Stochastic Optimisation - Problem Sheet 2

Dom Hutchinson

October 28, 2020

Question 3.

Question 3. (a)
Show that, if arm 2 is played by the above algorithm in round $s+1$ (i.e. $I(s+1)=2$) then one of the following statements must be true.

i). $N_2(s) < \dfrac{2\alpha\ln(s)}{\Delta^2}$

ii). $\hat{\mu}_{2,N_2(s)} \geq \mu_2 + \sqrt{\dfrac{\alpha\ln(s)}{2N_2(s)}}$

Answer 3. (a)
This is a proof by contradiction.
Suppose $I(s+1)=2$ but that none of the statements above hold. Then

$$\begin{array}{rrcll} & \hat{\mu}_{2,N_2(s)} - \sqrt{\dfrac{\alpha\ln(s)}{2N_2(s)}} & < & \mu_2 & \text{by not ii)} \\ & & = & \mu_1 - \Delta & \text{by def. of } \Delta \\ & & \leq & \mu_1 - \sqrt{\dfrac{2\alpha\ln(s)}{N_2(s)}} & \text{by not i)} \\ \implies & \hat{\mu}_{2,N_2(s)} + \sqrt{\dfrac{2\alpha\ln(s)}{N_2(s)}} - \sqrt{\dfrac{\alpha\ln(s)}{2N_2(s)}} & < & \mu_1 & \\ \implies & \hat{\mu}_{2,N_2(s)} + \left(\sqrt{2} - \dfrac{1}{\sqrt{2}}\right)\sqrt{\dfrac{\alpha\ln(s)}{N_2(s)}} & < & \mu_1 & \\ \implies & \hat{\mu}_{2,N_2(s)} + \sqrt{\dfrac{\alpha\ln(s)}{2N_2(s)}} & < & \mu_1 & \\ \implies & i_2(s) & < & \mu_1 & \end{array}$$

This means $I(s+1)=1$, which is a contradiction. Thus at least one of i) or ii) must be true.$\square$

Question 3. (b)
Recall that $N_2(t) = \sum_{s=1}^{t}\{I(S)=2\}$. For an arbitrary positive integer u and any $t\in\mathbb{N}$ explain why

$$N_2(t) \leq u + \sum_{s=u+1}^{t}\big\{\{N_2(s-1)\geq u\} \text{ and } \{I(s)=2\}\big\}$$

Fix $t,u\in\mathbb{N}$. We have two possibilities

Case 1 $N_2(t) \leq u$ (i.e. Arm two has not been played u times yet). The result trivially holds in this case.

Case 2 $\exists\ s \in [1,t]$ such that $N(s) > u$ (i.e. Arm two has been played at least u times).
Let s^* denote the smallest such s. Then it must be true that $N(s^*-1) = u$ and $s^* \geq u+1$.
Hence

$$\begin{array}{rcll} N(t) & = & \sum_{s=1}^{s^*-1} I(s) + \sum_{s=s^*}^{t} I(s) & \\ & = & N(s^*-1) + \sum_{s=s^*}^{t} I(s) \underbrace{\{N(s-1) \geq u\}}_{\text{true for all in sum}} & \\ & \leq & u + \sum_{s=u+1}^{t} \{N(s-1) \geq u\} & \text{since } s^* \geq u+1 \end{array}$$

Thus the result holds in all cases. □

Question 3. (c)
Define $u = \lceil (2\alpha \ln(t))/\Delta^2 \rceil$. Using the answers to parts `(a)` and `(b)`, and relevant probability inequalities, show that

$$\mathbb{E}[N_2(t) \leq u + \sum_{s=u+1}^{t} e^{-\alpha \ln(s)}$$

Use this to show that $\mathbb{E}[N_2(t)] \leq u + \dfrac{1}{\alpha - 1}$.

Answer 3. (c)
We have

$$\mathbb{E}[N_2(t)] \leq u + \sum_{s=u+1}^{t} e^{-\alpha \ln(s)}$$

Taking expectations of both sides

$$\begin{array}{rcl} \mathbb{E}[N_2(t)] & \leq & u + \displaystyle\sum_{s=u+1}^{t} \mathbb{P}\left(\{N_2(s-1) \geq u\} \text{ and } \{I(s) = 2\}\right) \\ & \leq & u + \displaystyle\sum_{s=u}^{t-1} \mathbb{P}\left(\{N_2(s) \geq u\} \text{ and } \{I(s+1) = 2\}\right) \end{array}$$

If $N_2(s) \geq u$ and $I(s+1) = 2$ then

$$\hat{\mu}_{2,N_2(s)} \geq \mu_2 + \sqrt{\frac{\alpha \ln(s)}{2N_2(s)}} \text{ by a)}$$

Thus

$$\mathbb{E}(N_2(t)) \leq u + \sum_{s=u}^{t-1} \mathbb{P}\left(\hat{\mu}_{2,N_2(s)} \geq \mu_2 + \sqrt{\frac{\alpha \ln(s)}{2N_2(s)}}\right) \quad (1)$$

