Comprehensive Exam: Statistical theory

Boston University, 2022

Instructions: This is a closed book exam. You are not allowed a crib sheet or a calculator. Please answer problems 1–2 (MA 581), 3–4 (MA 582), and 5–6 (MA 583) in separate blue books. ALL answers need to include an explanation, even if this is not explicitly asked in the question.

Instructions for Statistics Students: You need to answer four (and exactly four) out of six problems in the exam. You need to answer at least one problem from each of 581, 582, and 583. If you answer more than four problems, the lowest four scores will be used to compute your total. If you do not want us to grade any part of your answer, please cross it out completely.

Instructions for Biostatistics Students: You need to answer all of questions 1–4 (MA 581 and 582). You do not need to answer questions 5–6.

1. The number of students who register for a certain difficult course is distributed Poisson with mean 6. Due to its difficulty, each student has only an 75% chance of finishing that course. Students work independently of one another but have identical talent.

   (a) What is the probability that no one finishes this course?

   (b) Last year, that course had no students finishing. Find the probability that exactly 8 registered for that course last year. Interpret your answer appropriately.

   (c) This course has been run 3 times (under different instructors teaching independently of each other but with identical teaching skills). What is the probability that a total sum of 15 students altogether registered for the three offerings of that course?

2. Let X be a rv with pdf (density) given by $f(x) = c/x^3$, for $1 < x < 3$, where c is a constant.

   (a) Determine the value of c.

   (b) Find $\mathbb{P}(1.5 < X < 2)$.

   (c) Find the cdf F of X.

   (d) Compute $\mathbb{P}(X > 1.5|X < 2)$.

   (e) Compute $\mathbb{E}(X)$.

   (f) Compute stdev(X).

   (g) Compute $\mathbb{E}[X^3 cos(X)]$ exactly.

   (h) Compute the failure (or hazard) rate $R(t)$ of X, for $1 < t < 2$. (Recall $R(t) = f(t)/\mathbb{P}(X > t)$ for $t > 0$).

   (i) Suppose that 10 students each independently generate their own value of X. Find the probability that exactly 6 of those students generate a value greater than 1.5.

   (j) Suppose Y is a rv such that X and Y are iid (where X is the rv of this problem). Compute stdev$(4X - 3Y + 8)$.

3. Consider the Exp(θ) family, where $\theta > 0$ (and note that here $\mathbb{E}(X) = \theta$).

   (a) Show that no two different parameter values could give exactly the same pdf (this is "regularity condition R0").

   (b) Is the parameter space here open (this is "regularity condition R1")? Why or why not?

   (c) Is the support of the pdf here independent of the parameter ("parameter-free" - this is "regularity condition R2")? Why or why not?

   (d) Find the maximum likelihood estimator (MLE) of θ, call it Y_n here.

   (e) Show whether or not Y_n is unbiased for θ.

   (f) Show whether or not Y_n is a consistent estimator for θ.

   (g) Show whether or not Y_n is asymptotically normal, and if it is, identify its asymptotic normal variance (ANV).

   (h) Find $I(\theta)$, Fisher's Information, for this family – Is your MLE(θ) efficient? Why or why not?

   (i) Find the MLE of θ

   (j) Show whether or not it is biased.

   (k) Find a function g so that $n^{1/2}(g(Y_n) - g(\theta))$ is asymptotically standard normal ("variance stabilization").

4. We say the rv X has the D distribution with parameter $\theta > 0$ (written $X \sim D(\theta)$) if X has pdf $f(x, \theta) = 2x/\theta^2$, for $0 < x < \theta$, and $f(x, \theta) = 0$, elsewhere. Consider the parameterized D family $\{D(\theta) : \theta > 0\}$.

   (a) For each of the "regularity conditions" R0, R1, and R2, determine if the D family here satisfies that condition or not. (Refer to problem 3 here for a reminder of what those regularity conditions are, even if you do not work #3.)

   (b) Let Y_n be the maximum of the random sample of size n. Show that Yn is a consistent estimator of θ.

   (c) Find the pdf of Y_n. (Hint: Find its cdf first.)

   (d) Show that Y_n is NOT an unbiased estimator of θ. Show how to correct it to create an unbiased estimator of θ.

   (e) Show that $n(\theta - Y_n)$ converges in distribution, and find its asymptotic distribution explicitly.

5. Let B be a Bernoulli[p] random variable, $\Lambda = 1 + B$, and N_t be the counting process of a doubly stochastic Poisson process with rate Λ.

   (a) Compute $E[N_1]$.

   (b) Compute $E[\Lambda | N_1 = 1]$.

   (c) Compute $E[N_2 | N_1 = 1]$.

   (d) Compute $E[N_1 | N_2 = 1]$.

6. Let $X_1, X_2, \ldots$ be a 3-state, discrete time Markov chain with the following transition probability matrix:

$$P = \begin{pmatrix} 0 & 1-a & a \\ 1 & 0 & 0 \\ 0 & 0.1 & 0.9 \end{pmatrix}, \text{ where } a \in \{0, 0.1\}.$$

   (a) For the case $a = 0$: Label each state as transient, null recurrent, or positive recurrent and provide the period of each state. How many stationary distributions does this chain have? Compute a stationary distribution, if one exists. Will this chain approach a stationary distribution from any initial distribution? Starting from each state, what is the expected number of steps to return to that state?

   (b) Repeat part (a) for the case $a = 0.1$.