# STA 504 Mathematical Statistics with Calculus Review

**Midterm Exam #1**

9/30/2024

**Time: 120 Minutes**

**Please Print**: ________________ __________________
(First Name) (Last Name)

**Instructions**

- This is a closed-book test. Textbooks, 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.** Assume you have an imbalanced coin with a probability of observing H to be p. That is, Pr(H) = p. Let Y = outcome when flipping the coin just once. Assuming the numerical encoding of Y to be Y = 0, if T is observed, otherwise Y = 1.

1. Characterize the distribution of random variable Y by providing a table, a chart, or a function.

| **Y** | **P**(Y=y) |
|---|---|
| **0** | 1-**p** |
| **1** | p |

2. Calculate the expectation and variance of Y.

   $E[Y] = 0\times(1-p) + 1\times p = p = \mu$
   $V[Y] = E[Y^2] - \mu^2 = 0^2\times(1-p) + 1^2 p - p^2 = p - p^2 = p(1-p)$

3. If flipping that same imbalanced coin 10 times, what is the probability that you will observe 3 Heads?

   This is a binomial distribution problem with a preset number of trials = 10. Let X = number of heads to be observed, then

   $$P(X=3) = [10!/(3!7!)]p^3(1-p)^7 = 120\, p^3(1-p)^7$$

**Problem 2**. A company that is in the insurance business determines its premium amount based on the number of claims and the amount claimed per year. So, to evaluate its premium amount, the insurance company will determine the average number of claimed amounts per year. Then based on that average, it will also determine the minimum and the maximum number of claims that can reasonably be filed in the year. Based on the maximum number of the claim amount and the cost and profit from the premium, the insurance firm will determine what kind of premium amount will be good to break even. Therefore, the key information needed for evaluating its premium amount is the average number of claims **per year**.

1. Let's say the average number of claims handled by an insurance company per day is 5. Let X = number of claims handled on a randomly selected day, what is the probability that the claims to be handed in that day are between 1 and 10 inclusive ( $1 \le X \le 10$)?

   X =Poisson random variable with a mean of 5 per day.

   $$P(1 \le X \le 10) = \sum_{x=1}^{10} \frac{5^x e^{-5}}{x!} = \left(\frac{5^1}{1!} + \frac{5^2}{2!} + \cdots + \frac{5^9}{9!} + \frac{5^{10}}{10!}\right) e^{-5}$$
   $$= (5.000 + 12.500 + 20.833 + 26.042 + 26.042 + 21.701 + 15.501 + 9.688 + 5.382 + 2.691) e^{-5}$$
   $$\approx 145.38 \times e^{-5} \approx 0.97956$$

2. Let Y = total number of claims handled in a year. Assuming the insurance company handles claims 365 days a year. What is the probability that the total number of claims to be handled in a randomly selected year is at most 3285?

Clearly, Y = 365X. With this and $P(1 \le X \le 10) = 0.97956$ in mind, we have

$$\begin{aligned} P(Y \le 3285) = P(365X \le 3285) &= P(X \le 9) \\ &= P(X=0) + P(1 \le X \le 9) + P(X=10) - P(X=10) \\ &= P(X=0) + P(1 \le X \le 10) - P(X=10) \\ &= \frac{5^0 e^{-5}}{0!} + 0.9795668 - \frac{5^{10} e^{-5}}{10!} \approx 0.968172. \end{aligned}$$

**Problem 3.** Let Y be a random variable representing a lifetime with the following density function

$$f(y) = \begin{cases} ky(y+1), & x \in [0,1] \\ 0, & otherwise \end{cases}$$

1. Find the value of $k$ to make $f(y)$ a valid density function.

Set $\int_0^1 ky(y+1)dy = 1$, we have $k\int_0^1 (y^2+y)dy = 1$ which yields $k\left(\frac{y^3}{3}+\frac{y^2}{2}\right)\Big|_0^1 = 1$. That is, $k\left(\frac{1}{3}+\frac{1}{2}\right) = 1$. Therefore, $k = \frac{6}{5}$.

