Prof. Dr. Leif Döring
André Ferdinand, Sara Klein

Reinforcement Learning

2. Excercise Sheet

1. Sub-Gaussian random variables

Recall Definition 1.2.3. of a σ-sub-Gaussian random variable X.

a) Show that every σ-sub-Gaussian random variable satisfies $\mathbb{E}[X] = 0$ and $\mathbb{V}[X] \leq \sigma^2$.
Solution:
Let X be a σ-sub-Gaussian random variable. Then by Fubini

$$\sum_{t\geq 0} \frac{\lambda^t}{t!}\mathbb{E}[X^t] = \mathbb{E}\Big[Xe^{\lambda X}\Big] \leq e^{\frac{\lambda^2\sigma^2}{2}} = \sum_{t\geq 0} \frac{\lambda^{2t}\sigma^{2t}}{2^t t!}. \tag{1}$$

We follow that

$$\lambda\mathbb{E}[X] + \frac{\lambda^2}{2}\mathbb{E}[X^2] \leq \frac{\lambda^2\sigma^2}{2} + g(\lambda), \tag{2}$$

for

$$g(\lambda) = \sum_{t\geq 2} \frac{\lambda^{2t}\sigma^{2t}}{2^t t!} - \sum_{t\geq 3} \frac{\lambda^t}{t!}\mathbb{E}[X^t].$$

Note that $g \in o(\lambda^2)$ because

$$\lim_{\lambda\to 0} \frac{g(\lambda)}{\lambda^2} = \sum_{t\geq 2} \lim_{\lambda\to 0} \frac{\lambda^{2t}\sigma^{2t}}{2^t t!} - \sum_{t\geq 3} \lim_{\lambda\to 0} \frac{\lambda^t}{t!}\mathbb{E}[X^t] = 0,$$

where we used that both sums are finite due to the finiteness of exp.
Finally for $\lambda > 0$ dividing (2) by $1/\lambda$ and taking the limits $\lambda \downarrow 0$ leads to

$$\mathbb{E}[X] \leq \frac{\lambda\sigma^2}{2} + \frac{g(\lambda)}{\lambda} - \frac{\lambda}{2}\mathbb{E}[X^2] \to 0, \quad \lambda \downarrow 0$$

and for $\lambda < 0$ similarly

$$\mathbb{E}[X] \geq \frac{\lambda\sigma^2}{2} + \frac{g(\lambda)}{\lambda} - \frac{\lambda}{2}\mathbb{E}[X^2] \to 0, \quad \lambda \uparrow 0.$$

Hence, $\mathbb{E}[X] = 0$.
Rewriting (2) once again and deviding by λ^2 results in

$$\mathbb{E}[X^2] \leq 2\Big(\frac{\sigma^2}{2} + \frac{g(\lambda)}{\lambda^2}\Big) \to \sigma^2, \quad \lambda \to 0,$$

which proofs the second claim.

b) Suppose X is σ-sub-Gaussian. Prove that cX is $|c|\sigma$-sub-Gaussian.

Solution:

We have

$$M_{cX}(\lambda) = \mathbb{E}\left[e^{\lambda cX}\right] \leq e^{\frac{(c\lambda)^2\sigma^2}{2}} = e^{\frac{\lambda^2(c\sigma)^2}{2}}.$$

Thus, cX is $|c|\sigma$-sub-Gaussian.

c) Show that $X_1 + X_2$ is $\sqrt{\sigma_1^2 + \sigma_2^2}$-sub-Gaussian if X_1 and X_2 are independent σ_1-sub-Gaussian and σ_2-sub-Gaussian random variables.

Solution:

We have

$$\begin{aligned}
M_{X_1+X_2}(\lambda) &= \mathbb{E}\left[e^{\lambda(X_1+X_2)}\right] = \mathbb{E}\left[e^{\lambda X_1}e^{\lambda X_2}\right] \\
&= \mathbb{E}\left[e^{\lambda X_1}\right]\mathbb{E}\left[e^{\lambda X_2}\right] \\
&\leq e^{\frac{\lambda^2\sigma_1^2}{2}} e^{\frac{\lambda^2\sigma_2^2}{2}} \\
&= \exp(\frac{\lambda^2(\sqrt{\sigma_1^2+\sigma_2^2})^2}{2}).
\end{aligned}$$

where the thrid equality follows from independence. This proofs the claim.

d) Show that a Bernoulli-variable is $\frac{1}{2}$-sub-Gaussian.

