EE278 Statistical Signal Processing **Stanford, Autumn 2023**

Homework 2

Due: Thursday, October 12, 2023, 1:00 pm on Gradescope

Please upload your answers timely to Gradescope. Start a new page for every problem. For the programming/simulation questions you can use any reasonable programming language. Comment your source code and include the code and a brief overall explanation with your answers.

1. Repeat part (a) and (b) of Problem 4 from HW 1 by using Hoeffding's inequality to get a reasonable bound on the sequence length n.

2. Sub-Gaussian random variables
   Let $X_1, X_2, \ldots, X_n$ be independent (but not necessarily identically distributed) and sub-Gaussian with variance proxies $\sigma_1^2, \ldots, \sigma_n^2$ respectively. Show that $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n a_i X_i$ is sub-Gaussian and provide an expression for its variance proxy.

3. McDiarmid's Inequality

   Let A be a $n \times n$ matrix with each entry chosen to be 0 or 1 with probability 1/2 independently of the other entries. Let $f : \mathbb{R}^{n\times n} \to \mathbb{R}$ be the function $f(A) = \mathrm{trace}(A^T A)$.

   a) What is $\mathbb{E}[f(A)]$?

   b) Using McDiarmid's inequality provide a bound on $\Pr\{f(A) \geq (1+\epsilon)\mathbb{E}[f(A)]\}$ for $\epsilon > 0$.

4. Estimating mean and variance of a Gaussian distriibution
   Let $X_1, X_2, \ldots, X_n$ be i.i.d. Gaussian observations with mean μ and variance σ^2.

   (a) Suppose the variance σ^2 is known and we want to estimate the mean of the distribution by taking the average of the samples, i.e. our estimate for the mean is given by

   $$\hat{X}_{mean} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

   Show that $\mathbb{E}[\hat{X}_{mean}] = \mu$. An estimator that satisfies this property is called an unbiased estimator of the unknown parameter.

   (b) Provide an (exponentially decaying) upper bound on the probability that the estimate $\hat{X}_{mean}$ deviates from the true mean by some given $t > 0$, i.e.

   $$\Pr\left\{|\hat{X}_{mean} - \mu| \geq t\right\}.$$

(c) Suppose now that the mean μ is known and we want to estimate the variance of the distribution by using $\hat{X}_{variance} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$ as our estimate. Show that $\mathbb{E}[\hat{X}_{variance}] = \sigma^2$, i.e. $\hat{X}_{variance}$ is an unbiased estimator of the variance.

(d) Provide an upper bound on the probability that the estimate $\hat{X}_{variance}$ deviates from the true variance by some given $t > 0$, i.e.

$$\Pr\left\{|\hat{X}_{variance} - \sigma^2| \geq t\right\}$$

for some given $t > 0$, by using Chebyshev's inequality. Hint: you can use the fact that $\mathbb{E}[W^4] = 3$ for $W \sim N(0,1)$.

(e) Let $Z = (X_1 - \mu)^2$. Compute the moment generating function of Z.

(f) Is Z sub-Gaussian? Hint: Recall that the moment generating function of an q -sub-Gaussian r.v. Y with mean μ has to satisfy $M_{Y-\mu}(s) \leq e^{s^2 q^2/2}$ for all $s \in \mathbb{R}$.

(g) Derive an exponentially decaying upper bound on

$$\Pr\left\{|\hat{X}_{variance} - \sigma^2| \geq t\right\},$$

by using the method in the derivation of the Hoeffding's inequality.

(h) Compare the decay rate as a function of t to the decay rate in part (b). Which one is easier to estimate: the mean or the variance?

5. Chernoff bound

The Chernoff bound is a concentration inequality specifically for Bernoulli random variables. In this exercise, you will first prove it and then investigate whether or not it is more powerful than the Hoeffding's inequality applied to Bernoulli random variables. The exact statment of the Chernoff bound is as follows: let $X_1, X_2, \ldots, X_n$ be independent Bernoulli random variables with mean p. Let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, then

$$\Pr\left\{\bar{X}_n - p \geq t\right\} \leq e^{tn}\left(\frac{p}{p+t}\right)^{(t+p)n}.$$

a) Prove the above bound by following the steps in the proof of Hoeffding's inequality and use the fact that for Bernoulli random variables $M_{X_i}(s) = 1 + p(e^s - 1) \leq e^{p(e^s-1)}$.

b) If $p = 1/n$ and $t = \frac{\log n}{n}$, how does this probability decay as a function of n? Linear/ quadratic/ polynomial/ faster than polynomial/exponential?

c) Now assume we apply Hoeffding's inequality to bound $\Pr\left\{\bar{X}_n - p \geq t\right\}$ instead of the Chernoff bound above. When $p = 1/n$ and $t = \frac{\log n}{n}$ and $n \to \infty$, how does the bound from Hoeffding's inequality compare to the one from the previous part?