Autumn Semester 2020 **April 21, 2021**

IE 613: Online Machine Learning

Solutions: Assignment 2

Solution 1

(1) Given: X is σ- subgaussian.
To show: $\mathbb{E}[X] = 0$ and $V(X) \leq \sigma^2$.

Proof. We know X is σ- subgaussian if

$$\begin{aligned}
\mathbb{E}\left[e^{\lambda X}\right] &\leq e^{\frac{\lambda^2\sigma^2}{2}} \\
\Rightarrow \sum_{r\geq 0} \frac{\lambda^r \mathbb{E}[X^r]}{r!} &\leq \sum_{r\geq 0} \frac{(\sigma^2)^r(\lambda^2)^r}{2^r r!} \\
\Rightarrow \lambda\mathbb{E}[X] + \frac{\lambda^2\mathbb{E}\left[X^2\right]}{2!} &\leq \frac{\sigma^2\lambda^2}{2} + o(\lambda^2) \\
\Rightarrow \mathbb{E}[X] + \frac{\lambda\mathbb{E}\left[X^2\right]}{2!} &\leq \frac{\sigma^2\lambda^2}{2} + o(\lambda) \qquad (1)
\end{aligned}$$

Putting $\lambda > 0$ in eq. (1),

$$\Rightarrow \mathbb{E}[X] + \frac{\lambda\mathbb{E}\left[X^2\right]}{2!} \leq \frac{\sigma^2\lambda^2}{2} + o(\lambda)$$

Taking $\lambda \to 0$ we get

$$\Rightarrow \mathbb{E}[X] \leq 0 \qquad (2)$$

Putting $\lambda < 0$ in eq. (1),

$$\Rightarrow \mathbb{E}[X] + \frac{\lambda\mathbb{E}\left[X^2\right]}{2!} \geq \frac{\sigma^2\lambda^2}{2} + o(\lambda)$$

Taking $\lambda \to 0$ we get

$$\Rightarrow \mathbb{E}[X] \geq 0 \qquad (3)$$

Combining eq. (2) and eq. (3) we get

$$\mathbb{E}[X] = 0. \qquad (4)$$

Since $\mathbb{E}[X] = 0 \Rightarrow V(X) = \mathbb{E}\left[X^2\right]$.

$$\mathbb{E}\left[e^{\lambda X}\right] \leq e^{\frac{\lambda^2\sigma^2}{2}}$$

$$\Rightarrow \sum_{r\geq 0} \frac{\lambda^r \mathbb{E}[X^r]}{r!} \leq \sum_{r\geq 0} \frac{\sigma^{2r}\lambda^{2r}}{2^r r!}$$
$$\Rightarrow \frac{\lambda^2(\mathbb{E}\left[X^2\right] - \sigma^2)}{2} \leq \sum_{r\geq 1} \frac{\lambda^{2r}\sigma^{2r}}{2^r r!} - \sum_{r\geq 1} \frac{\lambda^r \mathbb{E}[X^r]}{r!}$$

By taking limits $\lambda \to 0$ we get,

$$\Rightarrow \mathbb{E}\left[X^2\right] \leq \sigma^2$$
$$\Rightarrow V(X) \leq \sigma^2 \qquad \text{Hence Proved.} \tag{5}$$

□

(2) Given: X is σ- subgaussian.
To show: cX is $|c|\sigma$-subgaussian for all $c \in \mathcal{R}$.

Proof. Let $Y = cX$ where X is σ-subgaussian and $c \in \mathcal{R}$.
$\Rightarrow \mathbb{E}[Y] = \mathbb{E}[cX] = c\mathbb{E}[X] = 0$ and $V(Y) = c^2V(X)$.
According to the definition of σ-subgaussian,

$$\begin{aligned}\mathbb{E}\left[e^{tY}\right] &= \mathbb{E}\left[e^{tcX}\right] \\ &= \mathbb{E}\left[e^{(ct)X}\right] \\ &\leq e^{\frac{(ct)^2\sigma^2}{2}} \\ &= e^{\frac{t^2(|c|\sigma)^2}{2}}\end{aligned}$$

$\Rightarrow cX$ is $|c|\sigma$-subgaussian for all $c \in \mathcal{R}$. Hence Proved. □

(3) Given: X_1 and X_2 are independent and σ_1 and σ_2 - subgaussian respectively.
To show: $X_1 + X_2$ is $\sqrt{\sigma_1^2 + \sigma_2^2}$ - subgaussian.

Proof. Let $Y = X_1 + X_2$ where X_1 and X_2 are independent.
According to eq. (4), $\mathbb{E}[Y] = \mathbb{E}[X_1] + \mathbb{E}[X_2] = 0$,
as both X_1 and X_2 are σ_1 and σ_2 - subgaussian respectively and $\mathbb{E}[X_1] = \mathbb{E}[X_2] = 0$.

$$\begin{aligned}V(Y) &= V(X_1) + V(X_2) + 2COV(X_1, X_2) \\ &= V(X_1) + V(X_2) && \text{[since } X_1 \text{ and } X_2 \text{ are independent]} \\ &\leq \sigma_1^2 + \sigma_2^2 && \text{[By eq. (5)]}\end{aligned}$$

According to the definition of σ-subgaussian, we get

$$\begin{aligned}\mathbb{E}\left[e^{tY}\right] &= \mathbb{E}\left[e^{tX_1+tX_2}\right] \\ &= \mathbb{E}\left[e^{tX_1}\right]\mathbb{E}\left[e^{tX_2}\right] && [X_1 \text{ and } X_2 \text{ are independent}]\end{aligned}$$

$$\leq e^{\frac{t^2\sigma_1^2}{2}} e^{\frac{t^2\sigma_2^2}{2}}$$
$$= e^{\frac{t^2(\sigma_1+\sigma_2)^2}{2}}$$

Therefore, $X_1 + X_2$ is $\sqrt{\sigma_1^2 + \sigma_2^2}$-subgaussian. Hence Proved. $\square$

Solution 2

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

To show: X is $\frac{(b-a)}{2}$ - subgaussian.

Proof. We know X is σ- subgaussian if

$$\mathbb{E}\left[e^{tX}\right] \leq e^{\frac{t^2\sigma^2}{2}} \tag{6}$$

Define: $\Psi(t) = \ln\left(\mathbb{E}\left[e^{tX}\right]\right)$ and we can compute that,
$\Psi'(t) = \frac{\mathbb{E}[Xe^{tX}]}{\mathbb{E}[e^{tX}]}$ and $\Psi"(t) = \frac{\mathbb{E}[X^2e^{tX}]}{\mathbb{E}[e^{tX}]} - \left\{\frac{\mathbb{E}[Xe^{tX}]}{\mathbb{E}[e^{tX}]}\right\}^2$

Thus $\Psi"(t)$ can be interpreted as the variance of the random variable X under the probability measure $dQ = \frac{e^{tX}}{\mathbb{E}[e^{tX}]}dP$.

But here $X \in [a, b]$ almost surely. So, under any probability we observe that,

$$X - \frac{a+b}{2} \leq \frac{(b-a)}{2}$$
$$\Rightarrow V(X) = V(X - \frac{a+b}{2}) \leq \mathbb{E}\left[X - \frac{a+b}{2}\right]^2 \leq \frac{(b-a)^2}{4} \tag{7}$$

The fundamental theorem of calculus yields,

$$\Psi(t) = \int_0^t (\int_0^s \Psi"(u)du)ds$$

Applying eq. (7) in $\Psi(t)$,

$$\Rightarrow \Psi(t) \leq \frac{t^2}{2}\frac{(b-a)}{4}^2$$

using $\Psi(0) = log(1) = 0$ and $\Psi'(0) = \mathbb{E}[X] = 0$

$$\Rightarrow \mathbb{E}\left[e^{tX}\right] \leq e^{\frac{t^2}{2}\frac{(b-a)}{4}^2} \tag{8}$$

$\Rightarrow X$ is $\frac{(b-a)}{2}$ - subgaussian. Hence Proved.