Solution:

Exactly as in the next exercise but with $a = 0$ and $b = 1$.

e) Show that every centered bounded random variable, say bounded below by a and above by b is $\frac{(b-a)}{2}$-sub-Gaussian.

Solution:

As $a \leq X \leq b$ we have almost surely

$$e^{\lambda X} \leq \frac{b-X}{b-a}e^{\lambda a} + \frac{X-a}{b-a}e^{\lambda b}.$$

We follow

$$\begin{aligned}
\mathbb{E}\left[e^{\lambda X}\right] &\leq \frac{b-\mathbb{E}[X]}{b-a}e^{\lambda a} + \frac{\mathbb{E}[X]-a}{b-a}e^{\lambda b} \\
&= \frac{b}{b-a}e^{\lambda a} - \frac{a}{b-a}e^{\lambda b} \\
&= \exp L(\lambda(b-a)),
\end{aligned}$$

where we used $\mathbb{E}[X] = 0$ and $L(h)$ is definied by

$$L(h) = \frac{ha}{(b-a)} + \log\left(1 + \frac{a-e^h a}{b-a}\right).$$

We will show that $L(h) \leq h^2/8$, then it follows

$$\mathbb{E}\left[e^{\lambda X}\right] \leq \exp L(\lambda(b-a)) \leq \exp(\frac{\lambda^2(b-a)^2}{8}),$$

which proofs that X is σ-sub-Gauss with $\sigma = \frac{(b-a)}{2}$.
So let us proof that $L(h) \leq h^2/8$. Therefore we first calculate the first and second derivative.

$$\nabla L(h) = \frac{a}{b-a} - \frac{e^h a}{b - e^h a},$$
$$\nabla^2 L(h) = -\frac{e^h ab}{(b-e^h a)^2}.$$

Note now, that

$$L(0) = 0,$$
$$\nabla L(0) = 0 \text{ and}$$
$$\nabla^2 L(h) = -\frac{e^h ab}{\underbrace{(b-e^h a)^2}_{\geq -4(be^h a)}} \leq \frac{e^h ab}{4e^h ab} \leq \frac{1}{4}.$$

By Taylor we know there exists $\theta \in [0,1]$ such that

$$L(h) = L(0) + h\nabla L(0) + \frac{1}{2}h^2 \nabla^2 L(h\theta) = \frac{1}{2}h^2 \nabla^2 L(h\theta).$$

As $\nabla^2 L(h) \leq \frac{1}{4}$, we have

$$L(h) \leq \frac{1}{2}h^2 \frac{1}{4} = \frac{h^2}{8}.$$

This conclues the proof.

2. Regret Bound

Recall the upper bound on the regret for ETC in the case of two arms from the first exercise sheet. Show that

$$R_n(\pi) \leq \Delta + C\sqrt{n}$$

for some model-free constant C so that, in particular, $R_n(\pi) \leq 1 + C\sqrt{n}$ for all bandit models with regret bound $\Delta \leq 1$ (for instance for Bernoulli bandits).
Hint: Use the same trick as in the proof of Theorem 1.2.10.
Solution:
We will first show that

$$R_n(\pi) \leq \min\{n\Delta, \Delta + \frac{4}{\Delta}\Big(1 + \max\{0, \log(\frac{n\Delta^2}{4})\}\Big)\}$$

by plugging $m^* = \max\{1, \lceil \frac{4}{\Delta^2} \log(\frac{n\Delta^2}{4})\rceil\}$ into the regret bound from the last exercise sheet

