Homework 2 - Solutions

Math for AI

Instructions: Please solve the following problems showing all your work. State any assumptions you make.

1. **Bespoke continuous RV**

   Let X be a random variable with pdf

$$f(x) = \begin{cases} 90x^8(1-x) & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

   a) Compute the cumulative distribution function of X.
   b) Compute $p(0.25 < X < 0.5)$
   c) Compute $E(X)$ and $Var(X)$.

   Solution:

   a) For $0 < x < 1$: $F(x) = \int_0^x 90t^8(1-t)dt = \int_0^x (90t^8 - 90t^9)dt = [10t^9 - 9t^{10}]_0^x = 10x^9 - 9x^{10}$.

$$F(x) = \begin{cases} 0 & x \leq 0 \\ 10x^9 - 9x^{10} & 0 < x < 1 \\ 1 & x \geq 1 \end{cases}$$

   b) $P(0.25 < X < 0.5) = F(0.5) - F(0.25) = [10(0.5)^9 - 9(0.5)^{10}] - [10(0.25)^9 - 9(0.25)^{10}] \approx 0.01074 - 0.00003 \approx 0.01071$.

   c) $E(X) = \int_0^1 x \cdot 90x^8(1-x)dx = \int_0^1 (90x^9 - 90x^{10})dx = [9x^{10} - \frac{90}{11}x^{11}]_0^1 = 9 - \frac{90}{11} = \frac{9}{11} \approx 0.8182$.
   $E(X^2) = \int_0^1 x^2 \cdot 90x^8(1-x)dx = \int_0^1 (90x^{10} - 90x^{11})dx = [\frac{90}{11}x^{11} - \frac{90}{12}x^{12}]_0^1 = \frac{90}{11} - \frac{15}{2} = \frac{180-165}{22} = \frac{15}{22}$.
   $Var(X) = E(X^2) - [E(X)]^2 = \frac{15}{22} - (\frac{9}{11})^2 = \frac{15}{22} - \frac{81}{121} = \frac{165-162}{242} = \frac{3}{242} \approx 0.0124$.

2. **Extreme value distribution**

   For what value of the parameter C is the function below a valid probability density function:

$$p(x) = Ce^{-x-e^{-x}}, x \in \mathbb{R}$$

   With the correct value for C this is the pdf of the extreme-value distribution. Find the cumulative distribution function of the extreme value distribution.

   Solution: To be a valid PDF, $\int_{-\infty}^{\infty} Ce^{-x-e^{-x}}dx = 1$. Let $u = e^{-x}$, then $du = -e^{-x}dx$. When $x \to \infty, u \to 0$. When $x \to -\infty, u \to \infty$. $\int_\infty^0 Ce^{-u}(-du) = C\int_0^\infty e^{-u}du = C[-e^{-u}]_0^\infty =$

$C(0-(-1)) = C$. Thus, $\mathbf{C = 1}$. The CDF is $F(x) = \int_{-\infty}^{x} e^{-t-e^{-t}} dt$. Using the same substitution $u = e^{-t}$: $F(x) = \int_{\infty}^{e^{-x}} -e^{-u} du = \int_{e^{-x}}^{\infty} e^{-u} du = [-e^{-u}]_{e^{-x}}^{\infty} = 0 - (-e^{-e^{-x}}) = e^{-e^{-x}}$.

3. **Expected area**

   An ecologist wishes to mark off a circular sampling region having radius $10m$. However, the radius of the resulting region is actually a random variable R with pdf

$$f(r) = \begin{cases} \frac{3}{4}[1-(10-r)^2] & 9 < r < 11 \\ 0 & \text{otherwise} \end{cases}$$

   What is the expected area of the resulting circular region?