$\square$

(2) Given: $X_1, X_2, .., X_n$ are independent and $X_t \in [a_t, b_t]$ almost surely with $a_t < b_t$ for all t.

To show: Using Cramer-Chernoff, we have to prove,

$$\mathbb{P}\left\{\sum_{t=1}^{n}(X_t - \mathbb{E}[X_t]) \geq \epsilon\right\} \leq \exp\left\{-\frac{2\epsilon^2}{\sum\limits_{t=1}^{n}(b_t - a_t)^2}\right\} \tag{9}$$

Proof. Let $Y = X - \mathbb{E}[X]$ (Centered random variables with $\mathbb{E}[Y] = 0$).

We know, e^x is a convex function.

$$\Rightarrow e^{\theta(\alpha b+(1-\alpha)a)} \leq \alpha e^{\theta b} + (1-\alpha)e^{\theta a} \qquad \text{[By Jensen's Inequality]}$$

Let $\alpha = \frac{Y-a}{b-a} \Rightarrow 1-\alpha = \frac{b-Y}{b-a}$ and $\alpha b + (1-\alpha)a = Y$

$$\Rightarrow e^{\theta Y} \leq \frac{Y-a}{b-a}e^{\theta b} + \frac{b-Y}{b-a}e^{\theta a}$$

Taking expectations on both sides,

$$\Rightarrow \mathbb{E}\left[e^{\theta Y}\right] \leq \frac{-a}{b-a}e^{\theta b} + \frac{b}{b-a}e^{\theta a}$$

Let $p = \frac{-a}{b-a} \Rightarrow 1-p = \frac{b}{b-a}$,

$$\begin{aligned}\Rightarrow \mathbb{E}\left[e^{\theta Y}\right] &\leq pe^{\theta b} + (1-p)e^{\theta a}\\ &= e^{\theta a}(1-p+pe^{\theta(b-a)})\\ &= e^{-\theta p(b-a)}(1-p+pe^{\theta(b-a)})\end{aligned}$$

Let $u = \theta(b-a)$

$$\begin{aligned}\Rightarrow \phi(u) &= -pu + log(1-p+pe^u)\\ \Rightarrow \mathbb{E}\left[e^{\theta Y}\right] &\leq e^{\phi(u)}\end{aligned} \tag{10}$$

From Taylor's expansion, $\phi(u) = \phi(0) + \phi'(o)u + \phi"(\zeta)\frac{u^2}{2}$ where $\zeta \in [0, u]$.

Now, $\phi(0) = -p.0 + log(1-p+p) = 0$ and $\phi'(0) = (-p + \frac{pe^u}{1-p+pe^u})|_{u=0} = 0$

$$\begin{aligned}\phi"(u) &= \frac{(1-p+pe^u)pe^u - pe^u pe^u}{(1-p+pe^u)^2}\\ &= \frac{pe^u(1-p+pe^u - pe^u}{(1-p+pe^u)^2}\\ &= \frac{(1-p)pe^u}{(1-p+pe^u)^2}\end{aligned}$$

Applying AM $\geq$ GM on two quantities $a = (1-p)$ and $b = pe^u$

$$
\begin{aligned}
\phi\text{"}(u) &\leq \frac{1}{4} \\
\Rightarrow \phi(u) &\leq \frac{u^2}{8} = \frac{\theta^2(b-a)^2}{8} \\
\Rightarrow \mathbb{E}\left[e^{\theta Y}\right] &\leq \exp\left\{\frac{\theta^2(b-a)^2}{8}\right\}
\end{aligned} \tag{11}
$$

Therefore,

$$
\begin{aligned}
\mathbb{P}\left\{\sum_{t=1}^{n}(X_t - \mathbb{E}\left[X_t\right]) \geq \epsilon\right\} &= \mathbb{P}\left\{\sum_{t=1}^{n} Y_t \geq \epsilon\right\} \\
&\leq e^{-\theta\epsilon}\mathbb{E}\left[e^{\theta\sum\limits_{t=1}^{n} Y_t}\right] \\
&\leq e^{-\theta\epsilon}\prod_{t=1}^{n}\mathbb{E}\left[e^{\theta Y_t}\right] \\
&\leq e^{-\theta\epsilon}\exp\left\{\sum_{t=1}^{n}\frac{\theta^2(b_t-a_t)^2}{8}\right\} \qquad \text{by eq. (11)} \\
&= f(\theta)
\end{aligned} \tag{12}
$$

Minimising $f(\theta)$ over θ, we get,

$$
\begin{aligned}
&f'(\theta) = 0 \\
&\Rightarrow \exp\left\{-\theta\epsilon + \sum_{t=1}^{n}\frac{\theta^2(b_t-a_t)^2}{8}\right\}\left(-\epsilon + \sum_{t=1}^{n}\frac{\theta(b_t-a_t)^2}{4}\right) = 0 \\
&\Rightarrow \theta = \frac{4\epsilon}{\sum\limits_{t=1}^{n}(b_t-a_t)^2} \qquad \text{and} \\
&f\text{"}(\theta) = \exp\left\{-\theta\epsilon + \sum_{t=1}^{n}\frac{\theta^2(b_t-a_t)^2}{8}\right\}\left(\sum_{t=1}^{n}\frac{(b_t-a_t)^2}{4}\right) \geq 0
\end{aligned}
$$

Putting the value of $\theta = \frac{4\epsilon}{\sum\limits_{t=1}^{n}(b_t-a_t)^2}$ in eq. (12) we get,

$$
\Rightarrow \mathbb{P}\left\{\sum_{t=1}^{n}(X_t - \mathbb{E}\left[X_t\right]) \geq \epsilon\right\} \leq \exp\left\{-\frac{2\epsilon^2}{\sum\limits_{t=1}^{n}(b_t-a_t)^2}\right\} \qquad \text{Hence Proved.}
$$

□

Solution 3

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

1. To prove: $\mathbb{E}[Z] \leq \sqrt{2\sigma^2 log(n)}$

   Proof. Let $\lambda > 0$. Then we have,

$$\begin{aligned}
exp(\lambda\mathbb{E}(Z)) &\leq \mathbb{E}(exp(\lambda Z)) \qquad \text{[By Jensen's inequality]} \\
&\leq \sum_{t=1}^{n} \mathbb{E}(\exp(\lambda X_t)) \\
&\leq n\exp(\frac{\lambda^2\sigma^2}{2}) \qquad \text{[From the definition of subgaussian r.v.]} \\
\implies \lambda\mathbb{E}(Z) &\leq log(n) + \frac{\lambda^2\sigma^2}{2} \\
\implies \mathbb{E}(Z) &\leq \frac{log(n)}{\lambda} + \frac{\lambda\sigma^2}{2} \qquad (13)
\end{aligned}$$

   Let us choose $\lambda = \frac{1}{\sigma}\sqrt{2log(n)}$. Plugging thus choice of λ in the above inequality (13), we get:

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

   □

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

   We observe that

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

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

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

   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

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

Solution 4

Given: 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. To show: $g(x) = \frac{1}{2} + \frac{x}{3!} + \frac{x^2}{4!} + \cdots = \frac{(exp(x)-1-x)}{x^2}$ is increasing.

