Statistics 134 - Instructor: Adam Lucas

Final Exam

Wednesday, May 9, 2018

Print your name: ______________________________

SID Number: ______________________________

Exam Information and Instructions:

- You will have 170 minutes to take this exam. Closed book/notes/etc. No calculator or computer.
- We will be using Gradescope to grade this exam. Write any work you want graded on the front of each page, in the space below each question. Additionally, write your SID number in the top right corner on every page.
- Please use a dark pencil (mechanical or #2), and bring an eraser. *If you use a pen and make mistakes, you might run out of space to write in your answer.*
- Provide calculations or brief reasoning in every answer.
- Unless stated otherwise, you may leave answers as unsimplified numerical and algebraic expressions, and in terms of the Normal c.d.f. Φ. Finite sums are fine, but simplify any infinite sums.
- Do your own unaided work. Answer the questions on your own. The students around you have different exams.

I certify that all materials in the enclosed exam are my own original work.

Sign your name: ______________________________

GOOD LUCK!

SID: ____________________

1. Yiming and Brian are playing darts. Suppose each dart's distance from the bullseye are independent standard Rayleigh random variables. If a dart is within 1 unit of the bullseye, 2 points are earned; if in 1 to 4 units of the bullseye, 1 point is earned; otherwise no points.

   (a) Find the expected point value of each dart throw.

   (b) Now let p_i represent the probability of scoring i points. Yiming and Brian take turns throwing; Yiming throws first. In terms of p_0, p_1, p_2, what is the probability Yiming is first to score at least one point?

   (c) Darts are thrown until one scores two points. Let X denote the number of points scored until stopping (not including the final throw). Find $E(X)$.

   Solutions:

   (a) Let D represent this point value. Then,

   $$E(D) = 2(1 - e^{-\frac{1}{2}}) + 1(e^{-\frac{1}{2}} - e^{-\frac{1}{2}(4)^2}) + 0$$

   (b) Let p_0 represent the probability of scoring no point.

   $$\begin{aligned} P(\text{Yiming scores first}) &= (1 - p_0) + (1 - p_0)p_0^2 + \dots \\ &= \sum_{i=0}^{\infty} p_0^{2i}(1 - p_0) \\ &= \frac{(1 - p_0)}{1 - p_0^2} = \frac{1}{1 + p_0} = \frac{1}{1 + e^{-\frac{1}{2}4^2}} \end{aligned}$$

   (c) Let Y represent the number of tosses until stopping, so $Y \sim$ Geom (p_2) on $0, 1, 2, \dots$.

   $$E(X) = E(E(X|Y)) = E\left(\frac{p_1}{p_0 + p_1} Y\right) = \frac{p_1(1 - p_2)}{p_2(p_0 + p_1)} = \frac{p_1}{p_2}$$

2. Beloved Stat 134 students Jonny and Ethan have very different study habits for the final. Jonny quits studying at a constant hazard rate $\lambda = \frac{1}{2}$ (i.e. average 2 hours). Ethan randomly quits anytime between 0 and 4 hours (also average 2 hours). You may assume Jonny and Ethan study independently. What is the chance the first quitter stops studying between 2 and 3 hours.

   Solution:

   Let $T_J \sim exp(\frac{1}{2})$ equal the time until Jonny quits and $T_E \sim U(0, 4)$ equal the time until Ethan quits. Let $T = min(T_J, T_E)$. The cdf of T is

   $$F(t) = 1 - P(T_E > t)P(T_J > t) = 1 - \left(\frac{4 - t}{4}\right)(e^{-t/2}) = 1 - 4e^{-t/2} + te^{-t/2}$$

   for $0 < t < 4$. The answer is $F(3) - F(2) = 2/e - e^{-3/2}$.

3. Phone calls arrive into a telephone exchange according to a Poisson arrival process at rate λ per unit time. This exchange serves three regions, and an incoming call gets routed to region i with probability p_i, for $i = 1, 2, 3$. Note that $p_1 + p_2 + p_3 = 1$.

   (a) Let N_t^i denote the numbers of calls routed to region i in time t starting from time 0. Find $P(N_t^1 = j, N_t^2 = k)$.

   (b) On a particular day, the telephone exchange malfunctions, and it fails to route each incoming call with probabilty q, independently of all other calls; such a "dropped" call is lost forever and it does not get transmitted to any region. On this unlucky day, what is the expected waiting time between the successive calls received by region 1?

