STA 504 Mathematical Statistics with Calculus Review

Midterm Exam #1

10/03/2022

Time: 120 Minutes

Please Print: ________________ ___________________
(First Name) (Last Name)

Instructions

- This is a closed-book test. Textbook, notes, and discussions are NOT allowed.
- You may use a calculator for the exam.
- Please show your detailed work to earn full credit.
- Partial credit will be granted to the key steps that reflect your correct reasoning even if your numerical answer is incorrect.

Problem 1. *Performance of a medical diagnostic test*. Consider the statistics in the report on a large-scale clinical trial of a newly developed medical diagnostic test for a genetic defect. Using the following notations in formulating the subsequent questions related to conditional probabilities.

D+ = having genetic defect
D- = having no genetic defect
T+ = positive test result
T- = negative test result

Assume that 1% of people have a certain genetic defect (D-). 90% of people with a known genetic defect reported a positive test result (T+|D+). 9.6% of people who do not have a genetic defect reported a positive test result (T+|D-).

1. Randomly select a person to receive the diagnostic test, what is the probability that the test result is positive?

Solution: What we want to find is P[T+]. Note that

$$P[T+] = P[T + \cap D+] + P[T + \cap D-]$$

$$= P[T + |D +]P[D +] + P[T + |D -]P[D -]$$

$$= 0.9 \times 0.01 + 0.096 \times (1 - 0.01) = 0.009 + 0.09504 = 0.10404 = 10.404\%$$

2. If a person gets a positive test result, what is the probability they actually have the genetic defect?

$$\mathrm{P[D + |T +]} = \frac{P[T + \cap D +]}{P[T +]} = \frac{0.9 \times 0.01}{0.10404} = \frac{0.009}{0.10404} = 0.0865 = 8.65\%$$

Problem 2. *Service Time Problem*. The number of customers arriving at a checkout counter in a department store follows a Poisson distribution at an average of seven per hour. That is,

$$P[Y = y] = \frac{7^y \mathrm{e}^{-7}}{y!}$$

Assume that it takes approximately ten minutes to serve each customer. Let Y be the number of customers arriving at a checkout counter within an hour. Then S = 10Y is the total service time during the 1-hour period. **Note that** $e^{-7} \approx 0.0009$.

1. Find the mean and variance of the total service time for customers arriving during a 1-hour period.

   Solution: Since S = 10Y, therefore,

$$E[S] = E[10Y] = 10\, E[Y] = 10 \times 7 = 70$$

$$V[S] = V[10Y] = 10^2 V[Y] = 100 \times 7 = 700$$

2. What is the probability that the total service time will be less than 30 minutes?

Solution:

$$P[S < 30] = P[10Y < 30] = P[Y < 3] = P[Y = 0] + P[Y = 1] + P[Y = 2]$$

$$= \frac{7^0 e^{-7}}{0!} + \frac{7^1 e^{-7}}{1!} + \frac{7^2 e^{-7}}{2!} = (8 + 24.5)e^{-7} \approx 0.03$$

The chance the total service time is less than 30 minutes is about a 3% chance.

Problem 3. *Quality and Maintenance Problem*. Suppose that a radio contains six transistors, two of which are defective.

1. Three transistors are selected at random, removed from the radio, and inspected. What is the probability that the three removed transistors include both defective ones?

Solution: This is a finite population with two types of elements. The number of defects among selected transistors follows a hypergeometric distribution.

$$P(Y = 2) = \frac{\binom{2}{2}\binom{4}{1}}{\binom{6}{3}} = \frac{1 \times 4}{\frac{5 \times 6 \times 4}{3 \times 2}} = 20\%$$

The probability of randomly selecting 3 transistors that include 2 defects is about 20%.

2. Instead of selecting three to remove from the radio, we want to know the minimum number of transistors that should be removed from the ratio so that the two defective ones will be removed with at least a 30% chance.

Solution: Let n be the total number of transistors to be removed so that two of them are defective with at least a 50% chance. Theoretical, we need to solve n from

$$P(Y = 2) = \frac{\binom{2}{2}\binom{4}{n-2}}{\binom{6}{n}} > 0.3$$

This is equivalent to

$$n(n-1) > 9$$

Solve for n from the above inequality, we have

$$n > \frac{1 + \sqrt{37}}{2} > 3.5$$

Therefore, the minimum number of transistors to be removed from the radio so that two defective transistors will be removed with at least 30%.

Note: You can also use trial-and-error approach to find the minimum number.

Problem 4.

Suppose that the waiting time for the first customer to enter a retail shop after 9:00 A.M. is a random variable *Y* with an exponential density function given by

$$f(y) = \begin{cases} \frac{1}{\theta} e^{-y/\theta}, & y > 0, \\ 0, & elsewhere. \end{cases},$$

Where θ is a positive constant.

1. Derive the CDF of Y.

   Solution:

   $If\ y < 0, F(y) = 0$. Next, we find $F(y)$ for $y > 0$.

   $$F(y) = \int_{-\infty}^{y} f(x)dx = \int_0^y \frac{1}{\theta} e^{-x/\theta} dx = -\int_0^y e^{-x/\theta} d(-x/\theta) = -e^{-x/\theta}\Big|_0^y = 1 - e^{-y/\theta}$$

   Therefore,

   $$F(y) = \begin{cases} 1 - e^{-y/\theta}, & y \geq 0, \\ 0, & y < 0. \end{cases}$$

2. Derive the expectation of Y. (Hint: it is a function of θ)

   Solution: we need to use both substitution and integral by parts

   $$E[Y] = \int_{-\infty}^{\infty} yf(y)dy = \int_0^\infty y\frac{1}{\theta} e^{-y/\theta} dy = -\int_0^\infty y\, e^{-y/\theta} d(-y/\theta)$$
   $$= -\int_0^\infty y\, d\left(e^{-y/\theta}\right) = -ye^{-y/\theta}\Big|_0^\infty + \int_0^\infty e^{-y/\theta}\, dy = \theta \int_0^\infty \frac{1}{\theta} e^{-y/\theta}\, dy = \theta$$

   Note that we used the 2nd requirement of the density function

   $$\int_0^\infty \frac{1}{\theta} e^{-y/\theta}\, dy = 1.$$

3. Find P(Y > u) using the CDF of Y.

   Solution: $P(Y > u) = 1 - P(Y < u) = 1 - F(u) = 1 - (1 - e^{-\frac{y}{\theta}}) = e^{-\frac{u}{\theta}}$

4. Find the conditional probability $P(Y > a + v \,|\, Y > v)$ with $a > 0$ and $v > 0$.

   Solution: We use the definition of conditional probability to find the answer.

$$P(Y > a + u \,|\, Y > u) = \frac{P[Y > (a+u) \cap Y > u]}{P[Y > u]} = \frac{P[Y > (a+u)]}{P[Y > u]} = \frac{e^{-\frac{a+u}{\theta}}}{e^{-\frac{u}{\theta}}} = e^{-\frac{a}{\theta}} = P[Y > a]$$

Note: The above condition basically says that the random variable is memoryless. For example, I have two cars (2 and 20 years old respectively) that follow the above distribution. Assume that both cars are in good working condition. I want to know the probability of 5-year survival of these two-cars.

Probability of surviving 5 more years of the 2-year-old car: $P(Y > 5 + 2 \mid Y = 2) = P(Y > 5)$

Probability of surviving 5 more years of the 20-year-old car: $P(Y > 5 + 20 \mid Y = 20) = P(Y > 5)$

They are equal! This means that chance of surviving 5 years for both cars are equal regardless of their age!