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

   $$g'(x) = \frac{x^2(\exp(x)-1) - 2x(\exp(x)-1-x)}{x^4}$$
   $$\implies x^3 g'(x) = xe^x - 2e^x + 2 + x \quad [h(x)\text{say}]$$

   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. Hence Proved. □

2. Let X be a random variable with $\mathbb{E}[X] = 0$ and $X \leq b$ almost surely. Show that $\mathbb{E}[\exp(X)] \leq 1 + g(b)V[X]$.

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

   $$\begin{aligned}
   \mathbb{E}(\exp(X)) &= 1 + \mathbb{E}(X) + \mathbb{E}(g(X)X^2) \\
   &\leq 1 + \mathbb{E}(g(X)X^2) \qquad [\text{Since } \mathbb{E}(X) = 0] \\
   &\leq 1 + \mathbb{E}(g(b)X^2) \qquad [\text{Since } x \leq b \quad a.s. \implies \mathbb{E}(X) \leq \mathbb{E}(b)] \\
   &\leq 1 + g(b)\mathbb{E}(X^2) \\
   &\leq 1 + g(b)V(X) \qquad [\text{Since } (\mathbb{E}(X))^2 = 0] \qquad \text{Hence Proved.}
   \end{aligned} \tag{14}$$

   □

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.

   Proof. Let us denote $g(\alpha) = (1+\alpha)log(1+\alpha) - \alpha - \frac{3\alpha^2}{6+2\alpha}$. Differentiating w.r.t. α we get,

   $$g'(\alpha) = log(1+\alpha) - \frac{(6+2\alpha)(6\alpha) - 6\alpha^2}{(6+2\alpha)^2}$$
   $$= log(1+\alpha) - \frac{6\alpha}{(6+2\alpha)} + \frac{6\alpha^2}{(6+2\alpha)^2}$$

$$
\begin{aligned}
g"(\alpha) &= \frac{1}{1+\alpha} - \frac{36}{(6+2\alpha)^2} + \frac{72\alpha}{(6+2\alpha)^3} \\
&= \frac{1}{1+\alpha} - \frac{36(6+2\alpha)}{(6+2\alpha)^3} + \frac{72\alpha}{(6+2\alpha)^3} \\
&= \frac{1}{1+\alpha} + \frac{72\alpha - 36(6+2\alpha)}{(6+2\alpha)^3} \\
&= \frac{(6+2\alpha)^3 - 216(1+\alpha)}{(6+2\alpha)^3(1+\alpha)}
\end{aligned}
$$

Let us now define,

$$
\begin{aligned}
h(\alpha) &= (6+2\alpha)^3 - 216(1+\alpha) \\
h'(\alpha) &= 6(6+2\alpha)^2 - 216\alpha \\
h"(\alpha) &= 24(6+2\alpha) > 0 \quad \alpha \geq 0 \\
\implies h'(\alpha) &\text{ is increasing and } h'(0) = 0 \\
\implies h'(\alpha) &\geq 0 \\
\implies h(\alpha) &\text{ is increasing and } h(0) = 0 \\
\implies h(\alpha) &\geq 0
\end{aligned}
$$

We know $g"(\alpha) = \frac{h(\alpha)}{(6+2\alpha)^3(1+\alpha)}$. Moreover, $(1+\alpha) \geq 0, (6+2\alpha)^3 \geq 0$ and $h(\alpha) \geq 0$ as $\alpha \geq 0$.

$$
\begin{aligned}
\implies g"(\alpha) &\geq 0 \\
\implies g'(\alpha) &\text{ is increasing and } g'(0) = 0 \\
\implies g'(\alpha) &\geq 0 \\
\implies g(\alpha) &\text{ is increasing and } g(0) = 0 \\
\implies g(\alpha) &\geq 0 \\
\implies (1+\alpha)log(1+\alpha) - \alpha &\geq \frac{3\alpha^2}{6+2\alpha}, \text{ for } \alpha \geq 0. \qquad \text{Hence Proved.}
\end{aligned}
$$

□

4. Let $\epsilon > 0$ and $\alpha = \frac{b\epsilon}{v}$, prove that

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

Proof. 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,

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

Using the result obtained in (14), we have,

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

So, we have

$$\mathbb{P}(S \geq \epsilon) = \mathbb{P}\Big(\sum_{t=1}^{n} Z_t > \epsilon\Big) \leq exp\Big(-\lambda\epsilon + g(\lambda b)\lambda^2 v\Big) \tag{15}$$

Since $\exp()$ is a convex function, we now minimize $f(\lambda) = -\lambda\epsilon + g(\lambda b)\lambda^2 v$
We have, $f(\lambda) = -\lambda\epsilon + \lambda^2 v\Big(\frac{e^{\lambda b}-1-\lambda b}{\lambda^2 b^2}\Big) = -\lambda\epsilon + \frac{v}{b^2}[e^{\lambda b} - 1 - \lambda b]$.
Differentiating w.r.t. λ,

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

Thus, we have

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

Then from (15), we have

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

Now, we had seen in Question 4.3 that $(1+\alpha)log(1+\alpha) - \alpha \geq \frac{3\alpha^2}{6+2\alpha}$. Therfore,

$$\begin{aligned} \mathbb{P}(S \geq \epsilon) &\leq exp\Big[-\frac{v}{b^2}[(1+\alpha)log(1+\alpha) - \alpha]\Big] \\ &\leq exp\Big[-\frac{v}{b^2}\frac{3\alpha^2}{6+2\alpha}\Big] \end{aligned}$$

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

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

$\square$

5. Use the previous result to show that

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

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

$$\log\Big(\frac{1}{\delta}\Big) = \frac{\epsilon^2}{2v+\frac{2}{3}b\epsilon}$$

$$\epsilon^2 - \frac{2}{3}\,b\epsilon\,log\Big(\frac{1}{\delta}\Big) - 2v\,log\Big(\frac{1}{\delta}\Big) = 0$$

The positive root of the above quadratic equation is given by

$$\epsilon = \frac{1}{2}\left(\frac{2}{3}b\,log\Big(\frac{1}{\delta}\Big) + \sqrt{\Big(\frac{2}{3}b\,log\Big(\frac{1}{\delta}\Big)\Big)^2 + 8v\,log\Big(\frac{1}{\delta}\Big)}\right)$$

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

$$\sqrt{\Big(\frac{2}{3}b\,log\Big(\frac{1}{\delta}\Big)\Big)^2 + 8v\,log\Big(\frac{1}{\delta}\Big)} \leq \sqrt{\Big(\frac{2}{3}b\,log\Big(\frac{1}{\delta}\Big)\Big)^2} + \sqrt{8v\,log\Big(\frac{1}{\delta}\Big)}$$

$$\frac{1}{2}\left(\frac{2}{3}b\,log\Big(\frac{1}{\delta}\Big) + \sqrt{\Big(\frac{2}{3}b\,log\Big(\frac{1}{\delta}\Big)\Big)^2 + 8v\,log\Big(\frac{1}{\delta}\Big)}\right) \leq \frac{1}{2}\left(\frac{2}{3}b\,log\Big(\frac{1}{\delta}\Big) + \frac{2}{3}b\,log\Big(\frac{1}{\delta}\Big) + 2\sqrt{2v\,log\Big(\frac{1}{\delta}\Big)}\right)$$

$$\frac{1}{2}\left(\frac{2}{3}b\,log\Big(\frac{1}{\delta}\Big) + \sqrt{\Big(\frac{2}{3}b\,log\Big(\frac{1}{\delta}\Big)\Big)^2 + 8v\,log\Big(\frac{1}{\delta}\Big)}\right) \leq \frac{2}{3}b\,\log\Big(\frac{1}{\delta}\Big) + \sqrt{2v\,log\Big(\frac{1}{\delta}\Big)}$$

Hence, we have

$$\mathbb{P}\left(S \geq \frac{2}{3}b\,\log\Big(\frac{1}{\delta}\Big) + \sqrt{2vlog\Big(\frac{1}{\delta}\Big)}\right) \leq \delta \qquad \text{Hence Proved.}$$

$\square$

Solution 5

<u>To show:</u> The regret of an optimally tuned Explore-then-Commit (ETC) algorithm for subgaussian 2-armed bandits with means $\mu_1, \mu_2 \in \mathcal{R}$ and $\Delta = \mu_1 - \mu_2$ is given as,

$$\mathcal{R}_T \le \min\left\{T\Delta, \Delta + \frac{4}{\Delta}\left(1 + \max\left\{0, log\left(\frac{T\Delta^2}{4}\right)\right\}\right)\right\}$$

satisfies $\mathcal{R}_T \le \Delta + C\sqrt{T}$ where $C > 0$ is a universal constant.

