STAT 184, PROBLEM SET 1

Matthew Qu
Due: September 29, 2022
matthewqu@college.harvard.edu

0. (**Collaborators and Acknowledgements**)
Collaborators: Kevin Huang, Katherine Zhou
Acknowledgements:

1. (**Bias of** $\hat{\mu}_t^{(k)}$)

Proof.

1. First, we expand $\mathbb{E}(r_0)$ with LOTE, conditioning on whether r_0 is smaller or larger than r_1:

$$\mathbb{E}(r_0) = \mathbb{E}(r_0 \mid r_0 < r_1)\mathbb{P}(r_0 < r_1) + \mathbb{E}(r_0 \mid r_0 > r_1)\mathbb{P}(r_0 > r_1) + \mathbb{E}(r_0 \mid r_0 = r_1)\mathbb{P}(r_0 = r_1).$$

Since r_0 and r_1 come from continuous distributions, $\mathbb{P}(r_0 = r_1) = 0$ and the third term vanishes. Furthermore, we know that unconditionally, $\mathbb{E}(r_0) = 0$. Therefore, it follows that

$$\mathbb{E}(r_0 \mid r_0 < r_1)\mathbb{P}(r_0 < r_1) = -\mathbb{E}(r_0 \mid r_0 > r_1)\mathbb{P}(r_0 > r_1).$$

Since r_0 and r_1 are iid, we also have that $\mathbb{P}(r_0 < r_1) = \mathbb{P}(r_0 > r_1) = \frac{1}{2}$. Thus, we conclude that

$$\mathbb{E}(r_0 \mid r_0 < r_1) = -\mathbb{E}(r_0 \mid r_0 > r_1).$$

2. Once again, let's condition on whether r_0 is smaller or larger than r_1 (leaving out the third term, which will be 0):

$$\mathbb{E}(\hat{\mu}_3^{(1)}) = \frac{1}{2}\mathbb{E}(\hat{\mu}_3^{(1)} \mid r_0 < r_1) + \frac{1}{2}\mathbb{E}(\hat{\mu}_3^{(1)} \mid r_0 > r_1).$$

Now, given $r_0 < r_1$, we have $\hat{\mu}_3^{(1)} = r_0$ because the first lever will not be pulled at time $t = 2$. Similarly, if $r_0 > r_1$, then $\hat{\mu}_3^{(1)} = \frac{1}{2}(r_0 + r_2)$, with $r_2 \sim \nu$. Therefore, the above expression becomes

$$\begin{aligned}\frac{1}{2}\mathbb{E}(\hat{\mu}_3^{(1)} \mid r_0 < r_1) + \frac{1}{2}\mathbb{E}(\hat{\mu}_3^{(1)} \mid r_0 > r_1) &= \frac{1}{2}\mathbb{E}(r_0 \mid r_0 < r_1) + \frac{1}{2}\mathbb{E}(\frac{1}{2}(r_0 + r_2) \mid r_0 > r_1)\\ &= \frac{1}{2}\mathbb{E}(r_0 \mid r_0 < r_1) + \frac{1}{4}\mathbb{E}(r_0 \mid r_0 > r_1) + \frac{1}{4}\mathbb{E}(r_2 \mid r_0 > r_1).\end{aligned}$$

From part (1), we have $\mathbb{E}(r_0 \mid r_0 > r_1) = -\mathbb{E}(r_0 \mid r_0 < r_1)$. In addition, r_2 is independent from r_0 and r_1, so $\mathbb{E}(r_2 \mid r_0 > r_1) = \mathbb{E}(r_2) = 0$. Putting these together, the above expression simplifies to $\frac{1}{4}\mathbb{E}(r_0|r_0 < r_1)$. Thus, it suffices to show that $\mathbb{E}(r_0|r_0 < r_1) < 0$. To see this, consider some constant $a \in \mathbb{R}$ in the support of $r_1 \sim \nu$. (If the support of ν has upper bound b, we can also assume $a < b$ because $\mathbb{P}(r_1 = b) = 0$.) If $a < 0$, then clearly $\mathbb{E}(r_0|r_0 < a) < 0$. Otherwise, for $a \geq 0$ note that

$$0 = \mathbb{E}(r_0) = \int_{-\infty}^{\infty} x\nu(x)\,dx = \int_{-\infty}^{a} x\nu(x)\,dx + \int_{a}^{\infty} x\nu(x)\,dx = \mathbb{E}(r_0|r_0 < r_1) + \int_{a}^{\infty} x\nu(x)\,dx.$$