Let $X_1, \dots, X_{N_2}$ be the random variables for each time arm 2 was played. Consider

$$\begin{array}{rcll} \mathbb{P}\left(\hat{\mu}_{2,N_2(s)} \geq \mu_2 + \sqrt{\dfrac{\alpha \ln(s)}{2N_2(s)}}\right) & = & \mathbb{P}\left(\dfrac{1}{N_2}\displaystyle\sum_{i=1}^{N_2} X_i \geq \mu_2 + \sqrt{\dfrac{\alpha \ln(s)}{2N_2(s)}}\right) & \\ & = & \mathbb{P}\left(\displaystyle\sum_{i=1}^{N_2}(X_i - \mu_2) \geq N_2\sqrt{\dfrac{\alpha \ln(s)}{2N_2(s)}}\right) & \\ & \leq & \exp\left(-2 \cdot N_2 \cdot \dfrac{\alpha \ln(s)}{2N_2(s)}\right) & \text{by Hoeffding's Ineq.} \\ & = & \exp(-\alpha \ln(s)) & \\ \implies \quad \mathbb{E}[N_2(t)] & \leq & u + \displaystyle\sum_{s=u+1}^{t} e^{-\alpha \ln(s)} & \text{by (1)} \end{array}$$

Further

$$\begin{aligned}
\mathbb{E}[N_2(t)] &\leq u + \sum_{s=u+1}^{t} e^{-\alpha\ln(s)} \\
&= u + \sum_{s=u+1}^{t} s^{-\alpha} \\
&\leq u + \int_u^\infty s^{-\alpha}ds \quad \text{since } \alpha > 1 \\
&= u + \left[\frac{s^{-\alpha+1}}{-\alpha+1}\right]_u^\infty \\
&= u - \frac{u^{-\alpha+1}}{-\alpha+1} \\
&= u + \frac{u^{-\alpha+1}}{\alpha+1}
\end{aligned}$$

By the definition of u, $u > 1$ thus $u^{-\alpha+1} < 1$ since $\alpha > 1$. Giving us

$$\mathbb{E}[N_2(t)] \leq u + \frac{1}{\alpha - 1}$$

Question 3. (d)
Use the answer to `(c)` to show that the regret of this algorithm is bounded above as

$$\mathcal{R}(T) \leq \frac{2\alpha\ln(T)}{\Delta} + \frac{\alpha}{\alpha-1}\Delta$$

Answer 3. (d)

$$\begin{aligned}
\mathcal{R}(T) &:= \Delta\mathbb{E}[N_2(t)] \\
&\leq \Delta\left(u + \frac{1}{\alpha-1}\right) && \text{by 3. (c)} \\
&\leq \Delta\left(\frac{2\alpha\ln(T)}{\Delta^2} + 1 + \frac{1}{\alpha-1}\right) && \text{by def. of } u \\
&= \frac{2\alpha\ln(T)}{\Delta} + \Delta\left(1 + \frac{1}{\alpha-1}\right) \\
&= \frac{2\alpha\ln(T)}{\Delta} + \frac{\Delta\alpha}{\alpha-1}
\end{aligned}$$

Question 4.
Consider a bandit with two independent Gaussian arms. Rewards on arm i constitute a sequence of iid $N(\mu_i, 1)$ random variables.

Question 4. (a)
Let $\hat{\mu}_{i,n}$ denote the sample mean reward on arm i after n plays of this arms. Using a resulting from Homework 1, show that

$$\mathbb{P}\left(\hat{\mu}_{i,n} < \mu_i + \sqrt{\frac{\alpha\ln(t}{2n}}\right) \leq \exp\left(-\frac{\alpha\ln(t)}{4}\right)$$

Express the last quantity as power of t.

Answer 4. (a)
Let $\hat{\mu}_{i,n}$ be the sample mean reward on arm i after n plays of that arms.
From *Problem Sheet 1 6b)*, for $X_i \overset{\text{iid}}{\sim} \text{Normal}(\mu, \sigma^2)$ and $\gamma > \mu_i$ we have that

$$\mathbb{P}(\hat{\mu} > \gamma) = \mathbb{P}\left(\sum_{i=1}^n X_i > n\gamma\right) \leq \exp\left(-n\frac{(\gamma-\mu)^2}{2\sigma^2}\right)$$

Applying this result to this scenario

$$\mathbb{P}(\hat{\mu}_{i,n} > \gamma) \leq \exp\left(-n\frac{(\gamma-\mu_i)^2}{2}\right)$$