Proof. The regret for Explore-then-Commit (ETC) for K-armed bandits is given as,

$$\mathcal{R}_T \le m \sum_{i=1}^{K} \Delta_i + (T - mK) \sum_{i=1}^{K} \Delta_i \exp\left(-\frac{m\Delta_i^2}{4}\right) \tag{16}$$

where $\Delta_i :=$ sub-optimal gaps $\mu_1 - \mu_i \quad \forall i = 1, 2, ..K; \quad \mu_i \in \mathcal{R}$.
Here $K = 2$. Let arm 1 is optimal.
$\Rightarrow \Delta_1 = 0$ and $\Delta = \Delta_2 = \mu_1 - \mu_2$.

$$\begin{aligned}
\mathcal{R}_T &\le m\Delta + (T - 2m)\Delta \exp\left(\frac{-m\Delta^2}{4}\right) \\
&\le m\Delta + T\Delta \exp\left(\frac{-m\Delta^2}{4}\right) \\
&= \psi(m)
\end{aligned}$$

We have to maximise $\psi(m)$

$$\begin{aligned}
&\Rightarrow \psi'(m) = 0 \\
\Rightarrow \Delta - T\Delta \exp\left(\frac{-m\Delta^2}{4}\right)\frac{\Delta^2}{4} &= 0 \\
&\Rightarrow m = \left\lceil \frac{4}{\Delta^2} log\left(\frac{T\Delta^2}{4}\right)\right\rceil \qquad \text{since } m \text{ takes positive values}
\end{aligned}$$

Moreover, if $0 \le \frac{T\Delta^2}{4} \le 1 \Rightarrow log\left(\frac{T\Delta^2}{4}\right) < 0$

$$\Rightarrow m < 0, \qquad \text{which is not possible}$$

$$\Rightarrow m = \max\left\{1, \left\lceil \frac{4}{\Delta^2} log\left(\frac{T\Delta^2}{4}\right)\right\rceil\right\} \tag{17}$$

The regret is given as,

$$\mathcal{R}_T \leq m\Delta + T\Delta \exp\left(\frac{-m\Delta^2}{4}\right)$$

Using $m = \left\lceil \frac{4}{\Delta^2} log\left(\frac{T\Delta^2}{4}\right)\right\rceil \Rightarrow \frac{4}{\Delta^2} log\left(\frac{T\Delta^2}{4}\right) \leq m \leq \left(1 + \frac{4}{\Delta^2} log\left(\frac{T\Delta^2}{4}\right)\right)$,

$$\begin{aligned}
&\leq \Delta + \frac{4}{\Delta}\log\left(\frac{T\Delta^2}{4}\right) + T\Delta \exp\left\{\frac{-\Delta^2}{4}\left(\frac{4}{\Delta^2}\log\left(\frac{T\Delta^2}{4}\right)\right)\right\} \\
&\leq \Delta + \frac{4}{\Delta}\log\left(\frac{T\Delta^2}{4}\right) + T\Delta \exp\left\{-\log\left(\frac{T\Delta^2}{4}\right)\right\} \\
&= \Delta + \frac{4}{\Delta}\log\left(\frac{T\Delta^2}{4}\right) + \frac{4}{\Delta} \\
&= \Delta + \frac{4}{\Delta}\left(1 + \log\left(\frac{T\Delta^2}{4}\right)\right)
\end{aligned}$$

If $0 \leq \frac{T\Delta^2}{4} \leq 1 \Rightarrow log\left(\frac{T\Delta^2}{4}\right) < 0$.

Hence the regret is given as,

$$\mathcal{R}_T \leq \Delta + \frac{4}{\Delta}\left(1 + \max\left\{0, \log\left(\frac{T\Delta^2}{4}\right)\right\}\right) \tag{18}$$

Moreover, if $T = 2m$

$$\Rightarrow \mathcal{R}_T \leq m\Delta \leq T\Delta \tag{19}$$

Combining eq. (15) and (16) we get,

$$\mathcal{R}_T \leq \min\left\{T\Delta, \Delta + \frac{4}{\Delta}\left(1 + \max\left\{0, \log\left(\frac{T\Delta^2}{4}\right)\right\}\right)\right\} \tag{20}$$

Let $\Delta = \frac{2}{\sqrt{T}}$.

$$\begin{aligned}
\mathcal{R}_T &\leq \min\left\{2\sqrt{T}, \frac{2}{\sqrt{T}} + 2\sqrt{T}\right\} \\
&\leq \max\left\{2\sqrt{T}, \frac{2}{\sqrt{T}} + 2\sqrt{T}\right\} \\
&= \begin{cases} 2\sqrt{T} = 0 + 2\sqrt{T} & \text{if } 2\sqrt{T} > \frac{2}{\sqrt{T}} + 2\sqrt{T} \\ \frac{2}{\sqrt{T}} + 2\sqrt{T} = \Delta + 2\sqrt{T} & \text{if } 2\sqrt{T} < \frac{2}{\sqrt{T}} + 2\sqrt{T} \end{cases}
\end{aligned} \tag{21}$$

Both the forms of eq. (18) are given as $\mathcal{R}_T \leq \Delta + 2\sqrt{T}$ where $C = 2$. Hence proved. □

Solution 6

To show: Let us fix $\delta \in (0,1)$. We have to modify the ETC algorithm to depend on δ and have to prove a bound on the pseudo-regret $\mathcal{R}_T = T\mu^* - \sum_{t=1}^{T} \mu_{A_t}$ of ETC algorithm that holds with probability $1-\delta$, where A_t is the arm chosen in the round t.

Proof. The regret for Explore-then-Commit (ETC) for K-armed bandits is given as,

$$\mathcal{R}_T \leq m \sum_{i=1}^{K} \Delta_i + (T - mK) \sum_{i=1}^{K} \Delta_i \exp\left(-\frac{m\Delta_i^2}{4}\right) \tag{22}$$

where $\Delta_i :=$ sub-optimal gaps $\mu_1 - \mu_i \quad \forall i = 1, 2, ..K; \quad \mu_i \in \mathcal{R}$ and $\mu_1 :=$ the optimal arm.

Let $\Delta = \max_{i \in [k]} \Delta_i$.

