

Homework 2

Problem 1

Suppose that the cumulative distribution function (CDF) of a random variable $F : \mathbb{R} \rightarrow \mathbb{R}$ is strictly increasing and continuous. Let U be a random variable with the uniform distribution on $(0, 1)$ and define

$$X := F^{-1}(U)$$

Show that X has CDF equal to F . This exercise shows us how to construct a random variable with given distribution, assuming that we have a uniform random variable.

Proof Since F is strictly increasing and continuous, it has an inverse function F^{-1} on its range, and F^{-1} is also strictly increasing. Thus for any $x, y \in \mathbb{R}$,

$$F^{-1}(y) \leq x \iff y \leq (F^{-1})^{-1}(x) = F(x)$$

Therefore for any $x \in \mathbb{R}$, we have

$$\{x \mid X(x) \leq x\} = \{x \mid F^{-1}(U(x)) \leq x\} = \{x \mid U(x) \leq F(x)\}$$

Therefore

$$\mathbb{P}(X \leq x) = \mathbb{P}(U(x) \leq F(x))$$

Since $U \sim \text{Unif}(0, 1)$ and for a CDF we have $F(x) \in [0, 1]$, we get

$$\mathbb{P}(U(x) \leq F(x)) = F(x)$$

Thus for all x $\mathbb{P}(X \leq x) = F(x)$, i.e., the CDF of X equals F .

Problem 2

A gas station fills its tank completely once a week. Let the weekly sales volume (in thousands of liters) be a random variable with density

$$f(x) = \begin{cases} a(1-x)^4, & x \in (0, 1), \\ 0, & \text{otherwise} \end{cases}$$

Find the constant a . What should be the tank capacity so that the probability of running out of fuel during a given week is $1/100$?

Sol. Since the density integrates to 1,

$$1 = \int_{-\infty}^{\infty} f(x) dx = \int_0^1 a(1-x)^4 dx = a \int_0^1 (1-x)^4 dx$$

Let $u = 1 - x$, then

$$\int_0^1 (1-x)^4 dx = \int_0^1 u^4 du = \frac{1}{5}$$

So $\frac{1}{5}a = 1$, which gives

$$a = 5$$

Now we look for the tank capacity c such that $\mathbb{P}(X > c) = \frac{1}{100}$.

Let the tank capacity be c (in thousands of liters). Running out of fuel in a week occurs when sales exceed c , i.e., the event $\{X > c\}$. We need

$$\mathbb{P}(X > c) = \frac{1}{100}$$

Since $a = 5$,

$$\mathbb{P}(X > c) = \int_c^1 5(1-x)^4 dx$$

Again let $u = 1 - x$, then we have

$$\int_c^1 5(1-x)^4 dx = 5 \int_{1-c}^0 u^4 (-du) = 5 \int_0^{1-c} u^4 du = 5 \cdot \frac{(1-c)^5}{5} = (1-c)^5$$

Therefore

$$(1-c)^5 = \frac{1}{100} \implies 1-c = 100^{-1/5} = 10^{-2/5} \implies c = 1 - 10^{-2/5}$$

So the tank capacity should be $1 - 10^{-2/5}$ thousand liters.

Problem 3

Let the random variable X have density

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

Find the probability density function of $Y := X^2$ and compute the probability $\mathbb{P}(2Y + 3 \leq 10)$.

Sol. Since $f_X(x) = 0$ for $|x| < 1$, we have $\mathbb{P}(|X| \geq 1) = 1$. Hence $Y = X^2 \geq 1$ almost surely, so $F_Y(y) = 0$ for $y < 1$ and therefore $f_Y(y) = 0$ for $y < 1$ (a.e.).

