Stat 330: Module 3 Homework Solution

Show all of your work, and upload this homework to Canvas.

1. Suppose a continuous random variable X has the following probability density function

$$f_X(x) = \begin{cases} kx(1-x) & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find the value of k that makes $f_X(x)$ a valid probability density function. (Recall a property that a PDF must have)

(b) Give the CDF, $F_X(x)$.

(c) Find $\mathbb{P}(0.5 \le X \le 1)$ using $f_X(x)$.

(d) Find $\mathbb{P}(0 \le X \le .75)$ using $F_X(x)$.

(e) Find $\mathbb{E}(X)$.

(f) Find $Var(X)$.

Answer:

(a) For a PDF to be valid, $\int_{-\infty}^{\infty} f_X(x)dx = 1$. We have:

$$\int_0^1 kx - kx^2 = 1$$

$$\frac{kx^2}{2} - \frac{kx^3}{3}\Big|_0^1 = \frac{k}{2} - \frac{k}{3}$$

$$\rightarrow \frac{k}{2} - \frac{k}{3} = 1 \rightarrow k = 6$$

So the final valid PDF is

$$f_X(x) = \begin{cases} 6x(1-x) & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

(b) $X < 0 \rightarrow F_X(x) = 0, x > 1 \rightarrow F_X(x) = 1$. For $x \in [0, 1]$ we have:

$$\begin{aligned} F_X(t) &= P(X \le t) \\ &= \int_0^t 6x(1-x)dx \\ &= 3x^2 - 2x^3\Big|_0^t \\ &= 3t^2 - 2t^3 \end{aligned}$$

$$F_X(t) = \begin{cases} 0 & t < 0 \\ 3t^2 - 2t^3 & 0 \le t \le 1 \\ 1 & t > 1 \end{cases}$$

(c) $\int_{.5}^1 6x(1-x)dx = 3x^2 - 2x^3\Big|_{.5}^1 = 0.50$

(d) $\mathbb{P}(0 \le X \le .75) = F_X(.75) - F_X(0) = 3(.75^2) - 2(.75^3) = 0.844$

(e) $\mathbb{E}(X) = \int_0^1 x f_X(x)dx = \int_0^1 6x^2(1-x)dx = 2x^3 - \frac{3}{2}x^4\Big|_0^1 = 0.50$

(f) We already have $\mathbb{E}(X)$, we just need $\mathbb{E}(X^2)$ and can use the short cut formula for variance.

$\mathbb{E}(X^2) = \int_0^1 x^2 f_X(x)dx = \int_0^1 6x^3(1-x)dx = \frac{6}{4}x^4 - \frac{6}{5}x^5\Big|_0^1 = 0.30$

So, $Var(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 = 0.30 - 0.50^2 = 0.05$

2. Suppose a continuous random variable X has the following probability density function:

$$f_X(x) = \begin{cases} \frac{3}{2}x^2 & -1 \le x \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

(a) Write down the cumulative distribution function (CDF) in functional form. (Make sure to cover all cases).

Answer: $X < -1 \to F_X(x) = 0, x > 1 \to F_X(x) = 1$. For $x \in [-1, 1]$ we have:

$$\begin{aligned} F_X(t) &= P(X \le t) \\ &= \int_{-1}^{t} \frac{3}{2}x^2 dx \\ &= \frac{x^3}{2}\Big|_{-1}^{t} \\ &= \frac{t^3}{2} + \frac{1}{2} \\ &= \frac{t^3+1}{2} \end{aligned}$$

$$F_X(t) = \begin{cases} 0 & t < -1 \\ \frac{t^3+1}{2} & -1 \le t \le 1 \\ 1 & t > 1 \end{cases}$$

(b) Use your CDF from part (a) to find $\mathbb{P}(|X - \frac{1}{2}| < \frac{1}{4})$. (Write an expression in terms of F_X first, then solve).

Answer:

$$\begin{aligned} \mathbb{P}\left(|X - \frac{1}{2}| < \frac{1}{4}\right) &= \mathbb{P}\left(\frac{-1}{4} < X - \frac{1}{2} < \frac{1}{4}\right) \\ &= \mathbb{P}\left(\frac{1}{4} \le X \le \frac{3}{4}\right) \\ &= F_X(.75) - F_X(.25) \\ &= \left(\frac{.75^3+1}{2}\right) - \left(\frac{.25^3+1}{2}\right) \\ &= .203 \end{aligned}$$

(c) Find the value t such that $\mathbb{P}(X \le t) = 0.80$.

