Solutions to Module 8 Practice Problems: Continuous Random Variables

1. A store manager wishes to study the relationship between the outside temperature $\widetilde{t}$ and the number of customers $\widetilde{c}$ visiting her store. Suppose the temperature can take three possible values: -10, 0 or 10. The number of customers visiting the store is either 100 or 200. The following table gives the joint probability mass function of t and c :

		c	
	$f(t,c)$	**100**	**200**
	−10	0	0.4
t	**0**	0.2	0.2
	10	0.1	0.1

(a) Derive the marginal probabilities of $\widetilde{t}$ and $\widetilde{c}$.

		c		
	$f(t,c)$	**100**	**200**	$f_t(t)$
	−10	0	0.4	0.4
t	**0**	0.2	0.2	0.4
	10	0.1	0.1	0.2
	$f_c(c)$	0.3	0.7	

(b) Derive $E(\widetilde{t})$, $E(\widetilde{t}^2)$ and $Var(\widetilde{t})$. Do the same for $\widetilde{c}$.

$E(\widetilde{t}) = -2$, $E(\widetilde{t}^2) = 60$, $Var(\widetilde{t}) = 56$

$E(\widetilde{c}) = 170$, $E(\widetilde{c}^2) = 31000$, $Var(\widetilde{c}) = 2100$.

(c) Derive $E(\widetilde{t}\widetilde{c})$. Then derive $Cov(\widetilde{t}, \widetilde{c})$ and $Corr(\widetilde{t}, \widetilde{c})$.

$E(\widetilde{t}\widetilde{c}) = -500$, $Cov(\widetilde{t}, \widetilde{c}) = -160$, $Corr(\widetilde{t}, \widetilde{c}) = -0.47$

(d) What can one say about the relationship between $\widetilde{t}$ and $\widetilde{c}$?

They are fairly strongly negatively related i.e. higher values of $\widetilde{t}$ are more likely to occur with lower values of $\widetilde{c}$, and vice-versa.

2. An analyst is interested in studying the relationship between the price of stocks A and B. Stock A can take three possible values: \$10, \$20 or \$30. Stock B can take three possible values: \$40, \$60 or \$80. The following table gives the joint probability mass function of the prices of A and B :

			b	
	$f(a,b)$	**40**	**60**	**80**
	10	0	0.2	0.1
a	**20**	0	0.2	0.2
	30	0.2	0.1	0

(a) Derive the marginal probabilities of the prices of A and B.

			b		
	$f(a,b)$	**40**	**60**	**80**	$f_a(a)$
	10	0	0.2	0.1	0.3
a	**20**	0	0.2	0.2	0.4
	30	0.2	0.1	0	0.3
	$f_b(b)$	0.2	0.5	0.3	1

(b) Derive $E(A)$, $E(A^2)$ and $Var(A)$. Do the same for B.

$E(A) = 20$, $E(A^2) = 460$, $Var(A) = 60$. $E(B) = 62$, $E(B^2) = 4040$, $Var(B) = 196$.

(c) Derive $E(AB)$. Then derive $Cov(A,B)$ and $Corr(A,B)$.

$E(AB) = 1180, Cov(A,B) = -60$, $Corr(A,B) = -0.55$

(d) What can one say about the relationship between the prices of the two stocks?

They are fairly strongly negatively related .

(e) What is $Cov(5 + 2A, 3 + 6B)$ and $Corr(5 + 2A, 3 + 6B)$?

$Cov(5 + 2A, 3 + 6B) = -720$, $Corr(5 + 2A, 3 + 6B) = -0.55$

(f) Can you think of two real stocks that you would expect to be positively correlated? What about two stocks which maybe negatively correlated?

3. Suppose X is a continuous random variable whose c.d.f. is given below:

$$
\begin{aligned}
F(x) &= 0 \quad \text{for } x < 0 \\
&= \frac{x^2}{100} \quad \text{for } 0 \leq x \leq 10 \\
&= 1 \quad \text{for } x > 10
\end{aligned}
$$

(a) What is the probability that (i) X is less than 5, (ii) X is between 7 and 9?

(i) $F(5) = \frac{1}{4} = 0.25$, (ii) $F(9) - F(7) = \frac{32}{100} = 0.32$

(b) Derive its p.d.f.

$$\begin{aligned} f(x) &= \frac{x}{50} \quad \text{for } x \in [0, 10] \\ &= 0 \quad \text{elsewhere} \end{aligned}$$

(c) Derive $E(X)$, $E(X^2)$ and the variance of X. [*This and many of the problems in this problem set require use of the following integration formula:* $\int x^n dx = \frac{x^{n+1}}{n+1}$]

$E(X) = \int_0^{10} \frac{x^2}{50} dx = 6.67$, $E(X^2) = \int_0^{10} \frac{x^3}{50} dx = 50$, $\text{Var}(X) = 5.51$

(d) [*Requires integration*] Derive $E(X^3)$.

