24400 HW2

Bin Yu

October 15, 2024

1 Question 1

Let $f(x)$ be the probability density function of a continuous random variable X:

$$f(x) = \begin{cases} 0 & x \leq 0, \\ x & 0 < x < 1, \\ \frac{c}{x^3} & 1 \leq x < \infty. \end{cases}$$

(a)

If $f(x)$ is a PDF, the total area under the curve must integrate to 1:

$$\int_{-\infty}^{\infty} f(x)\,dx = 1.$$

Thus,

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{0} 0\,dx + \int_{0}^{1} x\,dx + \int_{1}^{\infty} \frac{c}{x^3}\,dx.$$

$$\int_{-\infty}^{0} 0\,dx = 0$$

$$\int_{0}^{1} x\,dx = \left[\frac{x^2}{2}\right]_0^1 = \frac{1}{2}.$$

$$\int_{1}^{\infty} \frac{c}{x^3}\,dx = \left[-\frac{c}{2x^2}\right]_1^{\infty} = \frac{c}{2}.$$

Therefore,

$$\int_{-\infty}^{\infty} f(x)\,dx = 0 + \frac{1}{2} + \frac{c}{2} = 1.$$

$$c = 1.$$

(b)

The CDF is given by:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt.$$

For different values of x:

- If $x \leq 0$, $F(x) = 0$.
- If $0 < x < 1$,
$$F(x) = 0 + \int_0^x t\,dt = \frac{x^2}{2}.$$
- If $x \geq 1$,
$$F(x) = 0+\int_0^1 t\,dt+\int_1^x \frac{1}{t^3}\,dt = \frac{1}{2}+\left[-\frac{1}{2t^2}\right]_1^x = \frac{1}{2}+\frac{1}{2}\left(1-\frac{1}{x^2}\right) = 1-\frac{1}{2x^2}.$$

Thus, the CDF is:

$$F(x) = \begin{cases} 0 & x \leq 0, \\ \frac{x^2}{2} & 0 < x < 1, \\ 1 - \frac{1}{2x^2} & x \geq 1. \end{cases}$$

(c)

$$P(0.5 \leq X \leq 1.5) = F(1.5) - F(0.5).$$

From part (b), we know:

$$F(1.5) = 1 - \frac{1}{2(1.5)^2} = 1 - \frac{1}{2 \times 2.25} = \frac{7}{9} \approx 0.7778.$$

$$F(0.5) = \frac{(0.5)^2}{2} = \frac{0.25}{2} = 0.125.$$

Thus,

$$P(0.5 \leq X \leq 1.5) \approx 0.7778 - 0.125 \approx 0.6528.$$

Question 2

(a)

If $F(x)$ is a valid cumulative distribution function (CDF), it must be continuous at $x = \theta$. Thus, the limit as $x \to \theta$ from both sides should be equal:

$$\lim_{x\to\theta^-} F(x) = 0 = \lim_{x\to\theta^+} F(x) = 1 - c\theta^{-k}.$$

$$1 - c\theta^{-k} = 0.$$

$$c = \theta^k.$$

(b)

$$P(X > 3\theta) = 1 - P(X \leq 3\theta).$$

Using the given CDF $F(x)$:

$$P(X \leq 3\theta) = 1 - c(3\theta)^{-k}.$$

Substituting $c = \theta^k$ into the equation:

$$P(X \leq 3\theta) = 1 - \theta^k(3\theta)^{-k} = 1 - \left(\frac{1}{3}\right)^k.$$

Thus,

$$P(X > 3\theta) = 1 - [1 - \left(\frac{1}{3}\right)^k] = \left(\frac{1}{3}\right)^k.$$

(c)

To find the probability density function (PDF) $f(x)$, need to differentiate the CDF $F(x)$. Given:

$$F(x) = \begin{cases} 0, & \text{if } x < \theta, \\ 1 - cx^{-k}, & \text{if } x \geq \theta, \end{cases}$$

For $x < \theta$:

$$f(x) = \frac{d}{dx}F(x) = \frac{d}{dx}0 = 0.$$

For $x \geq \theta$:

$$f(x) = \frac{d}{dx}(1 - cx^{-k}) = 0 - c(-k)x^{-k-1} = ckx^{-k-1}.$$

Substituting $c = \theta^k$:

$$f(x) = k\theta^k x^{-k-1}, \quad x \geq \theta.$$

Therefore, the probability density function is:

$$f(x) = \begin{cases} 0, & \text{if } x < \theta, \\ k\theta^k x^{-k-1}, & \text{if } x \geq \theta. \end{cases}$$