- Case 1: $n = mk$

$$\mathcal{R}_T \leq m \sum_{i=1}^{K} \Delta_i$$

$$\Rightarrow \mathcal{R}_T \leq mk\Delta \qquad \text{since } \Delta = \max_{i \in [k]} \Delta_i \tag{23}$$

- Case 2: $n > mk$

  According to the regret decomposition lemma,

$$\mathcal{R}_T = \sum_{i=1}^{k} \Delta_i \mathbb{E}\left[N_i(T)\right] \tag{24}$$

$$\text{where } N_i(T) = \sum_{t=1}^{T} \mathcal{I}\left\{I_t = i\right\}$$

:= the number of times arm i is played over T rounds

$$\mathbb{P}\left\{N_i(T) > m\right\} = \mathbb{P}\left\{\hat{\mu}_i(mk) - \mu_i - \hat{\mu}_1(mk) + \mu_1 \geq \Delta_i\right\} \tag{25}$$

Theorem 1. *Assume $X_i - \mu$ are independent and σ- Subgaussian random variables. Then for all $\epsilon \geq 0$,*

$$\mathbb{P}\left\{\hat{\mu} \geq \mu + \epsilon\right\} \leq \exp - \frac{n\epsilon^2}{2^2}$$

$$\mathbb{P}\left\{\hat{\mu} \leq \mu - \epsilon\right\} \leq \exp - \frac{n\epsilon^2}{2^2}$$

Proposition 1:

$$\mathbb{P}\left\{\hat{\mu}_i(mk) - \mu_i - \hat{\mu}_1(mk) + \mu_1 \geq \Delta_i\right\} \leq \exp\left\{-\frac{m\Delta_i^2}{4}\right\} \tag{26}$$

Proof. Let $X_{i,s}$ are Subgaussian random variables.

$$\text{Let } Y_i = \hat{\mu}_i(mk) - \mu_i = \frac{1}{m}\sum_{s=1}^{m} X_{i,s} - \mu_i \sim \frac{1}{\sqrt(m)} \text{ Subgaussian}$$

$$Y_1 = \hat{\mu}_i(mk) - \mu_i = \frac{1}{m}\sum_{s=1}^{m} X_{i,s} - \mu_i \sim |-1|\frac{1}{\sqrt(m)} \text{ Subgaussian}$$

$$\Rightarrow Y_i - Y_1 \sim \sqrt{\frac{2}{m}} \text{ Subgaussian} \qquad \text{By Lemma (iii) of Subgaussian}$$

Therefore,

$$\mathbb{P}\left\{Y_i - Y_1 \geq \Delta_i\right\} \leq \exp\left\{-\frac{m\Delta_i^2}{4}\right\} \qquad \text{By Theorem (1) and } \epsilon = \Delta_i$$

$$\mathbb{P}\left\{\hat{\mu}_i(mk) - \mu_i - \hat{\mu}_1(mk) + \mu_1 \geq \Delta_i\right\} \leq \exp\left\{-\frac{m\Delta_i^2}{4}\right\}$$

□

Since we want this probability to be $\leq \delta$,

$$\mathbb{P}\left\{N_i(T) > m\right\} \leq \exp\left\{-\frac{m\Delta_i^2}{4}\right\} \leq \delta \tag{27}$$

$$\Rightarrow \exp\left\{-\frac{m\Delta_i^2}{4}\right\} \leq \delta$$

$$\Rightarrow m \geq \frac{4}{\Delta_i^2}\log\left(\frac{1}{\delta}\right)$$

Since $\Delta = \max_{i\in[k]} \Delta_i$,

$$\Rightarrow m \geq \frac{4}{\Delta^2}\log\left(\frac{1}{\delta}\right) \tag{28}$$

If eq. (22) happens for all i such that $N_i(T) \leq m$ holds with probability $1-\delta$, the eq. (21) is given by

$$\Rightarrow \mathcal{R}_T \leq \sum_{i=1}^{k} \Delta_i m \qquad \text{with probability } 1-\delta$$

$$\leq mk\Delta \qquad \text{with probability } 1-\delta$$

Thus combining both cases with probability $1-\delta$ the pseudo-regret is bounded by,

$$\Rightarrow \mathcal{R}_T \leq mk\Delta = \min\left\{n\Delta, \frac{4k}{\Delta}\log\left(\frac{1}{\delta}\right)\right\}$$

where $m = \min\left\{\frac{n}{k}, \frac{4}{\Delta^2}\log\left(\frac{1}{\delta}\right)\right\}$

Hence Proved.

$\square$

Solution 7

To show: Let us fix $\delta \in (0,1)$. We have to modify the ETC algorithm to depend on δ and have to prove a bound on the random regret $\mathcal{R}_T = T\mu^* - \sum_{t=1}^{T} X_{A_t}$ of ETC algorithm that holds with probability $1-\delta$, where A_t is the arm chosen in the round t.

Proof. We define random regret as,

$$\mathcal{R}_T = T\mu^* - \sum_{t=1}^{T} X_{A_t} \qquad \text{where } A_t \text{ is the arm chosen in the round } t.$$

Taking expectations on both the sides we get,

$$\mathbb{E}\left[\mathcal{R}_T\right] = T\mu^* - \sum_{t=1}^{T} \mu_{A_t}$$

Let us denote the pseudo-regret as $\bar{\mathcal{R}}_T$. According to the definition of pseudo-regret in Question 4 we get,

$$= \bar{\mathcal{R}}_T \tag{29}$$

By Markov's Inequality for any $\epsilon > 0$ we get,

$$\mathbb{P}\left\{\mathcal{R}_T \geq \epsilon\right\} \leq \frac{\mathbb{E}\left[\mathcal{R}_T\right]}{\epsilon}$$

where $\mathbb{E}\left[\mathcal{R}_T\right]$ denotes the pseudo-regret of ETC algorithm

$$= \frac{\bar{\mathcal{R}}_T}{\epsilon} \qquad \text{since } \epsilon > 0.$$

Since we want this probability to be $= \delta$, we get

$$\mathbb{P}\left\{\mathcal{R}_T \geq \frac{\bar{\mathcal{R}}_T}{\delta}\right\} \leq \delta \qquad \text{where } \frac{\bar{\mathcal{R}}_T}{\epsilon} = \delta$$

We can say that,

$$\Rightarrow \mathcal{R}_T \leq \frac{\bar{\mathcal{R}}_T}{\delta} \qquad \text{with probability } 1-\delta \tag{30}$$

Now we will use the bound of pseudo-regret of the ETC algorithm as derived in Question 4.

The pseudo-regret for Explore-then-Commit (ETC) for K-armed bandits is given as,

$$\bar{\mathcal{R}}_T \leq m \sum_{i=1}^{K} \Delta_i + (T - mK) \sum_{i=1}^{K} \Delta_i \exp\left(-\frac{m\Delta_i^2}{4}\right) \tag{31}$$

where $\Delta_i :=$ sub-optimal gaps $\mu_1 - \mu_i \quad \forall i = 1, 2, ..K; \quad \mu_i \in \mathcal{R}$ and $\mu_1 :=$ the optimal arm.

Let $\Delta = \max_{i \in [k]} \Delta_i$.