Note that $\int_a^\infty x\nu(x)\,dx$ has non-negative integrand because $a \geq 0$ and ν is a density. Furthermore, because a is in the support of r_1, there exists a neighborhood around r_1 such that ν is strictly positive. Therefore, $\int_a^\infty x\nu(x)\,dx > 0$, and it follows that $\mathbb{E}(r_0|r_0 < r_1) < 0$. This proves that $\mathbb{E}(\hat{\mu}_3^{(1)}) < 0$.

3. We will proceed in the same way as we did in part (2), but instead of conditioning on $r_0 < r_1$, we now have the condition $r_0 < a$, where a is a constant. We have

$$\mathbb{E}(\hat{\mu}_3^{(1)}) = \mathbb{E}(\hat{\mu}_3^{(1)} \mid r_0 < a)\mathbb{P}(r_0 < a) + \mathbb{E}(\hat{\mu}_3^{(1)} \mid r_0 > a)\mathbb{P}(r_0 > a).$$

Once again, given $r_0 < a$, we have $\hat{\mu}_3^{(1)} = r_0$, and given $r_0 > a$, we have $\hat{\mu}_3^{(1)} = \frac{1}{2}(r_0 + r_2)$. We know that $\mathbb{P}(r_0 < a) = a$, and the conditional distribution is $r_0 \mid r_0 < a \sim \mathrm{Unif}[0, a]$. Similarly,

$r_0 \mid r_0 > a \sim \text{Unif}[a, 1]$, and $r_2 \sim \text{Unif}$ independent of r_0 and r_1. Therefore, we have

$$
\begin{aligned}
\mathbb{E}(\hat{\mu}_3^{(1)}) &= \frac{a}{2} \cdot a + \frac{1}{2}\left(\frac{a+1}{2} + \frac{1}{2}\right)(1-a) \\
&= \frac{a^2}{2} + \frac{(a+2)(1-a)}{4} \\
&= \frac{a^2 - a + 2}{4}.
\end{aligned}
$$

Since $\mu^{(1)} = \frac{1}{2}$, the bias of $\hat{\mu}_3^{(1)}$ is $\frac{a^2-a}{4}$. Therefore, the arm is unbiased only in the degenerate cases $a = 0$ and $a = 1$, and downwards biased for all other $a \in (0, 1)$.

$\square$

2. (**ε-greedy algorithm**)

Proof.

1. We have $\mathbb{E}(\text{Regret}_T) = \sum_{t=0}^{T-1} \mathbb{E}(\mu^{(k^*)} - \mu^{(a_t)})$, where the expectation is taken over the randomness of a_t. Therefore, we can rewrite this as

$$\sum_{t=0}^{T-1} \mathbb{E}(\mu^{(k^*)} - \mu^{(a_t)}) = \sum_{t=0}^{T-1} \sum_{k=1}^{K} (\mu^{(k^*)} - \mu^{(k)}) \mathbb{P}(a_t = k).$$

Rewriting probabilities as expectations of indicators and rearranging terms, we get

$$\begin{aligned} \sum_{t=0}^{T-1} \sum_{k=1}^{K} (\mu^{(k^*)} - \mu^{(k)}) \mathbb{P}(a_t = k) &= \sum_{t=0}^{T-1} \sum_{k=1}^{K} (\mu^{(k^*)} - \mu^{(k)}) \mathbb{E}(\mathbf{1}_{a_t=k}) \\ &= \sum_{k=1}^{K} (\mu^{(k^*)} - \mu^{(k)}) \sum_{t=0}^{T-1} \mathbb{E}(\mathbf{1}_{a_t=k}) \\ &= \sum_{k=1}^{K} (\mu^{(k^*)} - \mu^{(k)}) \mathbb{E}(N_T^{(k)}). \end{aligned}$$