Question 3

(a)

Let $X \sim \text{Geometric}(0.8)$, with parameter $p = 0.8$. The probability that X is even is:

$$P(X = 2k) = (0.2)^{2k-1} \cdot 0.8$$

Summing over all possible even values of X:

$$P(X \text{ is even})$$

$$= \sum_{k=1}^{\infty} (0.2)^{2k-1} \cdot 0.8$$

$$= 0.8 \cdot 0.2 \cdot \sum_{k=1}^{\infty} (0.2^2)^{k-1}$$

$$= 0.16 \cdot \sum_{k=0}^{\infty} (0.04)^k$$

Since we have the formula: $\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}$, and here $x = 0.04$:

$$P(X \text{ is even}) = 0.16 \cdot \frac{1}{1 - 0.04} = 0.16 \cdot \frac{1}{0.96} = \frac{1}{6}.$$

Thus, $P(X \text{ is even}) = \frac{1}{6}$.

(b)

Let $X \sim \text{Binomial}(101, 0.5)$. The probability mass function is:

$$f(x) = \binom{n}{k} (1-p)^{n-k} p^k,$$

where $p = 0.5$, $n = 101$, and since we want to calculate $P(X \text{ is even})$,we can set $k = 2k$ for even values of X.

$$P(X \text{ is even})$$

$$= \sum_{k=0}^{50} \binom{101}{2k} (0.5)^{101-2k} (0.5)^{2k}.$$

$$= \sum_{k=0}^{50} \binom{101}{2k} (0.5)^{101}.$$

$$= (0.5)^{101} \sum_{k=0}^{50} \binom{101}{2k}.$$

Since $\binom{101}{2k} = \binom{101}{101-2k}$, we can infer:

$$\sum_{k=0}^{50} 2\binom{101}{2k}$$

$$= \sum_{k=0}^{50} [\binom{101}{2k} + \binom{101}{101-2k}]$$

$$= \sum_{k=0}^{101} \binom{101}{k}$$

$$= 2^{101}$$

(Each of the 101 elements can be in the subset or not in the subset) Thus,

$$P(X \text{ is even})$$

$$= \frac{1}{2^{101}} \cdot \frac{1}{2} \cdot \sum_{k=0}^{50} 2\binom{101}{2k}$$

$$= \frac{1}{2^{101}} \cdot \frac{1}{2} \cdot \sum_{k=0}^{101} \binom{101}{k}$$

$$= \frac{1}{2^{101}} \cdot \frac{1}{2} \cdot 2^{101}.$$

$$= \frac{1}{2}$$

Thus, $P(X \text{ is even}) = \frac{1}{2}$.

Question 4

(a)

For a uniform distribution over the interval $[a, b]$, the probability density function (PDF) is:

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & \text{for } x \in [a, b] \\ 0, & \text{otherwise.} \end{cases}$$

The expected value of a uniformly distributed random variable $X \sim \text{Uniform}(a, b)$ is:

$$E(X) = \int_a^b x \cdot \frac{1}{b-a}\, dx.$$

$$= \frac{1}{b-a} \int_a^b x\, dx.$$

$$= \frac{1}{b-a} \cdot \left(\frac{b^2}{2} - \frac{a^2}{2} \right)$$

$$= \frac{b^2 - a^2}{2(b-a)}.$$

$$= \frac{(b-a)(b+a)}{2(b-a)} = \frac{a+b}{2}.$$

For the uniform distribution $X \sim \text{Uniform}(0, t_i)$, substituting $a = 0$ and $b = t_i$:

$$E(X) = \frac{0 + t_i}{2} = \frac{t_i}{2}.$$

The expected value of the sum $S = X_1 + X_2 + \cdots + X_n$ is:

$$E(S) = E(X_1 + X_2 + \cdots + X_n) = E(X_1) + E(X_2) + \cdots + E(X_n).$$

$$E(S) = \frac{t_1}{2} + \frac{t_2}{2} + \cdots + \frac{t_n}{2} = \frac{t_1 + t_2 + \cdots + t_n}{2}.$$