Answer: $\mathbb{P}(X \le t) = 0.80 \Rightarrow \frac{t^3+1}{2} = 0.80 \Rightarrow t^3 = 0.60 \Rightarrow t = (0.6)^{\frac{1}{3}} \Rightarrow t = .843$

3. A supplier of kerosene has a 150-gallon tank that is filled at the beginning of each week. His weekly demand shows a relative frequency behavior that increases steadily up to 100 gallons and then levels off between 100 and 150 gallons. If X denotes weekly demand in hundreds of gallons, the relative frequency of demand can be modeled by

$$f_X(x) = \begin{cases} x & 0 \le x \le 1 \\ 1 & 1 < x \le 1.5 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find $F_X(x)$. (Remember to cover all cases)

(b) Find $\mathbb{P}(0.5 \leq X \leq 1.2)$.

(c) Find $\mathbb{E}(X)$.

Answer:

(a) $X < 0 \to F_X(x) = 0, x > 1.5 \to F_X(x) = 1$.

For $x \in [0,1]$ we have $\int_0^t x dx = \frac{x^2}{2}\Big|_0^t = \frac{t^2}{2}$

For $x \in [1,1.5]$ we have $\int_0^1 x dx + \int_1^t 1 dx = \frac{x^2}{2}\Big|_0^1 + x\Big|_1^t = t - \frac{1}{2}$

$$F_X(t) = \begin{cases} 0 & t < 0 \\ \frac{t^2}{2} & 0 \leq t \leq 1 \\ t - \frac{1}{2} & 1 < t \leq 1.5 \\ 1 & t > 1.5 \end{cases}$$

(b) $\mathbb{P}(0.5 \leq X \leq 1.2) = F_X(1.2) - F_X(.5) = (1.2 - \frac{1}{2}) - (.5^2/2) = 0.575$

(c) We have to be careful because the PDF has two parts:

$$\begin{aligned} \mathbb{E}(X) &= \int_0^{1.5} x f_X(x) dx \\ &= \int_0^1 x^2 dx + \int_1^{1.5} x dx \\ &= \frac{x^3}{3}\Big|_0^1 + \frac{x^2}{2}\Big|_1^{1.5} \\ &= \frac{23}{24} = .9583 \end{aligned}$$

4. You arrive at a bus stop at 10:00, knowing that the bus will arrive at some time uniformly distributed between 10:00 and 10:30. Let X = time you wait for the bus to arrive (in minutes). Thus, $X \sim \text{Unif}(0,30)$

(a) How many minutes do you expect to wait?

(b) What is the probability you that you have to wait longer than 10 minutes?

(c) If at 10:15 the bus has not yet arrived, what is the probability that you will have to wait at least an additional 10 minutes?

(d) Suppose you go grab a coffee right when you get to the stop at 10:00. What time do you need to be back at the stop, such that there is only a 10% chance you miss the bus?

Answer:

(a) $\mathbb{E}(X) = 15$ (minutes)

(b) $\mathbb{P}(X > 10) = \frac{20}{30} = \frac{2}{3}$

(c) $\mathbb{P}(X > 25 | X > 15) = \frac{\mathbb{P}(X>25)}{\mathbb{P}(X>15)} = \frac{5}{15} = \frac{1}{3}$

(d) We need to find x such that $\mathbb{P}(X \leq x) = .1 \Rightarrow \frac{x}{30} = .1 \Rightarrow x = 3$. Thus we need to be back at the bus stop by 10:03

5. A web page is accessed at an average of 20 times an hour. Assume that waiting time until the next hit has an exponential distribution.

(a) Determine the rate parameter λ of the distribution of the time until the first hit?

(b) What is the expected time between hits?

(c) What is the probability that the next hit is within 20 minutes?

(d) What is the distribution of the time until the second hit? (Give the name of the distribution and the value(s) of parameter(s).)

(e) Describe the distribution of the total waiting time for 5 hits? (Give the name of the distribution and the value(s) of parameter(s).)

(f) What is the expected total waiting time for 5 hits on the web page?

(g) What is the probability that there will be less than 5 hits in the first hour? (Hint: Consider Poisson distribution instead.)

Answer: Let X be the time until the next hit. It is the same as the time between hits.

(a) Let X be the time until the next hit. By the the description, the rate parameter (number of hits per hour) $\lambda = 20$. Alternatively, the expected waiting time between hits, $E[X] = 1/20 = 1/\lambda$ giving $\lambda = 20$.