   Solution: Area $A = \pi R^2$. We need $E(\pi R^2) = \pi E(R^2)$. Let $u = r - 10$, then $du = dr$. When $r = 9, u = -1$; when $r = 11, u = 1$. $E(R^2) = \int_9^{11} r^2 \frac{3}{4}[1-(10-r)^2]dr = \int_{-1}^{1} (u+10)^2 \frac{3}{4}(1-u^2)du$ $= \frac{3}{4}\int_{-1}^{1}(u^2+20u+100)(1-u^2)du = \frac{3}{4}\int_{-1}^{1}(-u^4-20u^3+99u^2+20u+100)du$ Odd powers integrate to 0 over $[-1,1]$: $= \frac{3}{4} \cdot 2\int_0^1(-u^4+99u^2+100)du = \frac{3}{2}[-\frac{1}{5}u^5+33u^3+100u]_0^1 = \frac{3}{2}(-\frac{1}{5}+33+100) = \frac{3}{2}(132.8) = 199.2$. Expected Area $= 199.2\pi \approx 625.80$ m^2.

4. **Assignment solving distribution**

   The length of time X (in days), needed by students in MAI to complete this assignment is a random variable with PDF given by

$$f(x) = \begin{cases} k(x^2+x) & 0 \le x < 1 \\ 0 & \text{otherwise} \end{cases}$$

   (a) Find the value of k that makes $f(x)$ a probability density function

   (b) Find the cumulative distribution function.

   (c) There are 38 students in our MAI class. What is the probability that eight or nine or them will complete this assignment in under half a day. Assume that completion times are independent.

   Solution:

   a) $\int_0^1 k(x^2+x)dx = k[\frac{x^3}{3}+\frac{x^2}{2}]_0^1 = k(\frac{1}{3}+\frac{1}{2}) = k(\frac{5}{6}) = 1 \implies k = 1.2$.

   b) For $0 \le x < 1$: $F(x) = \int_0^x 1.2(t^2+t)dt = 1.2(\frac{x^3}{3}+\frac{x^2}{2}) = 0.4x^3 + 0.6x^2$.

   c) Let p be the probability a student finishes in under half a day ($x < 0.5$): $p = F(0.5) = 0.4(0.5)^3 + 0.6(0.5)^2 = 0.05 + 0.15 = 0.2$. Let Y be the number of students. $Y \sim$ Binomial$(n = 38, p = 0.2)$. $P(Y = 8 \text{ or } Y = 9) = \binom{38}{8}(0.2)^8(0.8)^{30} + \binom{38}{9}(0.2)^9(0.8)^{29} \approx 0.1522 + 0.1438 = 0.2960$.

5. **Hotline**

   Suppose that the time it takes for a hotline operator to fill out an electronic form registering tips from concerned citizens, is uniformly between 1.5 and 11.5 minutes.

   (a) What are the mean and variance of the time it takes an operator to fill out the form?

(b) What is the probability that it will take less than two minutes to fill out the form?

(c) What is the probability that it will take between 5 minutes and 10 minutes to fill out the form?

(d) 99% of the forms take less than how many minutes to fill out?

(e) By the end of one an hour, an operator has completed 10 electronic forms. What is the probability that at least two of them took longer than 7 minutes to fill out? Assume that completion times are independent.

Solution: $X \sim \text{Uniform}(a = 1.5, b = 11.5)$. $f(x) = \frac{1}{11.5-1.5} = 0.1$.

a) $\mu = \frac{a+b}{2} = \frac{13}{2} = 6.5$ min. $\sigma^2 = \frac{(b-a)^2}{12} = \frac{10^2}{12} = \frac{100}{12} \approx 8.333$ min^2.

b) $P(X < 2) = \frac{2-1.5}{10} = \frac{0.5}{10} = 0.05$.

c) $P(5 < X < 10) = \frac{10-5}{10} = 0.5$.

d) $F(x) = 0.99 \implies \frac{x-1.5}{10} = 0.99 \implies x - 1.5 = 9.9 \implies x = 11.4$ minutes.

e) Let $p = P(X > 7) = \frac{11.5-7}{10} = 0.45$. Let $Y \sim \text{Binomial}(n = 10, p = 0.45)$. $P(Y \geq 2) = 1 - P(Y = 0) - P(Y = 1) = 1 - (0.55)^{10} - 10(0.45)(0.55)^9 \approx 1 - 0.0025 - 0.0207 = 0.9768$.