Thus, the expected value of the sum S is:

$$E(S) = \frac{t_1 + t_2 + \cdots + t_n}{2}, \quad t_i \in (0, \infty).$$

(b)

For each $i = 1, 2, \ldots, n$, define:

$$I_i = \begin{cases} 1, & \text{if } X_i \leq 1; \\ 0, & \text{if } X_i > 1. \end{cases}$$

Thus,

$$Y = I_1 + I_2 + \cdots + I_n.$$

The expected value of Y can be expressed as:

$$E[Y] = E[I_1 + I_2 + \cdots + I_n] = E[I_1] + E[I_2] + \cdots + E[I_n].$$

Since I_i are indicator variables,

$$E[I_i] = P(I_i = 1) = P(X_i \leq 1).$$

Therefore,

$$E[Y] = \sum_{i=1}^{n} P(X_i \leq 1).$$

Since X_i is uniformly distributed on $[0, t_i]$, its PDF is:

$$f_{X_i}(x) = \begin{cases} \frac{1}{t_i}, & \text{for } x \in [0, t_i]; \\ 0, & \text{otherwise.} \end{cases}$$

The CDF is:

$$F_{X_i}(x) = P(X_i \leq x) = \begin{cases} 0, & x < 0; \\ \frac{x}{t_i}, & 0 \leq x \leq t_i; \\ 1, & x > t_i. \end{cases}$$

To find out $P(X_i \leq 1)$, We need to compute $F_{X_i}(1)$:

- **When $t_i \leq 1$:**
  Since $x > t_i$:
  $$P(X_i \leq 1) = 1.$$

- **When $t_i > 1$:**
  $$P(X_i \leq 1) = F_{X_i}(1) = \frac{1}{t_i}.$$

Therefore, $P(X_i \leq 1)$ can be expressed as:

$$P(X_i \leq 1) = \min\left(1, \frac{1}{t_i}\right).$$

Thus,

$$E[Y] = \sum_{i=1}^{n} \min\left(1, \frac{1}{t_i}\right), \quad t_i \in (0, \infty)$$

Question 5

(a)

Since $X \in [0, 1)$:
When $X = 0$:

$$R = \frac{0}{1 - 0} = 0.$$

As X approaches 1:

$$R = \frac{X}{1 - X} \to \frac{1}{0^+} \to \infty.$$

Therefore, R is defined in support of $[0, \infty)$
The general formula for the PDF of a transformed variable $Y = g(X)$ is:

$$f_Y(y) = f_X(g^{-1}(y)) \cdot \left| \frac{d}{dy} g^{-1}(y) \right|$$

Given:

$$R = g(X) = \frac{X}{1-X}$$

To find the inverse function $g^{-1}(R)$, solve for X in terms of R:

$$R = \frac{X}{1-X} \quad \Rightarrow \quad R(1-X) = X \quad \Rightarrow \quad R - RX = X$$

$$R = X + RX \quad \Rightarrow \quad R = X(1+R) \quad \Rightarrow \quad X = \frac{R}{1+R}$$

$$X = g^{-1}(R) = \frac{R}{1+R}$$

The derivative of $g^{-1}(R)$ with respect to R:

$$\frac{d}{dR} g^{-1}(R) = \frac{d}{dR} \left(\frac{R}{1+R} \right) = \frac{(1+R) \cdot 1 - R \cdot 1}{(1+R)^2} = \frac{1}{(1+R)^2}$$

Since $X \sim \text{Uniform}[0, 1]$, its PDF is:

$$f_X(x) = \begin{cases} 1, & \text{if } 0 \leq x \leq 1; \\ 0, & \text{otherwise.} \end{cases}$$

$$f_R(R) = f_X\left(g^{-1}(R)\right) \cdot \left| \frac{d}{dR} g^{-1}(R) \right|$$

$$f_R(R) = f_X\left(\frac{R}{1+R} \right) \cdot \frac{1}{(1+R)^2}$$

Since $\frac{R}{1+R}$ lies within $[0, 1)$ for $R \geq 0$, $f_X\left(\frac{R}{1+R}\right) = 1$. Therefore:

$$f_R(R) = \frac{1}{(1+R)^2}, \quad \text{for } R \geq 0$$

The probability density function of $R = \frac{X}{1-X}$ is:

$$f_R(R) = \begin{cases} \frac{1}{(1+R)^2}, & R \geq 0; \\ 0, & \text{otherwise.} \end{cases}$$

