Homework 1
CSE 599i: Interactive Learning
Instructor: Kevin Jamieson
Due 11:59 PM on January 24, 2020

Probability
Concentration inequalities are at the heart of most arguments in statistical learning theory and bandits. Refer to [1] for more details.

1.1 (Markov’s Inequality) Let \(X\) be a positive random variable. Prove that \(\Pr(X > \lambda) \leq \frac{\mathbb{E}[X]}{\lambda}\).

1.2 (Jensen’s Inequality) Let \(X\) be a random vector in \(\mathbb{R}^d\) and let \(\phi : \mathbb{R}^d \to \mathbb{R}\) be convex. Then \(\phi(\mathbb{E}[X]) \leq \mathbb{E}[\phi(X)]\). Show this inequality for the special case when \(X\) has discrete support. That is, for \(p_i \geq 0\) and \(\sum_{i=1}^{n} p_i = 1\), and \((x_1, \ldots, x_n) \in \mathbb{R}^n\) show that \(\phi(\sum_{i=1}^{n} p_i x_i) \leq \sum_{i=1}^{n} p_i \phi(x_i)\).

1.3 (Hoeffding’s Lemma) Let \(X\) be a random variable with \(\mathbb{E}[X] = 0\) and \(X \in [a, b]\) almost surely. Show that for any \(\lambda \geq 0\), \(\log(\Pr[e^{\lambda X}]) \leq \frac{\lambda^2 (b-a)^2}{8}\). Hint\(^2\)

1.4 (Sub-exponential concentration) For \(i = 1, \ldots, n\) let \(X_i\) be an independent, positive random variable that satisfies \(\Pr(X_i > t) \leq e^{-t/a_i}\) for some \(a_i \in \mathbb{R}^n_+\). Show that there exists a universal constant \(c > 0\) such that \(\Pr(\sum_{i=1}^{n} (X_i - a_i) \geq t) \leq \exp(-c \min\{\frac{t}{\|a\|_2^2}, \frac{t}{\|a\|_{\infty}}\})\).

The Upper Confidence Bound Algorithm.
Consider the following algorithm for the multi-armed bandit problem.

\[\text{Algorithm 1: UCB}\]
\[
\text{Input:} \quad \text{Time horizon } T, \text{ 1-subGaussian arm distributions } P_1, \ldots, P_n \\
\text{Initialize:} \quad \text{At any time let } T_i(t) \text{ denote the number of times } i \text{ has been pulled at time } t \text{ and let } T_i = T_i(T). \text{ Pull each arm once.} \\
\text{for:} \quad t = n + 1, \ldots, T \\
\quad \text{Choose } I_t = \arg \max_{i=1, \ldots, n} \hat{\mu}_i, T_i(t) + \sqrt{\frac{2 \log(2nT^2)}{T_i(t)}} \\
\quad \text{Observe } X_{I_t,t} \sim P_{I_t}. \quad T_i(t+1) \leftarrow T_i(t) + 1, \text{ update } \hat{\mu}_{I_t, T_i(t+1)}
\]

In the following exercises, we will compute the regret of the UCB algorithm and show it matches the regret bound from lecture. Without loss of generality, assume that the best arm is \(\mu_1\). For any \(i \in [n]\), define the sub-optimality gap \(\Delta_i = \mu_1 - \mu_i\). Define the regret at time \(T\) as \(R_T = \mathbb{E}[\sum_{i=1}^{T} \mu_n^* - \mu_i] = \sum_{i=1}^{n} \Delta_i \mathbb{E}[T_i]\).

2.1 Consider the event 
\(\mathcal{E} = \bigcap_{i \in [n]} \bigcap_{s \leq T} \left\{ |\hat{\mu}_{i,s} - \mu_i| \leq \sqrt{\frac{2 \log(2nT^2)}{s}} \right\} \).
Show that \(\Pr(\mathcal{E}) \geq 1 - \frac{1}{T}\).