$$R_n \leq m^*\Delta + (n - 2m^*)\Delta \exp(-\frac{m^*\Delta^2}{4}).$$

This leads to

$$\begin{aligned}
R_n &\le m^*\Delta + (n-2m^*)\Delta \exp(-\frac{\Delta^2}{4}\max\{1, \lceil \frac{4}{\Delta^2}\log(\frac{n\Delta^2}{4})\rceil\}) \\
&= m^*\Delta + (n-2m^*)\Delta \min\{\exp(-\frac{\Delta^2}{4}), \underbrace{\exp(-\frac{\Delta^2}{4}\lceil \frac{4}{\Delta^2}\log(\frac{n\Delta^2}{4})\rceil)}_{\le \exp(-\frac{\Delta^2}{4}\frac{4}{\Delta^2}\log(\frac{n\Delta^2}{4}))\le \frac{4}{\Delta^2 n}}\} \\
&\le m^*\Delta + \min\{(n-2m^*)\Delta\exp(-\frac{\Delta^2}{4}), (n-\underbrace{2m^*}_{>0})\Delta\frac{4}{\Delta^2 n}\} \\
&\le m^*\Delta + \min\{(n-2m^*)\Delta\exp(-\frac{\Delta^2}{4}), \frac{4}{\Delta}\} \\
&\le \min\Big\{m^*\Delta + (n-2m^*)\Delta\underbrace{\exp(-\frac{\Delta^2}{4})}_{\le 1}, \frac{4}{\Delta} + \underbrace{\max\{1, \lceil \frac{4}{\Delta^2}\log(\frac{n\Delta^2}{4})\rceil\}\Delta}_{\le (1+\max\{0, \frac{4}{\Delta^2}\log(\frac{n\Delta^2}{4})\})\Delta}\Big\} \\
&\le \min\Big\{\underbrace{-m^*\Delta}_{\le 0} + n\Delta, \Delta + \frac{4}{\Delta}\big(1+\max\{0, \log(\frac{n\Delta^2}{4})\}\big)\Big\} \\
&\le \min\Big\{n\Delta, \Delta + \frac{4}{\Delta}\big(1+\max\{0, \log(\frac{n\Delta^2}{4})\}\big)\Big\}.
\end{aligned}$$

Using this we can devide in the cases $\Delta \le \sqrt{\frac{c}{n}}$ *and* $\Delta > \sqrt{\frac{c}{n}}$, *for some constant* $c > 0$ *which we specify later. Thus, in the first case* $\Delta \le \sqrt{\frac{c}{n}}$ *we have*

$$R_n \le \min\Big\{n\Delta, \Delta + \frac{4}{\Delta}\big(1+\max\{0, \log(\frac{n\Delta^2}{4})\}\big)\Big\} \le n\Delta \le \sqrt{cn}.$$

For the second case we consider the second term and rewrite

$$\frac{4}{\Delta}\big(1+\max\{0, \log(\frac{n\Delta^2}{4})\}\big) \le 4(\frac{1}{\Delta} + \frac{\log(\frac{n\Delta^2}{4})}{\Delta}).$$

We define $f(x) = \frac{\log(\frac{nx^2}{4})}{x}$, *and prove* $f(x) \le 2$ *for* $x \ge \sqrt{\frac{e^2 4}{n}}$. *If this is true we have for the second case with* $c = e^2 4$ *that*

$$\begin{aligned}
R_n &\le \Delta + 4(\frac{1}{\Delta} + \frac{\log(\frac{n\Delta^2}{4})}{\Delta}) \\
&\le \Delta + 4(\sqrt{\frac{n}{c}} + 2) \le \Delta + \sqrt{n}(8 + \frac{4}{\sqrt{c}}) = \Delta + \sqrt{n}(8 + \frac{2}{e}).
\end{aligned}$$

Now to our claim. We have

$$f'(x) = \frac{2 - \log(\frac{nx^2}{4})}{x^2}$$

and so $f'(x) \le 0$ *iff*

$$\log(\frac{nx^2}{4}) \ge 2 \quad \Leftrightarrow \quad x \ge \sqrt{\frac{e^2 4}{n}}.$$

Thus f *decreases in* $[\sqrt{\frac{e^2 4}{n}}, \infty)$ *and so* $f(x) \le f(\sqrt{\frac{e^2 4}{n}}) = 2$.
Coosing $C = 8 + \frac{2}{e}$ *concludes the proof, as for the first case with* $c = e^2 4$ *we have* $R_n \le 2e\sqrt{n} \le \Delta + C\sqrt{n}$ *and for the second case also* $R_n \le \Delta + C\sqrt{n}$.