6. **Dangerous driving**

The distance between major cracks in a highway follows an exponential distribution with a mean of five miles.

(a) What is the probability that there are two or three major cracks in a 10-mile stretch of the highway?

(b) What is the probability that the distance between two successive cracks is greater than 10-miles?

(c) What is the probability that the first major crack occurs between 12 and 15 miles of the start of inspection?

(d) Given that there are no cracks in the first five miles inspected, what is the probability that there are no major cracks in the next 10 miles inspected?

Solution: Mean distance $\beta = 5$, so rate $\lambda = 1/5 = 0.2$ cracks/mile.

a) For a 10-mile stretch, let N be the number of cracks. $N \sim \text{Poisson}(\mu = \lambda L = 0.2 \cdot 10 = 2)$. $P(N = 2 \text{ or } 3) = \frac{e^{-2}2^2}{2!} + \frac{e^{-2}2^3}{3!} = e^{-2}(2 + \frac{8}{6}) = e^{-2}(3.333) \approx 0.451$.

b) Let X be the distance. $X \sim \text{Exp}(0.2)$. $P(X > 10) = e^{-0.2(10)} = e^{-2} \approx 0.1353$.

c) $P(12 < X < 15) = e^{-0.2(12)} - e^{-0.2(15)} = e^{-2.4} - e^{-3} \approx 0.0907 - 0.0498 = 0.0409$.

d) By the memoryless property of the exponential distribution: $P(X > 15 | X > 5) = P(X > 10) = e^{-0.2(10)} = e^{-2} \approx 0.1353$.

7. **Symmetric Distribution and CLT**

Let X be a continuous random variable with the probability density function:

$$f(x) = \begin{cases} \frac{3}{4}(1 - x^2) & -1 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

(a) Compute $E(X)$ and $Var(X)$.

(b) Suppose we take a random sample of $n = 75$ independent observations of X. Let $\bar{X}$ be the sample mean. Use the Central Limit Theorem to approximate $P(\bar{X} > 0.05)$.

Solution:

a) $E(X) = \int_{-1}^{1} x \cdot \frac{3}{4}(1 - x^2)dx = 0$ (Odd function over symmetric interval). $Var(X) = E(X^2) = \int_{-1}^{1} x^2 \cdot \frac{3}{4}(1-x^2)dx = \frac{3}{4}[\frac{x^3}{3} - \frac{x^5}{5}]_{-1}^{1} = \frac{3}{4}[(\frac{1}{3} - \frac{1}{5}) - (-\frac{1}{3} + \frac{1}{5})] = \frac{3}{4}[\frac{2}{15} \cdot 2] = \frac{1}{5} = 0.2.$

b) By CLT, $\bar{X} \approx N(\mu, \sigma^2/n) = N(0, 0.2/75) = N(0, 0.002667)$. $\sigma_{\bar{X}} = \sqrt{0.2/75} \approx 0.05164$. $P(\bar{X} > 0.05) = P(Z > \frac{0.05-0}{0.05164}) = P(Z > 0.968) \approx 1 - 0.8335 = 0.1665$.

8. **Uniform distribution and CLT**

Suppose we take a random sample of $n = 90$ independent observations from the uniform distribution on the interval $[-1, 1]$. Determine the probability that the sample mean is in the range $[-0.1, 0.1]$.

Solution: For $U(-1, 1)$: $\mu = 0$, $\sigma^2 = \frac{(1-(-1))^2}{12} = \frac{4}{12} = \frac{1}{3}$. For $n = 90$, $\bar{X} \approx N(0, \frac{1/3}{90}) = N(0, \frac{1}{270})$. $\sigma_{\bar{X}} = \sqrt{1/270} \approx 0.06086$. $P(-0.1 < \bar{X} < 0.1) = P(\frac{-0.1}{0.06086} < Z < \frac{0.1}{0.06086}) = P(-1.643 < Z < 1.643)$. Using Z-tables: $\Phi(1.643) - \Phi(-1.643) \approx 0.9498 - 0.0502 = 0.8996$.