(b) Since $X \sim Exp(20)$, we have $E[X] = 1/\lambda = 1/20 = .05$ (hours)

(c) Need $P(X \leq 20/60)$ where $X \sim Exp(20)$. Using the cdf for the exponential distribution $F_X(.333) = 1 - e^{-.333\times 20} = 1 - 0.00127 = 0.9987$

(d) The time until the second hit is $Y = X_1 + X_2$ where $X_1 \sim Exp(20)$ and $X_2 \sim Exp(20)$ and X_1 and X_2 are independent. Thus $Y \sim Gamma(2, 20)$.

(e) The waiting time for 5 hits is $W = \sum_{i=1}^{5} = X_i$; so $W \sim Gamma(5, 20)$.

(f) We need $E[W]$ which is given by $k/\lambda = 5/20 = 0.25$ hour, as one might expect!

(g) Let N be the number of hits in the first hour and assume $N \sim Poi(20 \times 1)$. The we need $P(N < 5) = P(N \leq 4) = 0.000017$.

6. In lecture, we talked about the memoryless property of the exponential distribution. Let $Y \sim \text{Exp}(\lambda)$. For $t > s$, we stated that

$$\mathbb{P}(Y \leq s + t | Y \geq s) = \mathbb{P}(Y \leq t)$$

Prove that this is true.

Answer:

$$\begin{aligned}
\mathbb{P}(Y \leq s + t | Y \geq s) &= \frac{\mathbb{P}(s < Y \leq s + t)}{\mathbb{P}(Y \geq s)} \\
&= \frac{F_Y(s+t) - F_Y(s)}{e^{-\lambda s}} \\
&= \frac{(1 - e^{-\lambda(s+t)}) - (1 - e^{-\lambda s})}{e^{-\lambda s}} \\
&= \frac{(e^{-\lambda s}) - (e^{-\lambda(s+t)})}{e^{-\lambda s}} \\
&= 1 - e^{-\lambda t} \\
&= \mathbb{P}(Y \leq t)
\end{aligned}$$

7. The amount of time a postal clerk spends with his customer can be modeled using an exponential distribution. On average, the clerk spends 5 minutes with a customer. Let $X =$ the amount of time (in minutes) a postal clerk spends with his customer.

(a) Give the distribution of X and the value for its parameter.

(b) Give the probability density function (PDF) and the cumulative distribution function (CDF) of X.

(c) What is the probability that the clerk spends less than 5 minutes with a customer?

(d) If the clerk hasn't finished assisting the customer in 2 minutes, what is the probability that he spends less than 5 minutes with the customer?

Answer:

(a) $E(X) = \frac{1}{\lambda} = 5$ min $\rightarrow \lambda = \frac{1}{5} = 0.2$
$X \sim Exp(0.2)$

(b)

$$f(x) = \begin{cases} 0 & \text{for } x \leq 0 \\ 0.2e^{-0.2x} & \text{for } x > 0 \end{cases}$$

$$F_X(t) = \begin{cases} 0 & \text{for } t \leq 0 \\ 1 - e^{-0.2t} & \text{for } x > 0 \end{cases}$$

(c) $P(X < 5) = F_X(5) = 1 - e^{-0.2(5)} = 0.6321$

(d) $P(X < 5|X > 2) = P(X < 3)$ (by the memoryless property of exponential distribution)

$$P(X < 5|X > 2) = P(X < 3) = F_X(3) = 1 - e^{-0.2(3)} = 0.4512$$

Or we can calculate the conditional probability $P(X < 5|X > 2)$ directly

$$\begin{aligned} P(X < 5|X > 2) &= \frac{P((X < 5) \cap (X > 2))}{P(X > 2)} \\ &= \frac{P(2 < X < 5)}{P(X > 2)} \\ &= \frac{F_X(5) - F_X(2)}{1 - F_X(2)} \\ &= \frac{[1 - e^{-0.2(5)}] - [1 - e^{-0.2(2)}]}{1 - [1 - e^{-0.2(2)}]} \\ &= 0.4512 \end{aligned}$$

8. An online company that typically averages four customers per hour is offering a prize for the 25th customer. Based on past experience, the company expects a 50% increase in customers due to the competition. Define a random variable T as:

$$T = \text{time till the 25th customer (in hours)}$$

(a) What distribution should we use to model T? Give the name and parameter values.
Answer: Since T is the time till a certain number of occurrences, we should use $T \sim \text{gamma}(\alpha = 25, \lambda = 6)$

(b) What is the expected time till the 25th customer?
Answer: $\mathbb{E}(T) = \frac{\alpha}{\lambda} = \frac{25}{6} = 4.167$ (hours)