2.2 Conditioned on event \(\mathcal{E}\), show that \(T_i < \frac{8 \log(2nT^2)}{\Delta_i^2}\) for \(i \neq 1\).

2.3 Show that \(\mathbb{E}[T_i] \leq \frac{8 \log(2nT^2)}{\Delta_i^2} + 1\). When \(n \leq T\), conclude by showing that \(R_T \leq \sum_{i=1}^{n} \left(\frac{24 \log(2T)}{\Delta_i} + \Delta_i\right)\).

Thompson Sampling.
We consider the following Bayesian setting. Assume that we have access to \(n\) arm distributions \(\nu_1(\theta), \ldots, \nu_n(\theta)\)

\(^1\)Last updated to correct for typos January 13, 2020 at 9:17 AM
\(^2\)For any \(X \in [a, b]\) we can write \(X = (1-p)a + pb\) for \(p = \frac{X-a}{b-a} \in [0,1]\). Apply Jensen’s inequality. This can be tricky feel free to get as far as you can.
each supported on $[-1,1]$, where $\theta \in \mathbb{R}$ is a real parameter, and the mean of the $i$-th distribution is $\mu_i(\theta)$. We also assume access to a prior $p_0(\theta)$.

### Algorithm 1: Thompson Sampling

**Input:** Time horizon $T$, arm distributions $\nu_1, \ldots, \nu_n$

Let $p_t(\cdot | I_t, X_{i,1}, \ldots, I_{t-1}, X_{i,t-1})$ be the posterior distribution on $\theta$ at time $t$.

**for:** $t = 1, \ldots, T$

Sample $\theta_t \sim p_t$

Choose $I_t = \arg \max_{1 \leq i \leq n} \mu_i(\theta_t)$

Observe $X_{I_t, t} \sim \nu_{I_t}$. $T_{I_t}(t+1) \leftarrow T_{I_t}(t) + 1$

Compute exact posterior update $p_t$

Denote the $\sigma$-algebra generated by the observations at time $t$ by $\mathcal{F}_t = \sigma(I_1, X_{i,1}, \ldots, I_{t-1}, X_{i,t-1})$ (if you are unfamiliar with $\sigma$-algebras, don’t worry too much - conditioning on the $\sigma$-algebra just means conditioning on the choices of arms and the rewards observed). The *Bayesian Regret* of an algorithm is

$$BR_T = \mathbb{E}_{\theta \sim p_0} \left[ \sum_{t=1}^{T} \mu^* - \mu_{I_t}(\theta) \right]$$

where $i^* = \arg \max_{1 \leq i \leq n} \mu_i(\theta)$ (it’s a random variable depending on $\theta \sim p_0(\theta)$).

3.1 Let the good event be

$$\mathcal{E} = \bigcap_{i=1}^{n} \bigcap_{t \leq T} \left\{ |\hat{\mu}_{i,t}(t) - \mu_i| \leq \sqrt{\frac{2 \log(2/\delta)}{t}} \right\}$$

Show that $\mathbb{P}(\mathcal{E}^c) \leq nT\delta$.

3.2 (Key idea.) Argue that $\mathbb{P}(i^* = |\mathcal{F}_{t-1}) = \mathbb{P}(I_t = \mathcal{F}_{t-1})$.

3.3 Define $U_t(i) = \hat{\mu}_{i,t}(t) + \sqrt{\frac{2 \log(2/\delta)}{t}}$. Using the above, show that $\mathbb{E}[\mu_{i^*} - \mu_{I_t} | \mathcal{F}_{t-1}] = \mathbb{E}[\mu_{i^*} - U_t(i^*) | \mathcal{F}_{t-1}] + \mathbb{E}[U_t(I_t) - \mu_{I_t} | \mathcal{F}_{t-1}]$. Conclude that $BR_T = \mathbb{E}[(\sum_{t=1}^{T} \mu_{i^*} - U_t(i^*)) + \sum_{t=1}^{T} U_t(I_t) - \mu_{I_t}]$. Hint$^3$.