   Solutions:

   (a) we have independence of N_t^1 and N_t^2. N_t^1 is $Pois(\lambda p_1 t)$, and N_t^2 is $Pois(\lambda p_2 t)$ by Poisson thinning. Answer is $e^{-(p_1+p_2)\lambda t}(p_1\lambda t)^j(p_2\lambda t)^k)/j!k!$

   (b) The successive waiting time between calls of relay station 1 is distributed as $exp((1-q)p_1\lambda)$. So the expected time between calls is $1/((1-q)p_1\lambda)$

4. For each distribution of X indicated in the parts below, evaluate $Var(X)$ as a simple fraction,without involvingthe constants c_1, c_2, c_3.

   (a) X with values in $[0, 1]$ and density $f_X(x) = c_1 x^2(1-x)^3$ for $0 < x < 1$.

   (b) X with values in $[0, \infty)$ and density $c_2 x^4 e^{-3x}$ for $0 \le x < \infty$

   (c) X with possible values $0, 1, 2, \ldots$ and $P(X = k) = \frac{c_3}{k!}$ for $k = 0, 1, 2, \ldots$.

   Solutions:

   (a) $X \sim Beta(r = 3, s = 4)$ so $Var(X) = \frac{rs}{(r+s)^2(r+s+1)} = 3/98$

   (b) $X \sim Gamma(r = 5, \lambda = 3$ so $Var(X) = \frac{r}{\lambda^2} = 5/9$

   (c) $X \sim Pois(\mu = 1)$ so $Var(X) = \mu = 1$.

5. The joint density of X and Y is

$$f(x,y) = \begin{cases} \frac{4y}{x} & \text{for } 0 < x < 1 \text{ and } 0 < y < x \\ 0 & \text{otherwise} \end{cases}$$

   Find the following:

   (a) $E(XY)$

   (b) The marginal density of X

   (c) $E(Y|X)$

   Solutions:

(a) $E(XY) = 1/3$

(b) $f_X(x) = 2x$

(c) $E(Y|X) = 2X/3$

6. Let U follow a Uniform $(0, 1)$ distribution. Find the distribution of $-log(U)$.

Solution:

Let $S = -log(U) \Rightarrow U = e^{-S} \Rightarrow \frac{ds}{du} = -\frac{1}{U} \Rightarrow |\frac{ds}{du}| = \frac{1}{u}$

$$\begin{aligned} f_S(s) &= \frac{f_U(u)}{|\frac{ds}{du}|} \\ &= \frac{1}{\frac{1}{u}} = \frac{1}{\frac{1}{e^{-s}}} \\ &= e^{-s} \Rightarrow Exp(1) \end{aligned}$$

7. Suppose X and Y are independent and normally distributed with mean 0 and variance 1. Find the density of X/Y.

Solution:

As shown on page 383 of Pitman, $f_{X/Y}(z) = \frac{1}{\pi(z^2+1)}$.

8. There are 10 students preparing for a TV show audition. Eight of them can sing while five of them can dance (i.e. 3 students can sing and dance). All selections are made without replacement except part a.

   (a) If a director randomly selects with replacement 7 students what is the chance that she selects exactly 4 singers? Hint: a student is either a singer or not a singer.

   (b) Two students are randomly selected. What is the chance that the first student can only sing given that the second student can dance?

   (c) A student has one of three possible skills:

      i. sing only
      ii. dance only
      iii. sing and dance

   If 5 students are randomly selected how many different skills do you expect the 5 students will have?

   For example: if all 5 students can both sing and dance then the number of different skills is 1. If 2 students can sing and dance, two students can dance only and one student can sing only then the 5 students would have 3 different skills.

Solutions:

(a) By the binomial formula $\binom{7}{4}(8/10)^4(2/10)^3$.

(b) This is an application of Bayes' rule. This is the chance that the 1st student can sing only and the second student can dance divided by the chance that the second student can dance. This is (5/10)(5/9) divided by 5/10 which is 5/9.

(c) Let X=the number of different skills 5 students have. We can write

$$X = I_1 + I_2 + I_3$$

where,
I_1 is 1 if at least 1 student can sing only, and zero otherwise.
I_2 is 1 if at least 1 student can dance only, and zero otherwise.
I_3 is 1 if at least 1 student can sing and dance, and zero otherwise.

The probability that at least 1 student can sing only is 1 minus the probability that no student can sing only. Using the slot method this is $1 - \frac{5\times4\times3\times2\times1}{10\times9\times8\times7\times6} = 0.996$

Similarly, the probability that at least 1 student can dance only is 1 minus the probability that no student can dance only. Using the slot method this is $1 - \frac{8\times7\times6\times5\times4}{10\times9\times8\times7\times6} = 0.888$

Finally the probability that at least 1 student can sing and dance is 1 minus the probability that no student can sing and dance. Using the slot method this is $1 - \frac{7\times6\times5\times4\times3}{10\times9\times8\times7\times6} = 0.917$
Then

$$E(X) = E(I_1) + E(I_2) + E(I_3) = 0.996 + 0.888 + 0.917 = 2.81$$

9. People arrive at a store according to a Poisson process of rate $\lambda = 2$ per minute. Let T_i denote the arrival time in minutes of the i^{th} customer.

