Mathematical Statistics Qualifier Examination
Part I of the STAT AREA EXAM
August 19, 2020; 9:00 AM - 11:00 AM

Academic integrity is expected of all students at all times, whether in the presence or absence of members of the faculty. Understanding this, I declare that I shall not give, use, or receive unauthorized aid in this examination.

NAME: _____ **ID:** _____

Signature: _____

Instruction: There are 4 problems. You are required to solve them all. Please show detailed work for full credit. You need to turn in your exam by 11:05 am, and receive the questions for your applied statistics exam at 11:15 after a break.

- This exam is conducted via Zoom on August 19 from 9:00 to 11:00 am EDT.
- The entire Zoom meeting and chat messages are being recorded.
- This is a closed book, closed note exam.
- Hand calculators (or other computing devices) may not be used during the exam.
- You should join the Zoom meeting from two devices: Your computer/laptop/tablet (with webcam), and your smartphone (with camera).
- Audio should be muted and video must be kept on during the exam.
- Your computer webcam must fully show your face; your smartphone camera should show your hands and workspace, with the pages of paper being used for the exam.
- At the very beginning of the exam, during set up, you will be asked to do a brief "environment scan", showing the workspace where your computer is and the desk/table/floor where you will be writing your work.
- You are required to bring enough blank pieces of paper to use for the exam. You will show the blank pages at the beginning, during the "environment scan" on Zoom.
- You are not allowed to use the internet for any searching or communication with others, with the sole exception of communication privately with the proctors via Zoom chat (which is set so that your chats only go privately to hosts, not to others).
- It is recommended that you print the exam and write your answers on it. However, you can write your answers on your blank papers if you do not have a printer with you.
- After you finish the exam, scan your pages, ordered and oriented appropriately, into a single pdf file. Email the pdf file to **hongshik.ahn@stonybrook.edu** no later than 5 minutes after completion of the exam (i.e., **by 11:05 am EDT**).
- No students are allowed to leave the Zoom meeting until the exam is over.
  - If you finish the exam early, then submit your exam and remain in the Zoom meeting until the conclusion of the exam at 11:00 am EDT.
  - After submitting your exam, you can study for another exam or work on anything else, while staying in view in the Zoom meeting.
- If the answers are not submitted by 11:05 am EDT, the exam will not be graded, and a score of zero will be given.
- If you have a question during the exam, then send a chat message to the host privately.

1. Let $\{X_1, X_2, \ldots\}$ denote a sequence of independent and identically distributed random variables with zero mean and variance $\sigma^2 < \infty$. Find

$$\lim_{n\to\infty} \mathrm{E}(|X_1 + \cdots + X_n|/\sqrt{n}).$$

(Hint: Suppose that $\{Y_1, Y_2, \ldots\}$ is a sequence of non-negative random variables such that

(a) For $p > 1$, $\sum_{n=1,2,\ldots} \mathrm{E}({Y_n}^p) < c$ for some constant c,

(b) Y_n converges in distribution to a random variable Y.

Then $\mathrm{E}(Y)$ is finite, and $\lim_{n\to\infty} \mathrm{E}(Y_n) = \mathrm{E}(Y)$.)

2. To test COVID-19 efficiently with limited resources, batch tests are recommended by pooling multiple swab samples from each individual. If the batch is tested negative, then it can be deduced that all samples were negative. Otherwise, each sample needs to be tested individually. We conduct a hypothesis testing with null hypothesis H_0: The subject is healthy versus alternative hypothesis H_1: The subject is infected. Denote α as the probability of a type I error and β as the probability of a type II error for an individual test. Let the infection rate be p, and X the random variable denoting the number of positive cases in a batch of size n. Answer the following questions in terms of α, β and p.

(a) Find the probability that a batch is tested negative.

(b) Fill out the four probabilities in batch testing.

		True condition	
		No samples are infected	At least one sample is infected
Test Result	−	(A)	(B)
	+	(C)	(D)

(c) Find the false negative rate (one minus sensitivity) of batch testing.

3. Let $Y_1, \ldots, Y_n$ be independent random variables such that for $i = 1, \ldots, n$, $\mathrm{E}(Y_i|X = x_i) = \beta x_i$ and $\mathrm{Var}(Y_i|X = x_i) = \sigma^2 w_i^{-1}$, $w_i > 0$. If the conditional distribution of Y given x is Gamma with pdf

$$f(y|x) = \frac{y^{p-1}}{\Gamma(p) a_x^p} e^{-y/a_x},\ y > 0,\ p > 0,$$

where a_x is a function of x, and $w_i^{-1} = x_i^2$, prove that the MLE of β is also the weighted least squares estimator obtained by minimizing $\sum_{i=1}^n w_i (y_i - \beta x_i)^2$.

4. Let $X_1, \ldots, X_n$ be independently and identically distributed according to a continuous cdf $F(x)$. It is sometimes desirable to determine a random lower limit $L(X_1, \ldots, X_n)$ that has, with probability at least $1 - \alpha$, at least proportion β of the population to the right. Then $L(X_1, \ldots, X_n)$ is called a *one-sided nonparametric tolerance limit*. That is,

$$1 - \alpha \le P[L(X_1, \ldots, X_n) \le \xi_{1-\beta}] = P[F(L(X_1, \ldots, X_n)) \le 1 - \beta].$$

(a) Find the equation for determining the smallest sample size such that the s-th order statistic $X_{(s)}$ can serve as the tolerance limit.

(b) Find the smallest sample size such that $X_{(1)}$ is the tolerance limit.