3.4 Show that $BR_T \leq \delta T + \mathbb{E} \left[ \sum_{t=1}^{T} U_t(I_t) - \mu_{I_t} \right] \leq O(\delta T + \sqrt{Tn \log(1/\delta)})$. Hint$^4$

3.5 Make an appropriate choice of $\delta$ and state a final regret bound.

In general, giving frequentist bounds on the regret is significantly more difficult. We refer the interested reader to [2] and the tutorial [3] for more details. This exercise is motivated by [4].

---

$^3$Tower rule of expectation.

$^4$Apply Jensen’s to $\sum_{t=1}^{T} \sqrt{T_t}$. 

---

2
Empirical Experiments
Implement UCB, Thompson Sampling (TS), and Explore-then-Commit (ETC). Let \( P_i = \mathcal{N}(\mu_i, 1) \) for \( i = 1, \ldots, n \).

4.1 Let \( n = 10 \) and \( \mu_1 = 0.1 \) and \( \mu_i = 0 \) for \( i > 1 \). On a single plot, for an appropriately large \( T \) to see expected effects, plot the regret for the UCB, TS, and ETC for several values of \( m \).

4.2 Let \( n = 40 \) and \( \mu_1 = 1 \) and \( \mu_i = 1 - 1/\sqrt{i-1} \) for \( i > 1 \). On a single plot, for an appropriately large \( T \) to see expected effects, plot the regret for the UCB, TS, and ETC for several values of \( m \).

Lower Bounds on Hypothesis Testing
Consider \( n \) samples \( X_1, \ldots, X_n \sim P \) where \( P \in \{ P_0, P_1 \} \) (assume for simplicity that these are probability distributions on \( \mathbb{R} \)). A hypothesis test for \( H_0 : P = P_0, H_1 : P = P_1 \) is a function \( \phi(x_1, \ldots, x_n) : \mathbb{R}^n \to \{0, 1\} \) that takes the data as input and returns the null or the alternative. Assume that the density of \( P_i \) exists (think: \( p_i(x) = \frac{1}{\sqrt{2\pi \sigma_i^2}} e^{-(x-\mu_i)^2/2\sigma_i^2} \)). In this problem, we will lower bound the number of samples needed by any hypothesis test on a fixed number of samples. Convince yourself, at least intuitively, that any best-arm identification algorithm for two arms will take at least as many samples as this hypothesis test takes.

5.1 Show \( \inf_\phi \max \{ \mathbb{P}_0(\phi = 1), \mathbb{P}_1(\phi = 0) \} \geq \frac{1}{2} \int_{\mathbb{R}^n} \min(p_0(x), p_1(x)) dx \). Hint\(^5\).

5.2 Let’s continue on. Show \( \frac{1}{2} \int_{x \in \mathbb{R}} \min(p_0(x), p_1(x)) dx \geq \frac{1}{4} \left( \int_{x \in \mathbb{R}} \sqrt{p_0(x)p_1(x)} dx \right)^2 \). Hint\(^6\). 

5.3 One more step. Show \( \left( \int_{x \in \mathbb{R}} \sqrt{p_0(x)p_1(x)} dx \right)^2 \geq \exp \left( -\int_{x \in \mathbb{R}} \log \left( \frac{p_1(x)}{p_0(x)} \right) p_1(x) dx \right) \). Hint\(^7\).

5.4 The final quantity is known as the KL-Divergence between distributions. Now assume that \( P_0 = \mathcal{N}(\mu_01_n, I_n) \) and \( P_1 = \mathcal{N}(\mu_11_n, I_n) \) where \( I_n \) is the \( n \times n \) identity matrix and \( 1_n \in \mathbb{R}^n \) is the all ones vector. Show (or look up) \( KL(P_0||P_1) \).

5.5 Conclude that to achieve a test with a probability of error less than \( \delta \), then we necessarily have \( n \geq 2\Delta^{-2} \log(1/4\delta) \).

Remark: The art of lower bounds is well established and extensive in statistics. See [5] for more details in the hypothesis testing setting. In the bandit setting, see [6].

References

\(^5\) Bound the max below by the average
\(^6\) Note that for \( a, b > 0 \) we have \( ab = \min\{a, b\} \max\{a, b\} \).
\(^7\) Apply Cauchy Schwartz.
\(^8\) Jensen’s inequality.