For $y \geq 1$,

$$F_Y(y) = \mathbb{P}(X^2 \leq y) = \mathbb{P}(-\sqrt{y} \leq X \leq \sqrt{y}) = \int_{-\sqrt{y}}^{-1} \frac{1}{2x^2} dx + \int_1^{\sqrt{y}} \frac{1}{2x^2} dx$$

Compute each integral:

$$\int_1^{\sqrt{y}} \frac{1}{2x^2} dx = \frac{1}{2} \int_1^{\sqrt{y}} x^{-2} dx = \frac{1}{2} \left[-x^{-1} \right]_1^{\sqrt{y}} = \frac{1}{2} \left(1 - \frac{1}{\sqrt{y}} \right)$$

and similarly $\int_{-\sqrt{y}}^{-1} \frac{1}{2x^2} dx = \frac{1}{2} \left(1 - \frac{1}{\sqrt{y}} \right)$ since the function is even. Therefore, for $y \geq 1$,

$$F_Y(y) = 1 - \frac{1}{\sqrt{y}}$$

Combining both cases we have

$$F_Y(y) = \begin{cases} 0, & y < 1, \\ 1 - \frac{1}{\sqrt{y}}, & y \geq 1 \end{cases}$$

Notice that on $y \geq 1$, $F_Y(y)$ is differentiable (except on $y = 1$):

$$F'_Y(y) = \frac{d}{dy} \left(1 - y^{-1/2} \right) = \frac{1}{2} y^{-3/2}$$

So consider the function

$$g(y) = \begin{cases} \frac{1}{2y^{3/2}}, & y \geq 1, \\ 0, & y < 1. \end{cases}$$

Then for $x < 1$,

$$\int_{-\infty}^x g(y) dy = 0 = F_Y(x)$$

and for $x \geq 1$

$$\int_{-\infty}^x g(y) dy = \int_1^x \frac{1}{2y^{3/2}} dy = \left[-y^{-1/2} \right]_1^x = 1 - \frac{1}{\sqrt{x}} = F_Y(x)$$

This shows that Y is **absolutely continuous** and g is a **probability density** of Y . Hence

$$f_Y(y) = \begin{cases} \frac{1}{2y^{3/2}}, & y \geq 1, \\ 0, & y < 1 \end{cases}$$

Now we compute $\mathbb{P}(2Y + 3 \leq 10)$.

We have $2Y + 3 \leq 10 \iff Y \leq \frac{7}{2}$. Thus

$$\mathbb{P}(2Y + 3 \leq 10) = \mathbb{P}\left(Y \leq \frac{7}{2}\right) = F_Y\left(\frac{7}{2}\right) = 1 - \frac{1}{\sqrt{7/2}} = 1 - \sqrt{\frac{2}{7}}$$

Thus,

$$\mathbb{P}(2Y + 3 \leq 10) = 1 - \sqrt{\frac{2}{7}}$$

Problem 4

Let the random variable X have density f , which is symmetric about $\mu \in \mathbb{R}$, that is, $f(\mu + x) = f(\mu - x)$, for all $x \in \mathbb{R}$. Show that $\mathbb{P}(X \leq \mu) = \mathbb{P}(X \geq \mu)$. If in addition $\mathbb{E}|X| < \infty$, show that $\mathbb{E}(X) = \mu$. Can you use this observation if $X \sim N(0, 1)$?

Proof Since X has density f ,

$$\mathbb{P}(X \leq \mu) = \int_{-\infty}^{\mu} f(t) dt$$

Let $t = \mu - x$ so that $dt = -dx$. Then

$$\int_{-\infty}^{\mu} f(t) dt = \int_{\infty}^0 f(\mu - x)(-dx) = \int_0^{\infty} f(\mu - x) dx$$

Similarly,

$$\mathbb{P}(X \geq \mu) = \int_{\mu}^{\infty} f(t) dt = \int_0^{\infty} f(\mu + x) dx$$