2. In the (constant) ε-greedy algorithm, at any time t, the probability of choosing any arm is at least $\frac{\varepsilon}{K}$, which is the probability of exploring multiplied by the probability of randomly choosing an arm. That is, for all t and k, we have $\mathbb{P}(a_t = k) \geq \frac{\varepsilon}{K}$. It follows that

$$\begin{aligned} \mathbb{E}(N_T^{(k)}) &= \sum_{t=0}^{T-1} \mathbb{E}(\mathbf{1}_{a_t=k}) \\ &= \sum_{t=0}^{T-1} \mathbb{P}(a_t = k) \\ &\geq \frac{\varepsilon T}{K}. \end{aligned}$$

Therefore, from part (a) we have

$$\mathbb{E}(\text{Regret}_T) \geq \sum_{k=1}^{K} \frac{\varepsilon T}{K} (\mu^{(k^*)} - \mu^{(k)}).$$

Since we assume there exists some arm k_0 for which $\mu^{(k^*)} - \mu^{(k_0)} > 0$, we can conclude that

$$\mathbb{E}(\text{Regret}_T) \geq \frac{\varepsilon(\mu^{(k^*)} - \mu^{(k_0)})T}{K} > 0,$$

and thus we can choose the constant $C = \dfrac{\varepsilon(\mu^{(k^*)} - \mu^{(k_0)})}{K}$.

3. First, let us condition on the algorithm exploring with probability ε_t or exploiting with probability $1 - \varepsilon_t$:

$$\mathbb{E}(\mu^{k^*} - \mu^{a_t}) = \mathbb{E}(\mu^{k^*} - \mu^{a_t} \mid \text{Explore})\mathbb{P}(\text{Explore}) + \mathbb{E}(\mu^{k^*} - \mu^{a_t} \mid \text{Exploit})\mathbb{P}(\text{Exploit}).$$

Given that the algorithm explores, we can trivially bound the expectation by 1. Similarly, we can bound the probability of exploiting by 1 to remove additional factor of $1 - \varepsilon_t$. Thus, we have

$$\mathbb{E}(\mu^{k^*} - \mu^{a_t}) \leq \varepsilon_t + \mathbb{E}(\mu^{k^*} - \mu^{a_t} \mid \text{Exploit}).$$

Now, consider the action a_t given that we exploit. We know that it will choose the arm with the highest empirical mean. By definition, it follows that $\hat{\mu}^{a_t} - \hat{\mu}^{k^*} \geq 0$. Therefore, we have

$$\begin{aligned}\mathbb{E}(\mu^{k^*} - \mu^{a_t} \mid \text{Exploit}) &\leq \mathbb{E}(\mu^{k^*} - \mu^{a_t} + \hat{\mu}^{a_t} - \hat{\mu}^{k^*}) \\ &= \mathbb{E}((\mu^{k^*} - \hat{\mu}^{k^*}) + (\hat{\mu}^{a_t} - \mu^{a_t})).\end{aligned}$$

We will apply the uniform Hoeffding bound on the above expression, which provides a bound on the difference between the true and empirical means for all arms at once with probability $1 - \delta$. Keep in mind that $\mu^{k^*} - \mu^{a_t} \leq 1$. Therefore, when the bound fails, we can trivially bound the resulting expectation with 1, as before. Let H be the event that the Hoeffding bound fails. Therefore, we have

$$\begin{aligned}\mathbb{E}(\mu^{k^*} - \mu^{a_t} \mid \text{Exploit}) &= \mathbb{E}(\mu^{k^*} - \mu^{a_t} \mid \text{Exploit}, H)P(H) + \mathbb{E}(\mu^{k^*} - \mu^{a_t} \mid \text{Exploit}, H^c)P(H^c) \\ &\leq \delta + \mathbb{E}((\mu^{k^*} - \hat{\mu}^{k^*}) + (\hat{\mu}^{a_t} - \mu^{a_t}) \mid \text{Exploit}, H^c) \\ &\leq \delta + \mathbb{E}\left(\sqrt{\log(2Kt/\delta)/2N_t^{(k^*)}} + \sqrt{\log(2Kt/\delta)/2N_t^{(a_t)}}\right)\end{aligned}$$

