Homework 2

CSE 599i: Interactive Learning
Instructor: Kevin Jamieson
Due 11:59 PM on February 17, 2020

Confidence bounds

1.1 (Wald's identity) Let $X_1, X_2, \dots$ be a sequence of iid random variables. For $j \in \{0,1\}$, under $\mathbf{H}_j$ we have that $X_i \sim p_j$. Let $\mathbb{P}_j(\cdot), \mathbb{E}_j(\cdot)$ denote the probability and expectation under $\mathbf{H}_j$. Assume that the support of p_0 and p_1 are equal and furthermore, that $\sup_{x \in \text{support}(p_0)} \frac{p_1(x)}{p_0(x)} \le \kappa$. Fix some $\delta \in (0,1)$. If $L_t = \prod_{i=1}^t \frac{p_1(X_i)}{p_0(X_i)}$ and $\tau = \min\{t : L_t > 1/\delta\}$, we showed in class that the false alarm probability $\mathbb{P}_0(L_\tau > 1/\delta) \le \delta$. Assume that $\mathbb{E}_1[\tau] \le \infty$. Show that $\frac{\log(1/\delta)}{KL(p_1||p_0)} \le \mathbb{E}_1[\tau] \le \frac{\log(\kappa/\delta)}{KL(p_1||p_0)}$ where $KL(p_1|p_0) = \int p_1(x)\log(\frac{p_1(x)}{p_0(x)})dx$ is the Kullback Leibler divergence between p_1 and p_0.

1.2 (Method of Mixtures) Let $X_1, X_2, \dots$ be a sequence of iid random variables where $X_1 \sim \mathcal{N}(\mu, 1)$. Under $\mathbf{H}_0$ we have $\mu = 0$ and under $\mathbf{H}_1$ assume $\mu = \theta$.

- Define $L_t(\theta) := \exp(S_t\theta - t\theta^2/2)$ where $S_t = \sum_{s=1}^t X_s$. Show that $\prod_{i=1}^t \frac{p_1(X_i)}{p_0(X_i)} = L_t(\theta)$.
- Now suppose the sequence $X_1, X_2, \dots$ is still iid and $\mathbb{E}[X_1] = \mu$ in the above notation, but now the distribution of X_1 is unknown other than the knowledge that $\mathbb{E}[\exp(\lambda(X_1 - \mu))] \le e^{\lambda^2/2}$. Show that $L_t(\theta)$ defined above is a super-martingale under $\mathbf{H}_0$.
- Assume the setting of the previous step. Define $\bar{L}_t = \int_\theta L_t(\theta)h(\theta)d\theta$ where $h(\theta) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{\theta^2}{2\sigma^2}}$ and $\sigma > 0$. Show that $\bar{L}_t$ is a super-martingale under $\mathbf{H}_0$.
- Fix $\delta \in (0,1)$. Show that $\mathbb{P}_0(\exists t \in \mathbb{N} : \bar{L}_t > 1/\delta) \le \delta$.
- Conclude that if $Z_1, Z_2, \dots$ are iid random variables with $\mathbb{E}[\exp(\lambda(Z_1 - \mathbb{E}[Z_1])] \le \exp(\lambda^2/2)$, then for any $\sigma > 0$

$$\mathbb{P}\left(\exists t \in \mathbb{N} : |\frac{1}{t}\sum_{s=1}^t (Z_s - \mathbb{E}[Z_s])| > \sqrt{1 + \frac{1}{t\sigma^2}}\sqrt{\frac{2\log(1/\delta) + \log(t\sigma^2+1)}{t}}\right) \le \delta. \tag{1}$$

1.3 Problem 20.10 of [SzepesvariLattimore].

1.4 (Best arm identification) Consider an n-armed multi-armed bandit problem where the jth pull of the ith arm yields a random variable $X_{i,j} \sim \mathcal{N}(\theta_i, 1)$. The objective of the player is to strategically pull arms until getting to a point that they can predict the index of the arm with the highest mean, at which time they stop and output this estimated arm. Suppose an oracle told the player that θ, the true means they are playing against, is equal to $\Delta\mathbf{e}_j$ for some $j = 1, \dots, n$ where $\mathbf{e}_j$ is a vector of all zeros except a 1 in the jth location. Note that while $\Delta > 0$ is known to the player, which is the true j is unknown.