By symmetry $f(\mu - x) = f(\mu + x)$ for all x , hence the two integrals are equal, i.e. proved

$$\mathbb{P}(X \leq \mu) = \mathbb{P}(X \geq \mu)$$

If $\mathbb{E}|X| < \infty$, then $\mathbb{E}[X] = \mu$ for some $\mu \in \mathbb{R}$. We want to show that this μ is the same as the one in the symmetry condition. Consider $\mathbb{E}[X - \mu]$. Since $\mathbb{E}|X| < \infty$, we also have $\mathbb{E}|X - \mu| < \infty$, so the following integral is well-defined:

$$\mathbb{E}[X - \mu] = \int_{-\infty}^{\infty} (t - \mu)f(t) dt$$

Let $t = \mu + x$; then

$$\mathbb{E}[X - \mu] = \int_{-\infty}^{\infty} x f(\mu + x) dx$$

Define $g(x) := f(\mu + x)$. The symmetry condition $f(\mu + x) = f(\mu - x)$ implies that g is an even function, thus $xg(x)$ is an odd function. Since $\int |x|g(x) dx < \infty$, we may integrate over symmetric limits to get

$$\int_{-\infty}^{\infty} xg(x) dx = 0$$

Therefore $\mathbb{E}[X - \mu] = 0$, thus

$$\mathbb{E}[X] = \mu$$

Application to $X \sim N(0, 1)$: Since the standard normal density $\varphi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ satisfies $\varphi(0 + x) = \varphi(0 - x)$, so it is symmetric about $\mu = 0$. Hence

$$\mathbb{P}(X \leq 0) = \mathbb{P}(X \geq 0) = \frac{1}{2} \quad \text{and} \quad \mathbb{E}[X] = 0$$

Problem 5

An airline has observed that 5% of ticket holders do not show up for their flight. Today's flight has an airplane with 200 seats, and the airline has sold 203 tickets. What is the probability that the airline will not be able to accommodate a ticketed passenger? Assume that, for each passenger i , the event A_i that passenger i shows up is independent of all others, for $1 \leq i \leq 203$.

Sol. Let S be the number of passengers who show up. The condition indicates that S is a binomial random variable with parameters $n = 203$ and $p = 0.95$:

$$S \sim \text{Binomial}(n = 203, p = 0.95)$$

The airline cannot accommodate everyone exactly when more than 200 passengers show up, i.e.

$$\mathbb{P}(\text{cannot accommodate}) = \mathbb{P}(S \geq 201) = \sum_{k=201}^{203} \binom{203}{k} (0.95)^k (0.05)^{203-k}$$

Equivalently, letting $N := 203 - S$ be the number of no-shows, we have $N \sim \text{Binomial}(203, 0.05)$ and

$$\mathbb{P}(S \geq 201) = \mathbb{P}(N \leq 2) = \sum_{j=0}^2 \binom{203}{j} (0.05)^j (0.95)^{203-j}$$

Numerically we can calculate

$$\mathbb{P}(\text{cannot accommodate}) \approx 0.206\%$$

Problem 6

Consider a sequence of tosses of a fair die. We continue tossing until both outcomes 3 and 4 have appeared at least once. For example, one possible sequence of results is

$$5, 1, 1, 4, 6, 5, 4, 2, 6, 3,$$

and we then stop. Let X be the number of tosses required (in this example, $X = 10$). What is the expected value of the random variable X ?

Sol. We can decompose the waiting time into two stages.

Stage 1: wait until the first time we see either 3 or 4: On each toss, the probability to get a 3 or 4 is $2/6 = 1/3$. Hence the number of tosses T_1 until the first occurrence of $\{3, 4\}$ is geometric with success probability $1/3$, so

$$\mathbb{E}[T_1] = \frac{1}{1/3} = 3$$

Stage 2: after seeing one of them, wait until we see the other: Once 3 has appeared, each subsequent toss produces a 4 with probability $1/6$; otherwise we are still missing a 4. Thus the additional waiting time T_2 is geometric with success probability $1/6$, so

$$\mathbb{E}[T_2] = \frac{1}{1/6} = 6$$

Since $X = T_1 + T_2$, by linearity of expectation we get

$$\mathbb{E}[X] = \mathbb{E}[T_1] + \mathbb{E}[T_2] = 3 + 6 = 9$$