Homework 3

HDP KNU+ Fall 2022

Hints are in the back of this homework set.

As in the previous homework, $C, C_1, C_2, \dots$ and $c, c_1, c_2, \dots$ denote *positive absolute constants of your choice*. Thus, whenever you see C_1, you can replace it any positive constant you like, for example 10 or 100. Usually, you will find it helpful to choose either big values, like $C_1 = 100$, or small values, like $c = 1/100$, depending on the context. Ask yourself: is making C larger or smaller makes the statement easier to prove? Remember that these constants may not depend on anything, so $C_2 = \sqrt{n}$ is not a valid choice, for example.

The classical laws of probability theory, such as the law of large numbers, demonstrate the benefit of *averaging independent observations* X_i: the more observations we have, the more confident we become about the mean of the distribution. The next problem gives one more example of the benefit of averaging. Here you will find yourself in an unfamiliar territory where almost nothing is assumed about continuous random variables X_i. They may have infinite variance, and even their means do not need to exist!

Problem 1 (Small ball probabilities)

Let $X_1, \dots, X_N$ be *non-negative* independent random variables. Assume that the PDF (probability density function) of each X_i is uniformly bounded by 1.

(a) Check that each X_i satisfies

$$\mathbb{P}\left\{X_i \le \varepsilon\right\} \le \varepsilon \quad \text{for all } \varepsilon > 0.$$

(b) Show that the MGF (moment generating function) of each X_i satisfies

$$\mathbb{E}\exp(-tX_i) \le \frac{1}{t} \quad \text{for all } t > 0.$$

(c) Deduce that averaging increases the strength of (a) dramatically. Namely, show that

$$\mathbb{P}\left\{\frac{1}{N}\sum_{i=1}^{N} X_i \le \varepsilon\right\} \le (C\varepsilon)^N \quad \text{for all } \varepsilon > 0.$$

Problem 2 (Boosting randomized algorithms)

Imagine we have an algorithm for solving some decision problem (for example, the algorithm may answer the question: "is there a helicopter in a given image?"). Suppose each time the algorithm runs, it gives the correct answer independently with probability $\frac{1}{2} + \delta$ with some small margin $\delta \in (0, 1/2)$. In other words, the algorithm does just marginally better than a random guess.

To improve the confidence, the following "boosting" procedure is often used. Run the algorithm N times, and take the majority vote. Show that the new algorithm gives the correct answer with probability at least $1 - 2\exp(-c\delta^2 N)$. This is good because the confidence rapidly (exponentially!) approaches 1 as N grows.

Problem 3 (Optimality of Chernoff's inequality)

(a) Prove the following useful inequalities for binomial coefficients:

$$\left(\frac{n}{m}\right)^m \le \binom{n}{m} \le \sum_{k=0}^{m} \binom{n}{k} \le \left(\frac{en}{m}\right)^m$$

for all integers $m \in [1, n]$.

(b) Show that Chernoff's inequality is almost optimal. Namely, if S_N is a binomial random variable with mean μ, that is $S_N \sim \mathrm{Binom}(N, \mu/N)$, show that

$$\mathbb{P}\{S_N \ge t\} \ge e^{-\mu}(\mu/t)^t$$

for any integer $t \in \{1, \dots, N\}$ such that $t \ge \mu$.

Chernoff inequality we proved in class gives a bound on the one-sided tail $\mathbb{P}\{S_N \le t\}$. The next problem yields a similar bound on the other tail.

Problem 4 (The lower tail in Chernoff's inequality)

Let X_i be independent Bernoulli random variables with parameters p_i. Consider their sum $S_N = \sum_{i=1}^N X_i$ and denote its mean by $\mu = \mathbb{E}\, S_N$. Then, for any $t < \mu$, we have

$$\mathbb{P}\{S_N \le t\} \le e^{-\mu}\left(\frac{e\mu}{t}\right)^t.$$

TURN OVER FOR HINTS

Hints

Hint for Problem 1. Rewrite the inequality $\sum X_i \le \varepsilon N$ as $\sum(-X_i/\varepsilon) \ge -N$ and use the MGF method as in the proof of Hoeffding's inequality. Use part (a) to bound the MGF.

Hint for Problem 2. Apply a convenient form of Chernoff's inequality for S_N being the number of the wrong answers.

Hint for Problem 3. (a) To prove the upper bound, multiply both sides by the quantity $(m/n)^m$, replace this quantity by $(m/n)^k$ in the left side, and use the Binomial Theorem ("Newton's binomial").

(b) Since S_N is a binomial distribution, the smaller probability $\mathbb{P}\{S_N = t\}$ can be expressed using binomial coefficients. To lower bound the binomial coefficient, use part (a). To handle one of the remaining terms $(1 - \mu/N)^{N-t}$, check that the smaller quantity $(1 - \mu/N)^{N-\mu}$ is bounded below by $e^{-\mu}$.

Hint for Problem 4. Modify the proof of Chernoff's inequality.