- Due to the symmetry of the known problem setup, it is conceivable that the following algorithm is optimal: play every arm the same number of times (say, τ times), and then declare that the arm with the highest empirical mean is best. Provide a sufficient condition on τ such that such a procedure correctly identifies the location of the true j index with probability at least $1 - \delta$.
- Argue that if each arm is pulled the same number of times this value of τ up to a constant factor (ignoring dependence on δ) is necessary. Hint[1]
- The previous two parts suggest that any algorithm that pulls every arm the same number of times and identifies the best arm requires essentially $n\Delta^{-2}\log(n/\delta)$ total pulls. We will beat this with an adaptive procedure that requires just $O(n\Delta^{-2}\log(1/\delta))$ pulls.

  Algorithm: Initialize $S_1 = [n]$. At each around $t \ge 1$, while $|S_t| > 1$, pull every arm in S_t and then set $S_{t+1} = \{i \in S_t : \sum_{s=1}^t (X_{i,s} - \Delta) \ge -\Delta t/2 - \frac{\log(1/\delta)}{\Delta}\}$.

[1] If $Z_i \sim \mathcal{N}(0, \sigma^2)$ for $i = 1, \dots, n$ show that $\mathbb{P}(\max_{i=1,\dots,n} Z_i \ge \sqrt{2\sigma^2\log(n)}) > c$ for some absolute constant c.

We showed in class that if $Z_s \sim \mathcal{N}(0,1)$ then for any $\alpha > 0$ and $\rho \in (0,1)$ we have

$$\max\left\{\mathbb{P}\left(\bigcup_{t=1}^{\infty}\{\sum_{s=1}^{t} Z_s < -\frac{\alpha t}{2} - \frac{\log(1/\rho)}{\alpha}\}\right), \mathbb{P}\left(\bigcup_{t=1}^{\infty}\{\sum_{s=1}^{t} Z_s > \frac{\alpha t}{2} + \frac{\log(1/\rho)}{\alpha}\}\right)\right\} \leq \rho. \tag{2}$$

Conclude that if j is the index of the best-arm (so that $X_{j,s} \sim \mathcal{N}(\Delta, 1)$) then with probability at least $1-\delta$ arm j remains in S_t for all $t \geq 1$.

- For $i \neq j$ (so that $X_{i,s} \sim \mathcal{N}(0,1)$) define the random variables

$$\rho_i := \sup\left\{\rho \in (0,1) : \bigcap_{t=1}^{\infty}\{\sum_{s=1}^{t} X_{i,s} \leq \frac{\Delta t}{4} + \frac{\log(1/\rho)}{\Delta/2}\}\right\}.$$

If $T_i = \max\{t : i \in S_t\}$ show that $T_i \leq 4\Delta^{-2}\log(1/\delta) + 8\Delta^{-2}\log(1/\rho_i)$.

- Note that by (2) we have $\mathbb{P}(\rho_i \leq \rho) \leq \rho$ for any $\rho \in (0,1)$. Use this fact to show that with probability at least $1-\delta$ we have $\sum_{i=1}^{n} T_i \leq cn\Delta^{-2}\log(1/\delta)$ for some absolute constant $c > 0$. Hint[2].

In general, when the means are unknown, curved boundaries with a UCB-like algorithm [1] can be used to identify the best arm using just $O(\log(1/\delta)\sum_{i=2}^{n}\Delta^{-2}\log(\log(\Delta^{-2})))$ total pulls, where $\Delta_i = \max_j \theta_j - \theta_i$ using a similar analysis.