By defining $\gamma = \mu_i + \sqrt{\dfrac{\alpha\ln(t)}{2n}}$ with $\alpha > 0$.
Note that $\gamma > \mu_i$ so we can use the above inequality

$$\begin{aligned}
\mathbb{P}\left(\hat{\mu}_{i,n} > \mu_i + \sqrt{\frac{\alpha\ln(t)}{2n}}\right) &\leq \exp\left(-\frac{n}{2}\cdot\frac{\alpha\ln(t)}{2n}\right)\\
&= \exp\left(-\frac{\alpha\ln(t)}{4}\right)\\
&= t^{-\alpha/4}
\end{aligned}$$

Question 4. (b)
Explain in a few sentences why the same bound holds the probability of the event that $\hat{\mu}_{i,n} < \mu_i - \sqrt{\dfrac{\alpha\ln(t)}{2n}}$

Answer 4. (b)
The result from *Problem Sheet 1 6b)* is derived from the Chernoff Bound for IID random variables when $\left\{\sum X_i \geq nc\right\}$ and considers $\inf\limits_{\theta>0} e^{-n\theta c}(\mathbb{E}[e^{\theta X}])^n$. The result requires $c > \mu_i$ in order to fulfil the restriction on the infimum (i.e. $\theta > 0$).
To derive a similar result to *Question 4. (a)* for the event $\left\{\hat{\mu}_{i,n} < \mu_i - \sqrt{\frac{\alpha\ln(t)}{2n}}\right\}$ we define $c = \mu_i - \sqrt{\dfrac{\alpha\ln(t)}{2n}}$, meaning $c < \mu_i$ and thus $\theta < 0$, for the θ in the infimum.
The Chernoff Bound for this complementary event considers the infimum of the same expression, except with the restriction that $\theta < 0$ (rather than $\theta > 0$). Given our definition of c and the resulting value of θ, the same value for the infimum is found. Meaning the same bound is derived for both the event and its compliment.

Question 4. (c)
Replicate the analysis of the UCB algorithm to obtain a regret bound of the form $\mathcal{R}(T) \leq c_1 + c_2\ln(T)$ where c_1 and c_2 are constants that may depend on α, μ_1 and μ_2. Find explicit expressions for these constants.
The analysis will not work for all $\alpha > 1$. You will need α to be bigger than some other number. Find that number.

Answer 4. (c)
Assume WLOG $\mu_1 > \mu_2$ and define $\Delta = \mu_1 - \mu_2$. Let $N_2(t)$ be the number of times arm 2 is

played in the first t steps. Define $u_t = \left\lceil \frac{2\alpha\ln(t)}{\Delta^2} \right\rceil$. We have

$$N_2(t) \leq u + \sum_{s=u-1}^{t} \left(\{N_2(s-1) \geq u_t\} \text{ and } \{I(s)=j\}\right)$$

Taking expectations of both side we get

$$\mathbb{E}[N_2(t)] \leq u_t + \sum_{s=u_t}^{t-1} \mathbb{P}\left(\{N_2(s-1) \geq u_t\} \text{ and } \{I(s)=j\}\right)$$

By considering the two cases where the sub-optimal arm is played: $\hat{\mu}_1$ is significantly lower than μ_1; or $\hat{\mu}_2$ is significantly higher than μ_2.

$$\begin{array}{rcl}
\mathbb{E}[N_2(t)] & \leq & u_t + \displaystyle\sum_{s=u_t}^{t-1} \left[\mathbb{P}\left(\hat{\mu}_{1,N_1(s)} \leq \mu_1 - \sqrt{\frac{\alpha\ln(s)}{2N_2(s)}}\right) + \mathbb{P}\left(\hat{\mu}_{2,N_2(s)} > \mu_2 - \sqrt{\frac{\alpha\ln(s)}{2N_2(s)}}\right)\right] \\
& \leq & u_t + \displaystyle\sum_{s=u_t}^{t-1} 2t^{-\alpha/4} \text{ by } \textit{Question 4. (a)} \\
& \leq & u + \displaystyle\int_{u_t-1}^{\infty} 2t^{-\alpha/4}dt \\
& = & u_t + 2\left[\frac{t^{-\frac{\alpha}{4}+1}}{1-\frac{\alpha}{4}}\right]_{u_t-1}^{\infty} \\
& = & u_t - \dfrac{2(u_t-1)^{-\frac{\alpha}{4}+1}}{-\frac{\alpha}{4}+1} \\
& \leq & u_t + \dfrac{2}{\frac{\alpha}{4}-1} \\
& = & u_t + \dfrac{8}{\alpha-4} \\
& \leq & \dfrac{2\alpha\ln(t)}{\Delta^2} + 1 + \dfrac{8}{\alpha-4} \text{ by def. of } u_t \\
& = & \dfrac{2\alpha\ln(t)}{\Delta^2} + \dfrac{\alpha+4}{\alpha-4}
\end{array}$$

In this scenario $\mathcal{R}(T) = \Delta\mathbb{E}[N_2(T)]$. Thus, using the results above

$$\mathcal{R}(T) \leq \frac{2\alpha\ln(T)}{\Delta} + \Delta\frac{\alpha+4}{\alpha-4}$$

This requires $\alpha > 4$.