2. Find the CDF of the random variable Y.

$$F(y) = \int_0^y \frac{6}{5}(x^2+x)dx = \frac{6}{5}\left(\frac{x^3}{3}+\frac{x^2}{2}\right)\Big|_0^y = \begin{cases} 1, & x > 1; \\ \frac{2x^3+3x^2}{5}, & 0 \le x \le 1; \\ 0, & y < 0. \end{cases}$$

3. Find the expectation of Y.

$$E[Y] = \int_0^1 y\frac{6}{5}y(y+1)dy = \frac{6}{5}\int_0^1 (y^3+y^2)dy$$
$$= \frac{6}{5}\left(\frac{y^4}{4}+\frac{y^3}{3}\right)\Big|_0^1 = \frac{6}{5}\left(\frac{1}{4}+\frac{1}{3}\right) = \frac{6}{5}\times\frac{7}{12} = \frac{7}{10} = \mu$$

4. Find the variance of Y [Hint: find you can use any formula at your preference].

We use the alternative definition of variance: $V[Y] = E[Y^2] - \mu^2$

$$E[Y^2] = E[Y] = \int_0^1 y^2\frac{6}{5}y(y+1)dy = \frac{6}{5}\int_0^1 (y^4+y^3)dy$$
$$= \frac{6}{5}\left(\frac{y^5}{5}+\frac{4}{4}\right)\Big|_0^1 = \frac{6}{5}\left(\frac{1}{5}+\frac{1}{4}\right) = \frac{6}{5}\times\frac{9}{20} = \frac{54}{100}$$

$$V[Y] = E[Y^2] - \mu^2 = \frac{54}{100} - \left(\frac{7}{10}\right)^2 = \frac{5}{100} = \frac{1}{20}$$

5. Find conditional probability P[Y > 0.5 | Y< 0.75]

$$P[Y > 0.5 \mid Y < 0.75] = \frac{P[0.5 \leq Y < 0.75]}{P[Y < 0.75]} = \frac{P[Y < 0.75] - P[Y < 0.5]}{P[Y < 0.75]}$$

Note that

$$P(X < x) = F(x) = \begin{cases} 1, & x > 1; \\ \frac{2x^3 + 3x^2}{5}, & 0 \leq x \leq 1; \\ 0, & y < 0. \end{cases}$$

$$F(0.5) = \frac{2\times 0.5^3 + 3\times 0.5^2}{5} = \frac{1}{5} = 0.2, F(0.75) = \frac{2\times 0.75^3 + 3\times 0.75^2}{5} = \frac{2.53125}{5} = 0.506$$

$$P[Y > 0.5 \mid Y < 0.75] = \frac{P[Y < 0.75] - P[Y < 0.5]}{P[Y < 0.75]} = \frac{0.506 - 0.2}{0.506} \approx 0.605.$$

**Problem 4**. *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 a genetic defect
D- = having no genetic defect
T+ = positive test result
T- = negative test result

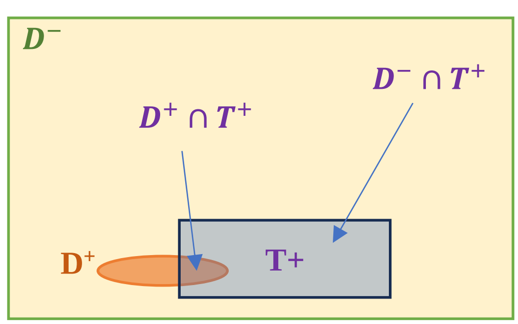
Assume that 1% of people have a certain genetic defect (D+). 90% of people with a known genetic defect reported a positive test result (T+|D+). 9.6% of people who do not have a genetic defect reported a positive test result (T+|D-).