Linear regression and experimental design
2.1 Exercise 20.2 of [SzepesvariLattimore]

Non-parametric bandits
4.1 Let $\mathcal{F}_{Lip}$ be a set of functions defined over $[0,1]$ such that for each $f \in \mathcal{F}_{Lip}$ we have $f : [0,1] \to [0,1]$ and for every $x, y \in [0,1]$ we have $|f(y) - f(x)| \leq L|y - x|$ for some known $L > 0$. At each round t the player chooses an $x_t \in [0,1]$ and observes a random variable $y_t \in [0,1]$ such that $\mathbb{E}[y_t] = f_*(x_t)$ where $f_* \in \mathcal{F}_{Lip}$. Define the regret of an algorithm after T steps as $R_T = \mathbb{E}\left[\sum_{t=1}^{T} f_*(x_\star) - f_*(x_t)\right]$ where $x_\star = \arg\max_{x\in[0,1]} f_*(x)$.

- Propose an algorithm, that perhaps uses knowledge of the time horizon T, that achieves $R_T \leq O(T^{2/3})$ regret (Okay to ignore constant, log factors).
- Argue that this is minimax optimal (i.e., unimprovable in general through the use of an explicit example, with math, but no formal proof necessary).

Experiments
5.1 Suppose we have random variables $Z_1, Z_2, \ldots$ that are iid with $\mathbb{E}[\exp(\lambda(Z_1 - \mathbb{E}[Z_1])] \leq \exp(\lambda^2/2)$. Then for any *fixed* $t \in \mathbb{N}$ we have the standard tail bound

$$\mathbb{P}\left(|\frac{1}{t}\sum_{s=1}^{t}(Z_s - \mathbb{E}[Z_s])| > \sqrt{\frac{2\log(2/\delta)}{t}}\right) \leq \delta. \tag{3}$$

The bound in (1) holds for all $t \in \mathbb{N}$ simultaneously (i.e., not for just a fixed t) at the cost of a slightly inflated bound. Let $\delta = 0.05$. Plot the *ratio* of the confidence bound of (1) to (3) as a function of t for values of $\sigma^2 \in \{10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 10^{0}\}$. What do you notice about where the ratio is smallest with respect to σ^2? Suppose you are observing a stream of iid Gaussian random variables $Z_1, Z_2, \ldots$ and you are trying to determine whether their mean is positive or negative. If someone tells you they think the absolute value of the mean is about .01, how would you choose σ^2 and use the above confidence bound to perform this test using as few total observations as possible?

5.2 Problem 21.1 of [SzepesvariLattimore].

[2] If $\mathbb{P}(U \leq u) \leq u$ then $\mathbb{P}(a_i \log(1/U) \geq \epsilon) \leq e^{-\epsilon/a_i}$. Use the sub-exponential tail-bound of homework 1.

5.3 Problem 21.7 of [SzepesvariLattimore].

5.4 Consider regret minimization for linear bandits with a fixed arm set $\mathcal{X}$ (i.e., it does not change over time), θ_* with $\|\theta_*\|_2 \leq 1$, and observations $y_t = \langle x_t, \theta_* \rangle + \epsilon_t$ where $\epsilon_t \sim \mathcal{N}(0,1)$. Implement the Thompson Sampling algorithm with a Gaussian prior with mean 0 and identity covariance (Section 36.3), LinUCB algorithm (Section 19.2 of text), and the elimination algorithm that uses G-optimal design (Section 22.0). For various values of n and $d \leq n$, construct $\mathcal{X}$ by drawing n vectors uniformly at random from $\{x \in \mathbb{R}^d : \|x\|_2 = 1\}$[3]. Use $\theta_* = \mathbf{e}_1$ for all experiments. Qualitatively describe the relative performance of the algorithms as a function of n and d. Use plots of the empirical regret to justify your claims.

References

[1] Kevin Jamieson, Matthew Malloy, Robert Nowak, and Sébastien Bubeck. lilucb: An optimal exploration algorithm for multi-armed bandits. In *Conference on Learning Theory*, pages 423–439, 2014.

[3] If $Z \sim \mathcal{N}(0, I_d)$ then $Z/\|Z\|_2$ is uniformly distributed on the unit sphere