Comprehensive Exam: Statistical theory

Boston University, 2023

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 enumerate 1–4 (MA 581 and 582). You do not need to answer enumerate 5–6.

1. Let $X_1, X_2, \ldots, X_{10} \overset{iid}{\sim} \text{Binom}(20, p)$ for $p \in (0,1)$ and let $S_k = \sum_{i=1}^{k} X_i$ for each $k = 1, 2, \ldots, 10$.

   (a) For all values of $x \in \mathbb{R}$, specify the probability mass function for S_{10}.

   (b) For all values of $x \in \mathbb{R}$, compute the conditional probability mass function for X_{10} given that $S_{10} = 100$.

   (c) Compute the mean and variance for S_k in terms of p and k.

   (d) Suppose that the parameter p is distributed according to a Uniform$(0,1)$ and that conditioned on p, it still holds that $X_1, X_2, \ldots, X_{10} \overset{iid}{\sim} \text{Binom}(20, p)$. Compute the (unconditional) mean and variance for S_k in terms of k.

   (e) Suppose that $p = 1/200$, specify the name and parameter value for an approximate distribution for S_{10}.

2. Let X be a random variable with p.d.f. given by $f(x) = c(a^2 - x^2)$ for $x \in [-a, a]$ and 0 otherwise, for some $a > 0$.

   (a) Compute the value of c.

   (b) Find the c.d.f. of X.

   (c) Find $\Pr[X > a/2 | X > 0]$.

   (d) Compute $\mathbb{E}[X]$.

   (e) Compute $\text{Var}[X]$.

   (f) Compute $\mathbb{E}[|X|]$.

   (g) Compute $\mathbb{E}[(X^2 - a^2)^{-1}]$.

   (h) Suppose that $X_1, \ldots, X_{1000}$ are i.i.d. each with density f. Give the name and parameter values for an approximate distribution for $\sum_{i=1}^{1000} X_i$.

3. Suppose X has an Exponential(θ) distribution, where $\theta > 0$ is the constant failure rate, and note that $\mathbb{E}(X) = 1/\theta$. Show all your work in solving the following:

   (a) Does this family satisfy all the regularity conditions for good MLE performance? Explain.

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

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

   (d) Show Y_n is a consistent estimator for θ.

   (e) Show Y_n is asymptotically normal, and if it is, identify its asymptotic normal variance explicitly.

   (f) Find the MLE of $\log \theta$. Show that it is asymptotically normal, and identify its asymptotic normal variance explicitly.

   (g) Show how to find a function g so that $n^{1/2}(g(Y_n) - g(\theta))$ is asymptotically standard normal for all values of $\theta > 0$.. (Hint: Use (f) here.)

4. We say the rv X has the W distribution with parameter $\theta > 0$ (written $X \sim W(\theta)$ if X has pdf

$$f(x, \theta) = \begin{cases} 5x^4/\theta^5, & \text{for } 0 < x < \theta \\ 0, & \text{elsewhere.} \end{cases}$$

   Consider the parameterized W family $\{W(\theta) : \theta > 0\}$. Show all your work in solving the following:

   (a) Show that the MLE of θ is the sample maximum.

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

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

   (d) Show that Y_n is NOT an unbiased estimator of θ

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

   (f) Find an unbiased estimator of θ, call it T_n . Show that T_n is a consistent estimator of θ

   (g) Show that $n(\theta - T_n)$ converges in distribution, and find its asymptotic distribution explicitly. (Hint: Use parts (e) and (f) here.)

5. Vehicles pass a point on the highway at a Poisson rate of one vehicle per minute. If 5% of the vehicles on the road are vans, then

   (a) What is the probability that at least one van passes by during an hour?

   (b) Given that 10 vans have passed by in an hour, what is the expected number of vehicles to have passed by in that time?

   (c) If 50 vehicles have passed by in an hour, what is the probability that five of them were vans?

6. A virus can exist in N different strains, numbered from 1 to N. At each generation, the virus mutates with probability $\alpha \in (0,1)$ to another strain which is chosen at random with equal probability. Let X_n denote the strain the virus is in after n generations, $n \geq 0$.

   (a) Why is $\{X_n : n \geq 0\}$ a Markov chain? What is the transition matrix of this Markov chain?

   (b) Justify why this Markov chain has a unique stationary distribution and compute it.

   (c) Compute the average number of generations to find the virus again in the same strain.