(a) What is the expected value of the third arrival T_3?

(b) What is the expected value of the third arrival conditional on no customers arriving by time 2?

(c) What is the probability that the third arrival T_3 is more than 1 minute after T_1?

Solutions:

(a) $T_3 \sim Gamma(r = 3, \lambda = 2)$ so $E(T_r) = r/\lambda = 3/2$.

(b) $E(T_r|T_1 > 2) = 2 + r/\lambda$. Hence $E(T_3|T_1 > 2) = 2 + 3/2 = 7/2$.

(c) $P(T_3 - T_1 > t) = P(T_2 > t) = \sum_{k=0}^{1} e^{-\lambda t}\frac{(\lambda t)^k}{k!}$. Hence $P(T_3 - T_1 > t) = e^{-2} + e^{-2}2 = 3e^{-2}$

10. Suppose we have a collection of n i.i.d variables, $X_1, X_2, \ldots, X_n \sim \mathcal{N}(0, \sigma^2)$. We are interested in the relationship between a single data point X_1 and the sample mean, $\overline{X} = \frac{1}{n}(X_1 + X_2 + \ldots + X_n)$.

    (a) Find $Corr(X_1, \overline{X})$. Use this to identify the joint distribution of $(X_1, \overline{X})$. Be sure to state all parameters. Write your answer as a simplified fraction.

    (b) Find $\mathbb{P}(\overline{X} > 0 \mid X_1 > 0)$. (Use the conditional probability rule.)

    (c) How large must X_1 be such that there is a 95% chance that the sample mean is greater than zero?

Solutions:

(a) We break up the covariance using bilinearity of expectations, noting most terms are zero by independence.

$$\begin{aligned} Cov(X_1, \overline{X}) &= \sum_{i=1}^{n} Cov(X_1, \frac{1}{n}X_i) \\ &= Cov(X_1, \frac{1}{n}X_1) \\ &= \frac{1}{n}Var(X_1) \\ &= \frac{\sigma^2}{n} \end{aligned}$$

Thus, $Corr(X_1, \overline{X}) = \frac{\sigma^2/n}{\sigma \cdot (\sigma/\sqrt{n})} = \frac{1}{\sqrt{n}}$. From this, we conclude $(X_1, \overline{X})$ is bivariate normal, with means $0, 0$, variances $\sigma^2, \frac{\sigma^2}{n}$, and correlation $\frac{1}{\sqrt{n}}$.

(b) Here, we must represent $X, \overline{X}$ in terms of standard bivariate normal variables. The σ^2 does not affect this answer due to symmetry, but we have to account for the $\sqrt{n}$ term. For simplicity, here let $\sigma^2 = 1$. Then,

$$\sqrt{n} \cdot \overline{X} = \frac{1}{\sqrt{n}}X_1 + \sqrt{\frac{n-1}{n}}Z$$

gives us the representation in terms of standard normals. From here, we find the probability both are positive, and apply the conditional probability.

$$\begin{aligned} P(X > 0, \overline{X} > 0) &= P(X > 0, \frac{1}{\sqrt{n}}X_1 + \sqrt{\frac{n-1}{n}}Z > 0) \\ &= P(X_1 > 0, Z > \frac{-1}{\sqrt{n-1}}X_1) \\ &= \frac{1}{4} + \frac{1}{2\pi}\arctan\left(\frac{1}{\sqrt{n-1}}\right) \\ \therefore \quad P(\overline{X} > 0 \mid X_1 > 0) &= 2\left(\frac{1}{4} + \frac{1}{2\pi}\arctan\left(\frac{1}{\sqrt{n-1}}\right)\right). \end{aligned}$$

(c) Here we must find the conditional distribution of $\overline{X}|X_1 = x$. This conditional distribution is normal with mean $\frac{x}{n}$ and variance $\frac{n-1}{n^2}\sigma^2$. Thus, we find x such that

$$\Phi\left(\frac{0 - x/n}{\sigma\sqrt{\frac{n-1}{n^2}}}\right) = 0.05$$

A bit of algebra yields the result $x = -\sigma\sqrt{n-1} \cdot \Phi^{-1}(0.05)$.