- Case 1: $n = mk$

$$\bar{\mathcal{R}}_T \leq m \sum_{i=1}^{K} \Delta_i$$
$$\Rightarrow \frac{\bar{\mathcal{R}}_T}{\delta} \leq \frac{mk\Delta}{\delta} \qquad \text{since } \Delta = \max_{i \in [k]} \Delta_i$$

- Case 2: $n > mk$

  According to the regret decomposition lemma,

$$\bar{\mathcal{R}}_T = \sum_{i=1}^{k} \Delta_i \mathbb{E}\left[N_i(T)\right] \tag{32}$$
$$\text{where } N_i(T) = \sum_{t=1}^{T} \mathcal{I}\left\{I_t = i\right\}$$

  := the number of times arm i is played over T rounds

$$\mathbb{P}\left\{N_i(T) > m\right\} = \mathbb{P}\left\{\hat{\mu}_i(mk) - \mu_i - \hat{\mu}_1(mk) + \mu_1 \geq \Delta_i\right\}$$

  Since we want this probability to be $\leq \delta$, so by Theorem 1 and Proposition 1 we get,

$$\mathbb{P}\left\{N_i(T) > m\right\} \leq \exp\left\{-\frac{m\Delta_i^2}{4}\right\} \leq \delta \tag{33}$$
$$\Rightarrow \exp\left\{-\frac{m\Delta_i^2}{4}\right\} \leq \delta$$
$$\Rightarrow m \geq \frac{4}{\Delta_i^2} \log\left(\frac{1}{\delta}\right)$$

  Since $\Delta = \max_{i \in [k]} \Delta_i$,

$$\Rightarrow m \geq \frac{4}{\Delta^2} \log\left(\frac{1}{\delta}\right) \tag{34}$$

If eq. (30) happens for all i such that $N_i(T) \leq m$ holds with probability $1-\delta$, the eq. (29) is given by

$$\Rightarrow \bar{\mathcal{R}}_T \leq \sum_{i=1}^{k} \Delta_i m \qquad \text{with probability } 1-\delta$$
$$\Rightarrow \frac{\bar{\mathcal{R}}_T}{\delta} \leq \frac{mk\Delta}{\delta} \qquad \text{with probability } 1-\delta$$

Thus combining both cases with probability $1-\delta$ the pseudo-regret is bounded by,

$$\Rightarrow \frac{\bar{\mathcal{R}}_T}{\delta} \leq \frac{mk\Delta}{\delta} = \min\left\{\frac{n\Delta}{\delta}, \frac{4k}{\Delta}\frac{1}{\delta}\log\left(\frac{1}{\delta}\right)\right\}$$

where $m = \min\left\{\frac{n}{k}, \frac{4}{\Delta^2}\log\left(\frac{1}{\delta}\right)\right\}$

Therefore, according to eq. (27) the random regret is upper bounded by

$$\mathcal{R}_T \leq \min\left\{\frac{n\Delta}{\delta}, \frac{4k}{\Delta}\frac{1}{\delta}\log\left(\frac{1}{\delta}\right)\right\} \tag{35}$$

where $m = \min\left\{\frac{n}{k}, \frac{4}{\Delta^2}\log\left(\frac{1}{\delta}\right)\right\}$

As compared with the pseudo-regret bound, the random regret bound of ETC algorithm is $\frac{1}{\delta}$ times the pseudo-regret bound where $\delta \in (0,1)$. Hence, we can conclude that the pseudo-regret bound is tighter than the random regret bound.

Hence Proved.

□

Solution 8

Given: 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 $A_t = \arg\max \hat{\mu}_i(t-1)$ with probability $1-\epsilon_t$ and otherwise chooses an arm uniformly at random.

(i) To prove: If $\epsilon_t = \epsilon > 0$, then

$$\lim_{T\to\infty} \frac{\mathcal{R}_T}{T} = \frac{\epsilon}{k}\sum_{i=1}^{k} \Delta_i \tag{36}$$

Proof. According to the regret decomposition lemma, the regret is given as

$$\mathcal{R}_T = \sum_{i=1}^{k} \Delta_i \mathbb{E}\left[N_i(T)\right]$$

where

$$N_i(T) = \sum_{t=1}^{T} \mathcal{I}\left\{I_t = i\right\} := \text{No. of times arm } i \text{ is played over } T \text{ rounds}$$
$$= 1 + \sum_{t=k+1}^{T} \mathcal{I}\left\{I_t = i\right\}$$
$$\mathbb{E}\left[N_i(T)\right] = 1 + \sum_{t=k+1}^{T} \mathbb{P}\left\{I_t = i\right\}$$
$$= 1 + \sum_{t=k+1}^{T} \mathbb{P}\left\{I_t = i | exploit\right\}(1-\epsilon) + \sum_{t=k+1}^{T} \mathbb{P}\left\{I_t = i | explore\right\}\epsilon$$
$$= 1 + \sum_{t=k+1}^{T} \mathbb{P}\left\{i = \arg\max_{j\in[k]} \hat{\mu}_j(t-1)\right\}(1-\epsilon) + \sum_{t=k+1}^{T} \frac{\epsilon}{k}$$
$$= 1 + \sum_{t=k+1}^{T} \mathbb{P}\left\{i = \arg\max_{j\in[k]} \hat{\mu}_j(t-1)\right\}(1-\epsilon) + (T-k)\frac{\epsilon}{k}$$
$$\leq 1 + \sum_{t=k+1}^{T} \mathbb{P}\left\{\hat{\mu}_i(t-1) \geq \max_{j\neq i} \hat{\mu}_j(t-1)\right\}(1-\epsilon) + T\frac{\epsilon}{k}$$
$$\Rightarrow \mathcal{R}_T \leq \sum_{i=1}^{k}\Delta_i + \sum_{i=1}^{k}\Delta_i \sum_{t=k+1}^{T} \mathbb{P}\left\{\hat{\mu}_i(t-1) \geq \max_{j\neq i} \hat{\mu}_j(t-1)\right\}(1-\epsilon) + \frac{T\epsilon}{k}\sum_{i=1}^{k}\Delta_i \tag{37}$$

We will bound $\mathbb{P}\left\{\hat{\mu}_i(t-1) \geq \max_{j\neq i} \hat{\mu}_j(t-1)\right\}$

$$\mathbb{P}\left\{\hat{\mu}_i(t-1) \geq \max_{j\neq i} \hat{\mu}_j(t-1)\right\} \leq \mathbb{P}\left\{\hat{\mu}_i(t-1) \geq \hat{\mu}_1(t-1)\right\}$$
$$= \mathbb{P}\left\{(\hat{\mu}_i(t-1) - \mu_i) - (\hat{\mu}_1(t-1) - \mu_1) \geq \mu_1 - \mu_i\right\}$$
$$= \mathbb{P}\left\{(\hat{\mu}_i(t-1) - \mu_i) - (\hat{\mu}_1(t-1) - \mu_1) \geq \Delta_i\right\}$$
$$\Rightarrow \mathbb{P}\left\{\hat{\mu}_i(t-1) \geq \max_{j\neq i} \hat{\mu}_j(t-1)\right\} \leq \exp\left\{\frac{-t\Delta_i^2}{4k}\right\}$$

$$\Rightarrow \sum_{t=k+1}^{T} \mathbb{P}\left\{\hat{\mu}_i(t-1) \geq \max_{j\neq i} \hat{\mu}_j(t-1)\right\} \leq \sum_{t=k+1}^{\infty} \mathbb{P}\left\{\hat{\mu}_i(t-1) \geq \max_{j\neq i} \hat{\mu}_j(t-1)\right\}$$

$$\leq \sum_{t=k+1}^{\infty} \exp\left\{\frac{-t\Delta_i^2}{4k}\right\} \qquad \text{By Proposition 1}$$

Let $\Delta_{min} = \min\{\Delta_i; \Delta_i > 0\}$ and $\Delta = \max_{i\in k} \Delta_i$

$$\leq \sum_{t=1}^{\infty} \exp\left\{\frac{-t\Delta_{min}^2}{4k}\right\} = \frac{1}{\exp\left\{\frac{\Delta_{min}^2}{4k}\right\} - 1} \tag{38}$$

Applying eq. (34) in eq. (35), we get

$$\mathcal{R}_T \leq \sum_{i=1}^{k} \Delta_i + \sum_{i=1}^{k} \frac{\Delta_i(1-\epsilon)}{\exp\left\{\frac{\Delta_{min}^2}{4k}\right\} - 1} + \frac{T\epsilon}{k} \sum_{i=1}^{k} \Delta_i$$

$$\Rightarrow \frac{\mathcal{R}_T}{T} \leq \frac{k\Delta}{T} + \frac{k\Delta(1-\epsilon)}{T\left(\exp\left\{\frac{\Delta_{min}^2}{4k}\right\} - 1\right)} + \frac{\epsilon}{k} \sum_{i=1}^{k} \Delta_i$$

$$\Rightarrow \lim_{T\to\infty} \frac{\mathcal{R}_T}{T} = \frac{\epsilon}{k} \sum_{i=1}^{k} \Delta_i \qquad \text{Hence Proved.}$$