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

We can see this from the Ven Diagram.

$$\boldsymbol{T^+ = (D^+ \cap T^+) \cup (D^- \cap T^+)}$$

$$\boldsymbol{P[T^+] = P(D^+ \cap T^+) + P(D^- \cap T^+)}$$



By the multiplicative rule, we have

$$\boldsymbol{P(D^+ \cap T^+) = P(T^+|D^+)P(D^+) = 0.90 \times 0.01 = 0.009}$$
$$\boldsymbol{P(D^- \cap T^+) = P(T^+|D^-)P(D^-) = 0.096 \times 0.99 = 0.09054}$$
$$\boldsymbol{P[T^+] = P[(D^+ \cap T^+) \cup (D^- \cap T^+)] = P(D^+ \cap T^+) + P(D^- \cap T^+)}$$
$$\boldsymbol{= 0.009 + 0.09054 = 0.10404}$$

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

$$\boldsymbol{P(D^+|T^+) = \frac{P(D^+ \cap T^+)}{P[T^+]} = \frac{0.009}{0.10404} = 0.0865}$$

**Bonus**. Understanding the distribution of customer waiting times is important to the manager of the service department to take suitable action and improve customer satisfaction. Let X be a random variable representing customers' waiting times (in minutes) for a specific division of an organization. Assume that the probability density function of X has the following form.

$$f(x) = \begin{cases} 0.1\, e^{-0.1x}, & \text{if } x > 0; \\ 0 & \text{if } x < 0. \end{cases}$$

1. Derive the cumulative probability distribution function (CDF).

   By definition, the CDF is given by

$$F(x) = \int_0^x 0.1\, e^{-0.1y}\, dy = -\int_0^x e^{-0.1y}\, d(-0.1y) = -e^{-0.1y}\Big|_0^x = 1 - e^{-0.1y}$$

2. Using the above CDF to find the probability that a randomly selected customer's waiting time for service is between 5 and 10 minutes.

$$P[5 < X < 10] = P[X < 10] - P[X < 5] = F(10) - F(5) = e^{-0.1\times 5} - e^{-0.1\times 10} = 0.2387$$

3. Given that a customer has already waited for 5 minutes, what is the probability that the customer still needs to wait for another 3 minutes?

$$P[X > 8|X > 5] = \frac{P[X > 8 \cap X > 5]}{P[X > 5]} = \frac{P[X > 8\,]}{P[X > 5]} = \frac{1 - P[X \le 8\,]}{1 - P[X \le 5]} = \frac{e^{-0.1\times 8}}{e^{-0.1\times 5}} = e^{-0.3} = 0.7408$$

4. Derive the expectation and variance of the waiting time [i.e., use the definition of a continuous random variable to calculate E[X] and V[X].

$$\mu = E[X] = \int_0^\infty y \times 0.1e^{-0.1y}\, dy = -\int_0^\infty ye^{-0.1y}\, d(-0.1y) = -\int_0^\infty y\, d(e^{-0.1y})$$

$$= -\left(ye^{-0.1y}\big|_0^\infty - \int_0^\infty e^{-0.1y}\, dy\right) = \int_0^\infty e^{-0.1y}\, dy$$

$$= -10\int_0^\infty e^{-0.1y}\,d(-0.1y) = -10e^{-0.1y}|_0^\infty = -(10e^{-0.1\times\infty} - 10e^{-0.1\times 0}) = 10.$$

$$E[X^2] = \int_0^\infty y^2 \times 0.1e^{-0.1y}\,dy = -\int_0^\infty y^2\,d(e^{-0.1y}) = -\left(y^2e^{-0.1y}|_0^\infty - \int_0^\infty e^{-0.1y}\,dy^2\right)$$
$$= \int_0^\infty e^{-0.1y}\,dy^2 = \int_0^\infty 2ye^{-0.1y}\,dy = \int_0^\infty 20y \times 0.1e^{-0.1y}\,dy = 20\int_0^\infty y \times 0.1e^{-0.1y}\,dy = 20\mu = 200$$

$$V[X] = E[X^2] - \mu^2 = 200 - 10^2 = 100.$$