EE5603: Concentration Inequalities **2024**

Exam 1: August 2024

Instructor: Shashank Vatedka

Instructions: *This is a closed-book exam. You are not permitted to refer to any material or discuss the problem with anyone. Malpractice will be severely punished. Please mention your ROLL Number and name clearly in the answer sheet.*

Justify all your statements clearly. You may use any result proved in class (but clearly state which results you are using), but everything else needs to be proved.

Question 1.1. Let X be a non-negative random variable with finite mean. Prove that

5 marks

$$\mathbb{E}[X] = \int_0^{\infty} \Pr[X > u]du$$

Question 1.2 (10+5pts). Let X be a Poisson random variable with mean $\alpha > 0$, i.e.,

$$p_X(i) = \frac{\alpha^i e^{-\alpha}}{i!}, \text{ for } i = 0, 1, \ldots$$

Find the moment generating function of X. Using this, find the best Chernoff upper bound for

$$\Pr[X - \alpha \geqslant \delta]$$

for any $\delta > 0$, and show that this bound decays as $c_1 e^{c_2 \delta}$ where c_1, c_2 are constants independent of δ.d

Question 1.3 (10+5pts). For $\sigma^2 > 0$, let

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/(2\sigma^2)}, \quad x \in \mathbb{R}$$

Compute the moment generating function $\mathbb{E}e^{tX}$ from first principles.

Using the MGF, compute $\mathbb{E}[X]$.

You must justify all your steps.

Question 1.4 (5+10pts). Suppose that we want to estimate the mean of a distribution from iid samples $X_1, \ldots, X_n$. We want to get estimates that have an error at most ϵ, i.e., if $f(X_1, \ldots, X_n)$ is one estimate, then we want $f(X_1, \ldots, X_n) \in (\mu - \epsilon, \mu + \epsilon)$ with probability at least $1 - \delta$. Here, ϵ, δ are parameters that decide the quality of the estimate.

Design an estimator that achieves $\delta = 3/4$ and arbitrary $\epsilon > 0$ using $n = c_1\sigma^2/\epsilon^2$ samples, where σ^2 is the variance of X_1 and c_1 is a universal constant.

Design an estimator that achieves arbitrary $\delta > 0, \epsilon > 0$ this using

$$n = c\frac{\sigma^2}{\epsilon^2} \log \frac{1}{\delta},$$

samples, where σ^2 is the variance of X_1 and c is a universal constant.

You may use any of the following results without proof (but state which result you use):

- Any of the concentration inequalities or bounds proved in class
- (B1) If $X_1, \ldots, X_n$ are iid Bernoulli(p) random variables, then for $0 < \delta < 1$

$$\Pr\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i - p\right| \geqslant \delta\right] \leqslant 2e^{-np\delta^2/3}$$

- (B2) If $X_1, \ldots, X_n$ are iid random variables where each X_i is 1 with probability 0.5 and -1 with probability 0.5, then for any $\delta > 0$ and $\underline{a} \in \mathbb{R}^n$,

$$\Pr\left[|\sum_{i=1}^{n} X_i| \geqslant \delta\right] \leqslant 2\exp\left(-\frac{\delta^2}{2\|\underline{a}\|_2^2}\right)$$