3. Upper bound on $\hat{Q}_a(t)$ for many samples

Suppose ν is a bandit model with 1-sub-gaussian arms. Show that under the UCB Algorithm $\hat{Q}_a(t) < Q_a + \Delta_a$ with probability $1-\delta$, given that $T_a(t) > \frac{2\log(1/\delta)}{\Delta_a^2}$.

Hint: Proof a generalized Hoeffding's inequality:

Suppose $X_1, X_2, \dots$ are iid random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with expectation μ such that X_1 is σ-sub-Gaussian. Assume further $T : \Omega \to \{1, 2, 3, \dots\}$ is a discrete random variable, almost surely finite, on the same probability space and independent of $X_1, X_2, \dots$.

Then it holds:

$$\mathbb{P}\Big(\frac{1}{T}\sum_{n=1}^{T} X_n - \mu \geq \sqrt{\frac{2\log(1/\delta)}{T}}\Big) \leq \delta.$$

Solution:

Proof: First we proof the generalized Hoeffding inequality. Assume $X_1, X_2, \dots$ are iid random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with expectation μ such that X_1 is σ-sub-Gaussian. Assume further $T : \Omega \to \{1, 2, 3, \dots\}$ is a discrete random variable, almost surely finite, on the same probability space and independent of $X_1, X_2, \dots$. Then by σ-additivity and independence we have

$$\begin{aligned}
&\mathbb{P}\Big(\frac{1}{T}\sum_{n=1}^{T} X_n - \mu \geq \sqrt{\frac{2\log(1/\delta)}{T}}\Big)\\
&= \mathbb{P}\Big(\Big\{\frac{1}{T}\sum_{n=1}^{T} X_n - \mu \geq \sqrt{\frac{2\log(1/\delta)}{T}}\Big\} \cap \Big\{\biguplus_{n=1}^{\infty}\{T = n\}\Big\}\Big)\\
&= \sum_{n=1}^{\infty}\mathbb{P}\Big(\Big\{\frac{1}{T}\sum_{n=1}^{T} X_n - \mu \geq \sqrt{\frac{2\log(1/\delta)}{T}}\Big\} \cap \Big\{T = n\Big\}\Big)\\
&= \sum_{n=1}^{\infty}\mathbb{P}(T = n)\,\mathbb{P}\Big(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu \geq \sqrt{\frac{2\log(1/\delta)}{n}}\Big)\\
&\leq \sum_{n=1}^{\infty}\mathbb{P}(T = n)\delta\\
&= \delta
\end{aligned}$$

where we used the normal Hoeffings inequality in the inequality.

Now note that $T_a(t) > \frac{2\log(1/\delta)}{\Delta_a^2}$ is equivalent to

$$\Delta_a > \sqrt{\frac{2\log(1/\delta)}{T_a(t)}}.$$

By the definition of $\hat{Q}_a(t) = \frac{1}{T_a(t)}\sum_{i=1}^{T_a(t)} Y_i$, ($(Y_i)$ iid realisation of Arm a), and as $T_a(t)$ is *independent of the rewards* $(\mathrm{Y}_i)_{i\in\mathbb{N}}$ *we have*

$$\begin{aligned}
&\mathbb{P}^\pi\Big(\big(\hat{Q}_a(t) < Q_a + \Delta_a\big) \cap \big(T_a(t) > \frac{2\log(1/\delta)}{\Delta_a^2}\big)\Big) \\
&\geq \mathbb{P}^\pi\Big(\hat{Q}_a(t) - Q_a < \sqrt{\frac{2\log(1/\delta)}{T_a(t)}}\Big) \\
&= 1 - \mathbb{P}^\pi\Big(\hat{Q}_a(t) - Q_a \geq \sqrt{\frac{2\log(1/\delta)}{T_a(t)}}\Big) \\
&\geq 1 - \delta.
\end{aligned}$$