Qualifying Exam Math 281AB

Spring 2024

Directions: please read carefully

The total points on this test exceed the maximum points attainable; this is to provide you flexibility in choosing the problems that best align with your strengths and preparation.

Total points possible: 70.

Maximum score: 50 points.

For those aiming for an **MS level pass**, your target is to accumulate up to **30** points. You may choose to address Problems 1 and 2 exclusively, or delve into a portion of Problem 3 to reach this benchmark.

If you're setting your sights on a **PhD- level** distinction, you should aim for a score **above 30** points and **up to 40** points.

However, if your aspirations are set on achieving a **PhD level pass** or higher, you are expected to gather **above 40**.

NOTE: Ensure that your computation is thoroughly detailed. Do not use examples or exercises from the book as statements. While you can use Theorems, Propositions, and Lemmas as statements, it is imperative to clearly define all relevant terms state the name, number and a book of the result you are using. Online resources should not be referenced or copied. Be meticulous in documenting every step of your work. Ensure your handwriting is clear and legible; illegible work will result in a deduction of points.

Name (Printed): ______________________________

Student ID Number: ______________________________

Signature: ______________________________

Problem 1 [5 points]

Is the following parametrization identifiable? We observe $X - Y$ where X and Y are independent $\mathcal{N}(\mu_1, \sigma^2)$ and $\mathcal{N}(\mu_2, \sigma^2)$ and parameter of interest here is $\theta = (\mu_1, \mu_2)$.

Problem 2 [5 points]

Is the following family of distributions an exponential family? (prove or disprove)

$$p(x, \theta) = 2\frac{x + \theta}{1 + 2\theta}, \qquad 0 < x < 1,\ \theta > 0.$$

Problem 3 [10 points]

Let $X_1, \cdots, X_n$ be a sample from a Poisson Distribution with the unknown parameter $\theta > 0$. Show that $\bar{X}$ is sufficient statistics for θ. Do so directly as well as using factorization theorem.

Problem 4 [20 points]

For a given n, let $X_1, \ldots, X_n$ be independent and identically distributed with distribution function

$$P(X_i \le t) = F(t).$$

Let $\hat{F}(x) = n^{-1} \sum_{i=1}^{n} I\{X_i \le x\}$ be the empirical distribution function. Here $\mathbb{I}$ denotes the indicator random variable that takes value 1 when the clause is correct and takes value 0 otherwise. Let

$$\mathcal{E}_n(x) = \sqrt{n}[\hat{F}(x) - F(x)], \qquad x \in \mathbb{R}.$$

(a) [10 points] Show that for fixed x_0, $\mathcal{E}_n(x_0)$ converges in distribution to a normal random variable with mean zero and variance $F(x_0)[1 - F(x_0)]$.

(b) [10 points] Find covariance for any $x_1 \le x_2$

$$\text{Cov}(\mathcal{E}_n(x_1), \mathcal{E}_n(x_2)).$$

Problem 6 [30 points]

Consider a random sample of observations $(X_i, Y_i)_{i=1}^n$ following a linear regression model

$$Y_i = \theta_0^\top X_i + e_i$$

for i.i.d. errors $e_1, \cdots, e_n$ that are independent of $X_1, X_2, \cdots, X_n$. Consider the following M-estimator of the parameter θ_0:

$$\hat{\theta} = \arg\max_{\theta} n^{-1} \sum_{i=1}^{n} w_i (Y_i - \theta^\top X_i)^2$$

where $w_i = g(X_i)$ are weights for each of the data points. Here, $g : \mathbb{R}^d \to \mathbb{R}_+$ is a given function of the covariates X_i.
NOTE: In this problem (all parts) you will need to identify a set of conditions for the weights w_i that are necessary to hold in addition to the ”classical” conditions used in the class lecture notes.

(a) [10 points] Determine the conditions on the weights w_i that ensure the consistency of the estimator, assuming that each error term e_i has an equal and known variance of 1. Additionally, provide a proof of consistency. Merely restating conditions from textbooks or lectures will not be sufficient to earn full or majority credit.

(b) [10 points] Determine the conditions on the weights w_i that ensure the asymptotic normality of the estimator, assuming that each error term e_i has an equal and known variance of 1. Additionally, provide a proof of asymptotic normality. Merely restating conditions from textbooks or lectures will not be sufficient to earn full or majority credit.

(c) [10 points] Prove asymptotic normality of the above estimator if we assume that e_i's are not identically distributed, i.e. that the variance of each e_i are NOT equal but are known to be σ_i: $\sigma_i \neq \sigma_j$. Here, you can rewrite $e_i = \sigma_i \epsilon_i$ with ϵ_i being i.i.d.