□

(ii) To prove: Let $\Delta_{min} = \min\{\Delta_i; \Delta_i > 0\}$ where $\Delta_i = \mu^* - \mu_i$ and $\epsilon_t = \min\left\{1, \frac{Ck}{t\Delta_{min}^2}\right\}$ where $C > 0$ is a sufficiently large universal constant. We have to prove that there exists a universal $C' > 0$ such that

$$\mathcal{R}_T \leq C' \sum_{i=1}^{k} \left(\Delta_i + \frac{\Delta_i}{\Delta_{min}^2} \log\max\left\{e, \frac{T\Delta_{min}^2}{k}\right\}\right) \tag{39}$$

Proof. According to the regret decomposition lemma,

$$\mathcal{R}_T = \sum_{i=1}^{k} \Delta_i \mathbb{E}\left[N_i(T)\right] \qquad \text{where} \tag{40}$$

$N_i(T) = \sum_{t=1}^{T} \mathcal{I}\{I_t = i\}$:= the number of times arm i is played over T rounds

$$= 1 + \sum_{t=k+1}^{T} \mathcal{I}\{I_t = i\}$$

$$\Rightarrow \mathbb{E}\left[N_i(T)\right] = 1 + \sum_{t=k+1}^{T} \mathbb{P}\{I_t = i\}$$

We will bound $\mathbb{E}\left[N_i(T)\right]$

$$\Rightarrow \mathbb{E}\left[N_i(T)\right] = 1 + \sum_{t=k+1}^{T} \mathbb{P}\left\{I_t = i | exploit\right\}(1-\epsilon_t) + \sum_{t=k+1}^{T} \mathbb{P}\left\{I_t = i | explore\right\}\epsilon_t$$
$$= 1 + \sum_{t=k+1}^{T} \mathbb{P}\left\{i = \arg\max_{j\text{ß}[k]} \hat{\mu}_j(t-1)\right\}(1-\epsilon_t) + \sum_{t=k+1}^{T} \frac{\epsilon_t}{k}$$

Since $\epsilon_t = \min\left\{1, \frac{Ck}{t\Delta_{min}^2}\right\}$, where $C > 0$ is a sufficiently large universal constant.

$$\leq 1 + \sum_{t=k+1}^{T} \mathbb{P}\left\{\hat{\mu}_i(t-1) \geq \max_{j\neq i} \hat{\mu}_j(t-1)\right\}(1-\epsilon_t) + \sum_{t=k+1}^{T} \frac{C}{t\Delta_{min}^2}$$
$$= 1 + \sum_{t=k+1}^{T} \mathbb{P}\left\{\hat{\mu}_i(t-1) \geq \max_{j\neq i} \hat{\mu}_j(t-1)\right\}(1-\epsilon_t) + \frac{C}{\Delta_{min}^2}\sum_{t=k+1}^{T} \frac{1}{t}$$
$$\leq 1 + \sum_{t=k+1}^{T} \mathbb{P}\left\{\hat{\mu}_i(t-1) \geq \max_{j\neq i} \hat{\mu}_j(t-1)\right\}(1-\epsilon_t) + \frac{C}{\Delta_{min}^2}\int_{1}^{T-k} \frac{ds}{s+k}$$
$$= 1 + \sum_{t=k+1}^{T} \mathbb{P}\left\{\hat{\mu}_i(t-1) \geq \max_{j\neq i} \hat{\mu}_j(t-1)\right\}(1-\epsilon_t) + \frac{C}{\Delta_{min}^2}\log\left(\frac{T}{k}\right)$$
$$\Rightarrow \mathbb{E}\left[N_i(T)\right] \leq 1 + \sum_{t=k+1}^{T} \mathbb{P}\left\{\hat{\mu}_i(t-1) \geq \max_{j\neq i} \hat{\mu}_j(t-1)\right\} + \frac{C}{\Delta_{min}^2}\log\left(\frac{T\Delta_{min}^2}{k}\right) \tag{41}$$

We will bound $\mathbb{P}\left\{\hat{\mu}_i(t-1) \geq \max_{j\neq i} \hat{\mu}_j(t-1)\right\}$

$$\mathbb{P}\left\{\hat{\mu}_i(t-1) \geq \max_{j\neq i} \hat{\mu}_j(t-1)\right\} \leq \mathbb{P}\left\{\hat{\mu}_i(t-1) \geq \hat{\mu}_1(t-1)\right\}$$
$$= \mathbb{P}\left\{(\hat{\mu}_i(t-1) - \mu_i) - (\hat{\mu}_1(t-1) - \mu_1) \geq \mu_1 - \mu_i\right\}$$
$$= \mathbb{P}\left\{(\hat{\mu}_i(t-1) - \mu_i) - (\hat{\mu}_1(t-1) - \mu_1) \geq \Delta_i\right\}$$
$$\Rightarrow \mathbb{P}\left\{\hat{\mu}_i(t-1) \geq \max_{j\neq i} \hat{\mu}_j(t-1)\right\} \leq \exp\left\{\frac{-t\Delta_i^2}{4k}\right\} \qquad \text{By Proposition 1}$$

Let $\Delta_{min} = \min\left\{\Delta_i; \Delta_i > 0\right\}$

$$\Rightarrow \mathbb{P}\left\{\hat{\mu}_i(t-1) \geq \max_{j\neq i} \hat{\mu}_j(t-1)\right\} \leq \exp\left\{\frac{-t\Delta_{min}^2}{4k}\right\} \tag{42}$$

Applying eq. (38) in eq. (37),

$$\Rightarrow \mathbb{E}\left[N_i(T)\right] \leq 1 + \sum_{t=k+1}^{T} \exp\left\{\frac{-t\Delta_{min}^2}{4k}\right\} + \frac{C}{\Delta_{min}^2}\log\left(\frac{T\Delta_{min}^2}{k}\right)$$
$$\leq 1 + \sum_{t=1}^{\infty} \exp\left\{\frac{-t\Delta_{min}^2}{4k}\right\} + \frac{C}{\Delta_{min}^2}\log\left(\frac{T\Delta_{min}^2}{k}\right)$$

$$
\begin{aligned}
&= 1 + \frac{1}{\exp\left\{\frac{\Delta_{min}^2}{4k}\right\} - 1} + \frac{C}{\Delta_{min}^2} \log\left(\frac{T\Delta_{min}^2}{k}\right) \\
&\leq 1 + \frac{4k}{\Delta_{min}^2} + \frac{C}{\Delta_{min}^2} \log\left(\frac{T\Delta_{min}^2}{k}\right) \qquad \text{since } e^x > 1 + x \\
&\leq C'\left(1 + \frac{1}{\Delta_{min}^2} + \frac{1}{\Delta_{min}^2} \log\left(\frac{T\Delta_{min}^2}{k}\right)\right)
\end{aligned}
$$

where C' is an universal constant appropriately chosen

If $0 \leq \frac{T\Delta_{min}^2}{k} \leq 1 \Rightarrow \log\left(\frac{T\Delta_{min}^2}{k}\right) < 0$

Therefore,

$$
\Rightarrow \mathbb{E}\left[N_i(T)\right] \leq C'\left(1 + \frac{1}{\Delta_{min}^2} \log \max\left\{e, \frac{T\Delta_{min}^2}{k}\right\}\right) \tag{43}
$$

Applying eq. (39) in eq. (36),

$$
\Rightarrow \mathcal{R}_T \leq C' \sum_{i=1}^{k} \left(\Delta_i + \frac{\Delta_i}{\Delta_{min}^2} \log \max\left\{e, \frac{T\Delta_{min}^2}{k}\right\}\right) \qquad \text{Hence Proved.}
$$

□

Solution 9

Given: 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.