Choosing $\delta = \varepsilon_t$ and rewriting, we have

$$\mathbb{E}(\mu^{k^*} - \mu^{a_t}) \leq \varepsilon_t + \sqrt{\log(2Kt/\varepsilon_t)/2}\,\mathbb{E}\left(\frac{1}{N_t^{(k^*)}} + \frac{1}{N_t^{(a_t)}}\right).$$

We substitute this inequality to conclude that

$$\mathbb{E}(\mu^{k^*} - \mu^{a_t}) \leq 2\varepsilon_t + \sqrt{\log(2Kt/\varepsilon_t)/2}\,\mathbb{E}\left(\frac{1}{N_t^{(k^*)}} + \frac{1}{N_t^{(a_t)}}\right).$$

4. We proceed in the same way as in part (2); the only difference is that now, ε_t depends on time t. Here, at time t we have

$$\begin{aligned}\mathbb{P}(a_t = k) &= \mathbb{P}(a_t = k \mid \text{Explore})\mathbb{P}(\text{Explore}) + \mathbb{P}(a_t = k \mid \text{Exploit})\mathbb{P}(\text{Exploit}) \\ &\geq \mathbb{P}(a_t = k \mid \text{Explore})\mathbb{P}(\text{Explore}) \\ &= \frac{\varepsilon_t}{K}.\end{aligned}$$

Therefore, we have

$$\mathbb{E}(N_t^{(k)}) = \sum_{\tau=0}^{t-1} \mathbb{E}(\mathbf{1}_{a_\tau = k}) = \sum_{\tau=0}^{t-1} \mathbb{P}(a_\tau = k) \geq \sum_{\tau=0}^{t-1} \frac{\varepsilon_\tau}{K} = \frac{1}{K}\sum_{\tau=0}^{t-1} \varepsilon_\tau,$$

as desired.

5. Note that for $a \leq 0$ and $\tau \geq 0$, we have that $\frac{d}{d\tau} = a(1+\tau)^{a-2} \leq 0$, with equality only at $a = 0$. Therefore, $(a+\tau)^a$ is non-increasing with respect to τ, which implies that

$$(1+\tau)^a \geq \int_{\tau+1}^{\tau+2} x^a \, dx$$

for all $\tau \geq 0$. Therefore, we see that

$$\begin{aligned}
\sum_{\tau=0}^{t-1}(1+\tau)^a &\geq \int_1^{t+1} x^a \, dx \\
&= \frac{1}{a+1} x^{a+1} \Big|_1^{t+1} \\
&= \frac{1}{a+1} \left((t+1)^{a+1} - 1\right).
\end{aligned}$$

This proves the inequality.

6. For the choice $a = -\frac{1}{3}$, we have

$$\sum_{t=0}^{T-1} \left((1+t)^a + \frac{1}{\sqrt{(1+t)^{a+1}}} \right) = \sum_{t=0}^{T-1} \frac{2}{(1+t)^{\frac{1}{3}}}.$$

We know from part (5) that $(1+t)^{-\frac{1}{3}}$ is a decreasing function. However, this time, we will upper bound the function with another integral:

$$(1+t)^{-\frac{1}{3}} \leq \int_t^{t+1} x^{-\frac{1}{3}} \, dx.$$

Therefore, we have

$$\begin{aligned}
\sum_{t=0}^{T-1} \frac{2}{(1+t)^{\frac{1}{3}}} &\leq 2 \int_0^T (1+x)^{-\frac{1}{3}} \, dx \\
&= 2 \left(\frac{3}{2}((T+1)^{\frac{2}{3}} - 1) \right) \\
&\in \tilde{O}(T^{\frac{2}{3}}).
\end{aligned}$$

Since we have upper bounded the expression with a function that is $\tilde{O}(T^{\frac{2}{3}})$, it follows that

$$\sum_{t=0}^{T-1} \left((1+t)^a + \frac{1}{\sqrt{(1+t)^{a+1}}} \right) \in \tilde{O}(T^{\frac{2}{3}})$$

for $a = -\frac{1}{3}$.

$\square$