(b)

Given $Y = X(1-X)$ and $X \in [0,1)$:

- **When $X = 0$:**

$$Y = 0(1-0) = 0$$

- **As X approaches 1 from below ($X \to 1^-$):**

$$Y = X(1-X) = 1 \cdot (1-1) = 0$$

- **Maximum of Y:**

$$\frac{dY}{dX} = 1 - 2X = 0 \quad \Rightarrow \quad X = \frac{1}{2}$$

$$Y_{\max} = \frac{1}{2}\left(1 - \frac{1}{2}\right) = \frac{1}{4}$$

The range of Y is:

$$Y \in [0, 0.25]$$

To find the inverse function $g^{-1}(Y)$, solve for X in terms of Y:

$$Y = X - X^2$$

$$X^2 - X + Y = 0$$

$$X = \frac{1 \pm \sqrt{1-4Y}}{2}$$

$$X_1 = \frac{1 + \sqrt{1-4Y}}{2}, \quad X_2 = \frac{1 - \sqrt{1-4Y}}{2}$$

Now calculate $\frac{dX}{dY}$ for each inverse transformation.
For X_1:

$$X_1 = \frac{1 + \sqrt{1-4Y}}{2}$$

$$\frac{dX_1}{dY} = \frac{1}{2} \cdot \frac{-4}{2\sqrt{1-4Y}} = \frac{-2}{2\sqrt{1-4Y}} = \frac{-1}{\sqrt{1-4Y}}$$

$$\left|\frac{dX_1}{dY}\right| = \frac{1}{\sqrt{1-4Y}}$$

For X_2:

$$X_2 = \frac{1 - \sqrt{1-4Y}}{2}$$

$$\frac{dX_2}{dY} = \frac{1}{2} \cdot \frac{4}{2\sqrt{1-4Y}} = \frac{2}{2\sqrt{1-4Y}} = \frac{1}{\sqrt{1-4Y}}$$

$$\left|\frac{dX_2}{dY}\right| = \frac{1}{\sqrt{1-4Y}}$$

Since:

$$f_Y(Y) = \sum_i f_X(g_i^{-1}(Y)) \cdot \left|\frac{d}{dY} g_i^{-1}(Y)\right|$$

Given that $X \sim \text{Uniform}[0,1]$, its PDF is:

$$f_X(X) = \begin{cases} 1, & 0 \le X \le 1; \\ 0, & \text{otherwise.} \end{cases}$$

$$f_Y(Y) = f_X(X_1) \cdot \left|\frac{dX_1}{dY}\right| + f_X(X_2) \cdot \left|\frac{dX_2}{dY}\right|$$

Since $X_1, X_2 \in [0,1]$ for $Y \in [0, 0.25]$, $f_X(X_1) = f_X(X_2) = 1$.
Therefore:

$$f_Y(Y) = \frac{1}{\sqrt{1-4Y}} + \frac{1}{\sqrt{1-4Y}} = \frac{2}{\sqrt{1-4Y}}, \quad \text{for } Y \in [0, 0.25]$$

Therefore,

$$f_Y(Y) = \begin{cases} \frac{2}{\sqrt{1-4Y}}, & Y \in [0, 0.25]; \\ 0, & \text{otherwise.} \end{cases}$$

Question 6

(a)

For an exponential distribution with parameter $\lambda > 0$:

$$f(t) = \lambda e^{-\lambda t}, \quad \text{for } t \ge 0.$$

$$F(t) = P(T \le t) = \int_0^t f(x)\, dx = \int_0^t \lambda e^{-\lambda x}\, dx.$$

$$F(t) = \int_0^t \lambda e^{-\lambda x}\, dx.$$

Let $u = -\lambda x$, then $du = -\lambda\, dx$.

$$\begin{cases} x = 0 & \Rightarrow u = 0, \\ x = t & \Rightarrow u = -\lambda t. \end{cases}$$

$$F(t) = \int_0^t \lambda e^{-\lambda x}\, dx = \lambda \int_0^t e^{-\lambda x}\, dx = \lambda \int_0^{-\lambda t} e^u \left(-\frac{1}{\lambda}\right) du.$$

$$F(t) = -\int_0^{-\lambda t} e^u\, du.$$

$$F(t) = \int_{-\lambda t}^{0} e^u \, du.$$

$$= [e^u]_{-\lambda t}^{0} = e^0 - e^{-\lambda t} = 1 - e^{-\lambda t}.$$

Using the definition of the hazard rate:

$$h(t) = \frac{f(t)}{1 - F(t)}.$$

$$h(t) = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda$$

(b)