(c) Suppose you are busy and can only enter the competition five hours after it opens. What is the probability that the competition will be over by the time you have a chance to join?
Answer: We want $\mathbb{P}(T < 5) = \mathbb{P}(T \leq 5) = \mathbb{P}(X \geq 25)$ where $X \sim \text{pois}(30)$

$\mathbb{P}(X \geq 25) = 1 - \mathbb{P}(X \leq 24) = .843$. (using the poisson cdf table)

9. For a Normal random variable X with $\mathbb{E}(X) = -3$ and $Var(X) = 4$, Compute

   (a) $\mathbb{P}(X \leq 2.39)$
   (b) $\mathbb{P}(\frac{X+3}{2} \leq -0.99)$
   (c) $\mathbb{P}(|X| \geq 2.39)$
   (d) $\mathbb{P}(|X+3| \geq 2.39)$
   (e) $\mathbb{P}(X < 5)$
   (f) $\mathbb{P}(|X| < 5)$
   (g) the value x such that $\mathbb{P}(X > x) = .33$

   Answer: We will convert everything to a $N(0,1)$ random variable and use the z-table

   (a) $\mathbb{P}(X \leq 2.39) = \mathbb{P}(Z \leq 2.70) = \Phi(2.70) = .997$
   (b) $\mathbb{P}(\frac{X+3}{2} \leq -0.99) = \mathbb{P}(X \leq -0.99) = \Phi(-0.99) = .161$
   (c) $\mathbb{P}(|X| \geq 2.39) = \mathbb{P}(X \leq -2.39) + \mathbb{P}(X \geq 2.39) = .623$
   (d) $\mathbb{P}(|X+3| \geq 2.39) = \mathbb{P}(|Z| \geq 1.20) = \mathbb{P}(Z \leq -1.20) + \mathbb{P}(Z \geq 1.20) = .23$
   (e) $\mathbb{P}(X < 5) = \mathbb{P}(Z < 4) \approx 1$
   (f) $\mathbb{P}(|X| < 5) = \mathbb{P}(-5 < X < 5) = \mathbb{P}(-1 < Z < 4) = \Phi(4) - \Phi(-1) = .841$
   (g) First we find the value z in the standard normal distribution such that $\mathbb{P}(Z \leq z) = .67$. That value is $z = .44$.

       Thus $\frac{x+3}{2} = .44 \Rightarrow x = 2(.44) - 3 = -2.12$

10. The price of a particular make of a 64GB iPad Mini among dealers nationwide is assumed to have a Normal distribution with mean $\mu = \$500$ and variance $\sigma^2 = 225$.

    (a) What is the probability that an iPad Mini of the same specs, chosen randomly from a dealer, will cost less than \$490?

    Answer:

$$\begin{aligned} P(X < 490) &= P\left(Z < \frac{490-500}{15}\right) \\ &= P(Z < -0.67) = \Phi(-0.67) = 0.2514 \end{aligned}$$

    (b) What is the probability that an iPad Mini of the same specs, chosen randomly from a dealer, will cost more than \$530?

    Answer:

$$\begin{aligned} P(X > 530) &= P\left(Z > \frac{530-500}{15}\right) \\ &= P(Z > 2.00) = 1 - \Phi(2.00) = 1 - 0.9772 = 0.0228. \end{aligned}$$

    (c) What is the probability that an iPad Mini of the same specs, chosen randomly from a dealer, will cost between \$490 and \$530?

    Answer:

$$\begin{aligned} P(490 < X < 530) &= P\left(\frac{490-500}{15} < Z < \frac{530-500}{15}\right) \\ &= P(-0.67 < Z < 2.00) \\ &= \Phi(2.00) - \Phi(-0.67) = 0.9772 - 0.2514 = 0.7258 \end{aligned}$$

(d) The manufacturer doesn't want dealers to markup these iPad Minis above the 90^{th} percentile of the price distribution. Approximately, what is the 90^{th} percentile of the distribution of the price of this iPad Mini?

Answer:

$$P(Z < z) = 0.90 \implies z \approx 1.28$$

Hence we have

$$\frac{X - 500}{15} = 1.28 \implies X = 500 + (1.28)(15) = 519.20$$

(e) What is the probability that the average price of 25 iPad Minis of the same specs, chosen randomly from a dealer, will be greater than \$510?

Answer: The average price of 25 computers of the same make chosen randomly from a dealer has mean $\mu = \$500$ and standard deviation $\sigma/\sqrt{n} = 15/\sqrt{25} = \$3$.

