase l', implying that our assumption was wrong. Therefore, if G_i occurs, we have $L_i(l_0) \leq L_i(u_i)$.
We now try to upper bound $\mathbb{P}(G_i^c)$. By definition,

$$G_i^c = \left\{ \mu_1 \geq \min_{l \in [l_0]} \text{UCB}_1(l, \delta) \right\} \cup \left\{ \hat{\mu}_{iL_i(u_i)} + \sqrt{\frac{2\log(1/\delta)}{L_i(u_i)}} \geq \mu_1 \right\} \tag{5}$$

Now, using a union bound and a concentration bound for sums of independent subgaussian random variables, we have,

$$\begin{aligned} \mathbb{P}\left(\mu_1 \geq \min_{l \in [l_0]} \text{UCB}_1(l, \delta) \right) &\leq \mathbb{P}\left(\bigcup_{s \in [l_0]} \left\{ \mu_1 \geq \hat{\mu}_{1L_i(s)} + \sqrt{\frac{2\log(1/\delta)}{L_i(s)}} \right\} \right) \\ &\leq \sum_{s=1}^{l_0} \mathbb{P}\left\{ \mu_1 \geq \hat{\mu}_{1L_i(s)} + \sqrt{\frac{2\log(1/\delta)}{L_i(s)}} \right\} \\ &\leq l_0 \, \delta \end{aligned} \tag{6}$$

Next, we bound the probability of the second set in Eq. 5. Here we assume that $L_(u_i)$ is chosen large enough that,

$$\Delta_i - \sqrt{\frac{2\log(1/\delta)}{L_i(u_i)}} \geq c\Delta_i \tag{7}$$

for some $c \in (0,1)$. Now, since $\mu_1 = \mu_i + \Delta_i$, and using one of the Concentration bounds, we get

$$\begin{aligned}
\mathbb{P}\left(\hat{\mu}_{iL_i(u_i)} + \sqrt{\frac{2\log(1/\delta)}{L_i(u_i)}} \geq \mu_1\right) &= \mathbb{P}\left(\hat{\mu}_{iL_i(u_i)} - \mu_i \geq (\mu_1 - \mu_i) - \sqrt{\frac{2\log(1/\delta)}{L_i(u_i)}}\right) \\
&= \mathbb{P}\left(\hat{\mu}_{iL_i(u_i)} - \mu_i \geq \Delta_i - \sqrt{\frac{2\log(1/\delta)}{L_i(u_i)}}\right) \\
&\leq \mathbb{P}\left(\hat{\mu}_{iL_i(u_i)} - \mu_i \geq c\Delta_i\right) \\
&\leq \exp\left(-\frac{L_i(u_i)c^2\Delta_i^2}{2}\right)
\end{aligned} \tag{8}$$

Combining equations (6) and (8), we then have:

$$\mathbb{P}(G_i^c) \leq l_0\,\delta + \exp\,(-\frac{L_i(u_i)c^2\Delta_i^2}{2}) \tag{9}$$

Substituting (9) in Eq. (4), we obtain

$$E[L_i(l_0)] \leq L_i(u_i) + n\left(l_0\,\delta + \exp\left(-\frac{L_i(u_i)c^2\Delta_i^2}{2}\right)\right) \tag{10}$$

An obvious question that remains is how do we choose $L_i(u_i)$. A natural choice is the smallest integer for which (7) holds, which is

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

Then, using an assumption that $\delta = 1/n\,l_0 = 1/[n\{\ \log_2(n+2) - 1\}]$ and the above choice of $L_i(u_i)$, we get the following expression for $\mathbb{E}[L_i(l_0)]$

$$\begin{aligned}
\mathbb{E}[L_i(l_0)] &\leq L_i(u_i) + 1 + n\,\exp\left[-\frac{c}{(1-c)^2}\log\,\left[n\{\log_2(n+2)-1\}\right]\right] \\
&= \left\lceil \frac{2\log\{n[\ \log_2(n+2)-1]\}}{(1-c)^2\Delta_i^2} \right\rceil + 1 + n^{1-\{n\ \log_2(n+2)-1\}^{c^*}}
\end{aligned}$$

where $c^* = \frac{c^2}{(1-c)^2}$ (c^* is in the exponent of the bracketed term).

Putting the above form of $\mathbb{E}[L_i(l_0)]$ into the regret decomposition result, we get the required upper bound on regret:

$$\begin{aligned}
R_n &\leq \sum_{i=1}^{k}\left(\left\lceil \frac{2\log\{n[\ \log_2(n+2)-1]\}}{(1-c)^2\Delta_i} \right\rceil + \Delta_i\big(1 + n^{1-\{n\ \log_2(n+2)-1\}^{c^*}}\big)\right) \\
\implies R_n &\leq \sum_{i=1}^{k} O\big(\log(n\,\log(n)\big) \quad \text{[We use c=1/2 and observe that the } 2^{nd} \text{ term} \to 0 \text{ as } n \to \infty]
\end{aligned}$$

The regret bound from UCB came out to be $R_n \leq \sum_{i=1}^{k} O(\log(n))$. Thus, we see that phased UCB performs worse than UCB algorithm, which is in line with our intuition.