Some basic notations in terms of phase:

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

- $N_i(l) = \sum_{p=1}^{l} \mathbb{I}\{A_p = i\}.2^p$: 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 slot n.
- $\hat{\mu}_i(l) = \hat{\mu}_{iN_i(l)} = \frac{1}{N_i(l)} \sum_{p=1}^{l} Y_{pi}$: Empirical mean of i^{th} arm after phase l,
- By regret decomposition result, regret can be written as

$$R_n = \sum_{i=1}^{k} \Delta_i E(T_i(n)) = \sum_{i=1}^{k} \Delta_i E(N_i(L)).$$

To prove:

1. State and prove a bound on the regret for this version of UCB.

Derivation of upper bound on regret

Proof. We would follow the same line of proof as given for UCB algorithm in the book "Bandit Algorithms" by Tor Lattimore.

Our derivation will follow by showing that $E[N_i(L)]$ is not too large for suboptimal arms i. Let us consider arm 1 is optimal, i.e., $\mu_1 = \max_{i\in[K]} \mu_i$. We proceed by decoupling the randomness from the behaviour of the UCB algorithm. We define an event G_i as follows:

$$G_i = \left\{ \mu_1 < \min_{l\in[L]} UCB_1(l,\delta) \right\} \cap \left\{ \hat{\mu}_{iN_i(u_i)} + \sqrt{\frac{2}{N_i(u_i)} log(1/\delta)} < \mu_1 \right\} \tag{44}$$

where $N_i(u_i) \in [n]$ 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 $N_i(u_i)$ rounds) from this arm is below the pay-off of the optimal arm.

We try to show the following:

(a) If G_i occurs, then arm i will be played at most $N_i(u_i)$ times till last phase L, i.e. $N_i(L) \le N_i(u_i)$

(b) The complement event G_i^c occurs with low probability.

Since $N_i(L) \le n$ no matter what, this means:

$$E[N_i(L)] = E[\mathbb{I}\{G_i\}N_i(L)] + E[\mathbb{I}\{G_i^c\}N_i(L)] \le N_i(u_i) + n\mathbb{P}(G_i^c) \tag{45}$$

First, we try to prove a part of what we have stated that $N_i(L) \leq N_i(u_i)$, 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 $N_i(L) \leq N_i(u_i)$ by contradiction. Let us suppose that $N_i(L) \geq N_i(u_i)$. Then, it implies that $\exists$ a phase $l' \in [L] \ni N_i(l'-1) = N_i(u_i)$ and action i was chosen in phase l'. Then, using the definition of G_i,

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

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 $N_i(L) \leq N_i(u_i)$.

We now try to upper bound $\mathbb{P}(G_i^c)$. By definition,

$$G_i^c = \left\{\mu_1 \geq \min_{l\in[L]} \text{UCB}_1(l,\delta)\right\} \cup \left\{\hat{\mu}_{iN_i(u_i)} + \sqrt{\frac{2\text{log}(1/\delta)}{N_i(u_i)}} \geq \mu_1\right\} \quad (46)$$

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

$$\begin{aligned}
\mathbb{P}\left(\mu_1 \geq \min_{l\in[L]} \text{UCB}_1(l,\delta)\right) &\leq \mathbb{P}\left(\bigcup_{s\in[L]} \left\{\mu_1 \geq \hat{\mu}_{1N_i(s)} + \sqrt{\frac{2\text{log}(1/\delta)}{N_i(s)}}\right\}\right) \\
&\leq \sum_{s=1}^{L} \mathbb{P}\left\{\mu_1 \geq \hat{\mu}_{1N_i(s)} + \sqrt{\frac{2\text{log}(1/\delta)}{N_i(s)}}\right\} \\
&\leq L\,\delta
\end{aligned} \quad (47)$$

Next, we bound the probability of the second set in Eq. 46. Here we assume that $N_i(u_i)$ is chosen large enough that,

$$\Delta_i - \sqrt{\frac{2\text{log}(1/\delta)}{N_i(u_i)}} \geq c\Delta_i \quad (48)$$

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

Concentration bounds, we get

$$\begin{aligned}
\mathbb{P}\bigg(\hat{\mu}_{iN_i(u_i)} + \sqrt{\frac{2\log(1/\delta)}{N_i(u_i)}} \geq \mu_1\bigg) &= \mathbb{P}\bigg(\hat{\mu}_{iN_i(u_i)} - \mu_i \geq (\mu_1 - \mu_i) - \sqrt{\frac{2\log(1/\delta)}{N_i(u_i)}}\bigg) \\
&= \mathbb{P}\bigg(\hat{\mu}_{iN_i(u_i)} - \mu_i \geq \Delta_i - \sqrt{\frac{2\log(1/\delta)}{N_i(u_i)}}\bigg) \\
&\leq \mathbb{P}\bigg(\hat{\mu}_{iN_i(u_i)} - \mu_i \geq c\Delta_i\bigg) \\
&\leq \exp\left(-\frac{N_i(u_i)c^2\Delta_i^2}{2}\right)
\end{aligned} \tag{49}$$

Combining equations (47) and (49), we then have:

$$\mathbb{P}(G_i^c) \leq L\,\delta + \exp\,(-\frac{N_i(u_i)c^2\Delta_i^2}{2}) \tag{50}$$

Substituting (50) in Eq. (45), we obtain

$$E[N_i(L)] \leq N_i(u_i) + n\left(L\,\delta + \exp\,\left(-\frac{N_i(u_i)c^2\Delta_i^2}{2}\right)\right) \tag{51}$$

An obvious question that remains is how do we choose $N_i(u_i)$. A natural choice is the smallest integer for which (48) holds, which is

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

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

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

where $c^* = \frac{c^2}{(1-c)^2}$ (c^* is in the exponent of the bracketed term). All that remains is to choose $c^* \in (0,1)$. The second term will contribute a polynomial dependence on n unless $c^2/(1-c)^2 1$. However, if c is chosen too close to 1, then the first term blows up. Somewhat arbitrarily we choose $c = \frac{1}{2}$, which leads to

$$\begin{aligned}
\mathbb{E}[N_i(L)] &\leq \left\lceil \frac{8\,\log\{n[\,\log_2(n+2)-1]\}}{\Delta_i^2} \right\rceil + 1 + n^{1-\{n\,\log_2(n+2)-1\}} \\
&= 2 + \frac{8\,\log\{n[\,\log_2(n+2)-1]\}}{\Delta_i^2} + n^{1-\{n\,\log_2(n+2)-1\}}
\end{aligned}$$

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

$$R_n \leq \sum_{i=1}^{k} \left(\frac{8\log\{n[\,\log_2(n+2)-1]\}}{\Delta_i} + \Delta_i\big(2 + n^{1-\{n\log_2(n+2)-1\}}\big) \right)$$
$$\implies R_n \leq \sum_{i=1}^{k} O\big(\log(n\,\log(n)\big)$$

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. □

2. How would the result change if the l^{th} phase had a length of $\lceil \alpha^l \rceil$ with $\alpha > 1$?

Proof. In this case, in l^{th} phase we will play the selected arm for $\lceil a^l \rceil$ with $\alpha > 1$ times. The regret bound will be of same order as obtained in Q. 9 (First part) with some constant. □