Hence the probability that the average price of 25 computers of the same make is greater than \$510 is

$$\begin{aligned} P(\bar{X} > 510) &= P\left(Z > \frac{510 - 500}{3}\right) \\ &= P(Z > 3.33) = 1 - \Phi(3.33) = 1 - 0.9995 = 0.0005 \end{aligned}$$

11. Suppose the installation time in hours for a software on a laptop has probability density function $f(x) = \frac{4}{3}(1 - x^3)$, $0 \leq x \leq 1$.

(a) Find the probability that the software takes between 0.3 and 0.5 hours to be installed on your laptop.

(b) Let $X_1, \ldots, X_{30}$ be the installation times of the software on 30 different laptops. Assume the installation times are independent. Find the probability that the *average* installation time is between 0.3 and 0.5 hours. Cite the theorem you use.

(c) Instead of taking a sample of 30 laptops as in the previous question, you take a sample of 60 laptops. Find the probability that the *average* installation time is between 0.3 and 0.5 hours. Cite the theorem you use.

Answer:

(a)

$$P(0.3 < X < 0.5) = \frac{4}{3}\int_0 .3^{0.5}(1 - x^3)dx = \frac{4}{3}\left(x - \frac{x^4}{4}\right)\Bigg|_{0.3}^{0.5} = 0.2485$$

(b) First $E(X) = \int_0^1 4x/3 \times (1 - x^3)dx = 0.4$. $Var(X) = E(X^2) - E(X)^2 = 2/9 - (2/5)^2 = 14/225$. By the central limit theorem,

$$\bar{X}_{30} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \equiv N\left(0.4, \frac{14/225}{30}\right) \equiv N(0.4, 7/3375)$$

$$\begin{aligned} P(0.3 < \bar{X}_{30} < 0.5) &= P\left(\frac{0.3 - 0.4}{\sqrt{7/3375}} < \frac{\bar{X}_{30} - 0.4}{\sqrt{7/3375}} < \frac{0.5 - 0.4}{\sqrt{7/3375}}\right) \\ &= P(-2.20 < Z < 2.20) \\ &= P(Z < 2.20) - P(Z < -2.20) \\ &= 0.9861 - 0.0139 \\ &= 0.9722 \end{aligned}$$

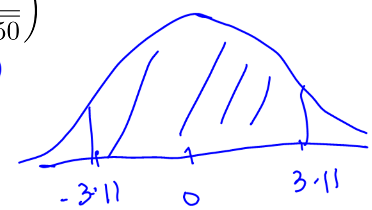
(c) $E(X) = 0.4$. $Var(X) = 14/225$. By the central limit theorem,

$$\bar{X}_{60} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \equiv N\left(0.4, \frac{14/225}{60}\right) \equiv N(0.4, 7/6750)$$

$$P(0.3 < \bar{X}_{\cancel{30}\,60} < 0.5) = P\left(\frac{0.3 - 0.4}{\sqrt{7/6750}} < \frac{\bar{X}_{30} - 0.4}{\sqrt{7/6750}} < \frac{0.5 - 0.4}{\sqrt{7/6750}}\right)$$

$\bar{X}_{60}$

$$\begin{aligned} &= \cancel{P(< Z <)}\ P(-3.11 < Z < 3.11) \\ &= P(Z < 3.11) - P(Z < -3.11) \\ &= 0.9991 - 0.0009 \\ &= 0.9982 \end{aligned}$$



12. Installation of some software package requires downloading 82 files. On average, it takes 15 seconds to download one file, with a variance of 16 sec^2. What is the (approximate) probability that the software is installed in less than than 20 minutes? (Use the Central Limit Theorem)

    Answer: Let S = sum of the 82 file download times. Then, from the Central Limit Theorem, we have $S \approx N(1200, 1312)$
    We want
    $$P(S < 1200) = P\left(Z \leq \frac{1200 - 1230}{\sqrt{1312}}\right) = \Phi(-.83) = 0.2033.$$

13. *Extra Credit (2 points)* Let X be a continuous random variable with CDF $F_X(t)$. Let $U \sim \text{unif}(0,1)$. Define a random variable Y as $Y = F_X^{-1}(U)$. Show that the CDF of Y is the same as the CDF of X, i.e. show $F_Y(t) = \mathbb{P}(Y \leq t) = \ldots = F_X(t)$. Fill in the dots to show this.

    Answer: $F_Y(t) = \mathbb{P}(Y \leq t) = \mathbb{P}(F_X^{-1}(U) \leq t) = \mathbb{P}(U \leq F_X(t)) = F_X(t)$ ■

    The key is that the unif(0,1) CDF has a unique property that F(t) = t. You get back whatever you plug in.