Given:
CDF of Weibull distribution, where $\quad k, \alpha > 0$:

$$F(t) = 1 - e^{-(t/\alpha)^k}, \quad \text{for } t \geq 0.$$

in support: $t \in [0, \infty)$.
The PDF is the derivative of the CDF:

$$f(t) = \frac{d}{dt} F(t).$$

$$f(t) = \frac{d}{dt} \left(1 - e^{-(t/\alpha)^k} \right)$$
$$= 0 - (-e^{-(t/\alpha)^k}) \cdot \frac{d}{dt} \left((t/\alpha)^k \right).$$
$$= e^{-(t/\alpha)^k} \cdot \frac{d}{dt} \left((t/\alpha)^k \right).$$

And

$$\frac{d}{dt} \left((t/\alpha)^k \right) = k \left(\frac{t}{\alpha} \right)^{k-1} \cdot \frac{1}{\alpha} = \frac{k}{\alpha^k} t^{k-1}.$$

Thus,

$$f(t) = \frac{k}{\alpha^k} t^{k-1} e^{-(t/\alpha)^k}, \quad \text{for } t \geq 0.$$

Using the definition:

$$h(t) = \frac{f(t)}{1 - F(t)}.$$

Substitute $f(t)$ and $1 - F(t)$:

$$h(t) = \frac{\frac{k}{\alpha^k} t^{k-1} e^{-(t/\alpha)^k}}{e^{-(t/\alpha)^k}} = \frac{k}{\alpha^k} t^{k-1}.$$

$$h(t) = \frac{k}{\alpha} \left(\frac{t}{\alpha} \right)^{k-1}.$$

(c)

Given the hazard rate function for the Weibull distribution:

$$h(t) = \frac{k}{\alpha}\left(\frac{t}{\alpha}\right)^{k-1}, \quad \text{for } t \geq 0.$$

We can see that the α is not related to whether $h(t)$ is decreasing over time, increasing over time, or constant over time, since both k/a and $(1/a)^{k-1}$ is constant and not related to t.
The behavior of the hazard rate $h(t)$ over time depends primarily on the k:

1. **Constant Hazard Rate:**
   When $k = 1$:
   $$h(t) = \frac{1}{\alpha}$$
   the hazard rate is **constant** over time.

2. **Increasing Hazard Rate:**
   When $k > 1$, the term $\left(\frac{t}{\alpha}\right)^{k-1}$ increases as t increases.
   Therefore, $h(t)$ **increases** over time.

3. **Decreasing Hazard Rate:**
   When $0 < k < 1$, the term $\left(\frac{t}{\alpha}\right)^{k-1}$ decreases as t increases.
   Therefore, $h(t)$ **decreases** over time.

Therefore,

$$h(t) = \begin{cases} \text{Decreasing over time,} & \text{if } 0 < k < 1; \\ \text{Constant over time,} & \text{if } k = 1; \\ \text{Increasing over time,} & \text{if } k > 1. \end{cases}$$

For all $k > 0$ and $\alpha > 0$.

(d)

(i) Monotone Increasing Hazard Rate

Example: The Probability of Developing Alzheimer's Disease
T could measure the time until an individual develops Alzheimer's disease. When an individual is younger, the probability of developing Alzheimer's is low because the brain is in good condition. As the individual ages, the likelihood of developing Alzheimer's increases due to factors like age-related cognitive decline and other health-related changes. The hazard rate $h(t)$ in this case starts low and increases over time, reflecting the growing risk of developing Alzheimer's as the individual gets older.

(ii) "U" Shaped Hazard Rate

Example: Human Mortality Rate Across a Lifetime
T could measure the time until death for humans across their entire lifespan. When people are infants, there will be a high hazard rate after birth due to vulnerability. After they grow up, there will be a low hazard rate as people are generally healthy when they are in their young and middle ages. When people are getting older, there will be an increasing hazard rate due to aging, where $h(t)$ is high initially, decreases to a low point, and then rises again.