$E(X^3) = \int_0^{10} \frac{x^4}{50} dx = 400$

(e) Derive the 75-th percentile of X, i.e. the particular number such that the probability of the random variable taking values lower than that number is 0.75.

$x = 8.66$

4. Suppose Y is a continuous random variable whose p.d.f. is given by:

$$\begin{aligned} f(y) &= cy \quad \text{for } y \in [10, 20] \\ &= 0 \quad \text{elsewhere} \end{aligned}$$

where c is a constant.

(a) [*Requires integration*] Show that its c.d.f. is given by:

$$\begin{aligned} F(y) &= 0 \quad \text{for } y < 10 \\ &= \frac{c(y^2 - 100)}{2} \quad \text{for } 10 \le y \le 20 \\ &= 1 \quad \text{for } y > 20 \end{aligned}$$

$F(y) = \int_{-\infty}^{y} cx dx = \int_{10}^{y} cx dx = c\frac{x^2}{2}|_{10}^{y} = \frac{c(y^2-100)}{2}$ for $y \in [0, 10]$

$F(y) = 0$ for $y < 0$, $F(y) = 1$ for $y > 20$,

(b) Show that c must $\frac{1}{150}$ for this to be a valid p.d.f.

$F(20) = 1 \Rightarrow \frac{c(400-100)}{2} = 1 \Rightarrow c = \frac{1}{150}$

(c) [*Requires integration*] Derive the mean and variance of Y.

$E(Y) = \int_{10}^{20} \frac{y^2}{150} dy = 15.556$, $E(Y^2) = \int_{10}^{20} \frac{y^3}{150} dy = 250$, $\text{Var}(Y) = 8.01$

(d) What is the probability that (i) Y is greater than 15, (ii) Y is between 18 and 20, (iii) Y is between 18 and 20, conditional on knowing that it is greater than 15?

(i) $1 - F(15) = 0.583$, (ii) $F(20) - F(18) = 0.253$, (iii) $\frac{P(18 \leq y \leq 20)}{P(15 \leq y)} = 0.434$

(e) [*Requires integration*] Derive $E(Y^\alpha)$, where α is a constant greater than 1.

$E(Y^\alpha) = \int_{10}^{20} \frac{y^{\alpha+1}}{150} dy = \frac{1}{150(\alpha+1)}[20^{\alpha+2} - 10^{\alpha+2}]$

5. [**Advanced**] Suppose Z is a continuous random variable whose p.d.f. is given below:

$$\begin{aligned} f(z) &= me^{-mz} \qquad \text{for } z \geq 0 \\ &= 0 \qquad \text{elsewhere} \end{aligned}$$

This is called the *exponential distribution* and is often used to model the amount of time till an event occurs. For example, the time till a battery dies or time till the next customer arrives.

(a) [*Requires the integration formula:* $\int e^{-mx} dx = -\frac{e^{-mx}}{m}$.] Show that its c.d.f. is given by:

$$\begin{aligned} F(z) &= 1 - e^{-mz} \qquad \text{for } z \geq 0 \\ &= 0 \qquad \text{for } z < 0 \end{aligned}$$

$F(z) = \int_0^z me^{-mx} dx = -e^{-mx}|_0^z = 1 - e^{-mz}$.

(b) Suppose $m = 1$. What is the probability that (i) Z is greater than 1, (ii) Z is greater than 2, (iii) Z is greater than 3. *Research if a command in Excel or Google Sheets can help you in doing these calculations.*

(i) $1 - F(1) = e^{-1} = 0.368$ (ii) $1 - F(2) = e^{-2} = 0.135$, (iii) $1 - F(3) = e^{-3} = 0.0498$

(c) What is the probability that Z is greater than 3, conditional on it being greater than 2?

$\frac{0.0498}{0.135} = 0.368$

Check that this probability is the same as the probability that Z is greater than 1 from (b)(i). This is called the memoryless property of the Exponential distribution i.e. the fact that it is greater than 2 does not impact the probability that it will increase by another 1. In other words, the fact that 2 customers arrived in the past has no impact on the time till the next one will arrive. This is written as $P(Z > k) = P(Z > t + k | Z > t)$.

6. X is a random variable which is uniformly distributed over the interval $[0, 10]$.

(a) Write the p.d.f. of X. What is its c.d.f.?

p.d.f.:

$$\begin{aligned} f(x) &= \frac{1}{10} \quad \text{for } x \in [0, 10] \\ &= 0 \quad \text{elsewhere} \end{aligned}$$

c.d.f.:

$$\begin{aligned} F(x) &= 0 \quad \text{for } x < 0 \\ &= \frac{x}{10} \quad \text{for } 0 \le x \le 10 \\ &= 1 \quad \text{for } x > 10 \end{aligned}$$

(b) What is E(X)? What is $E(10 + 2X)$?

5, 20

(c) What is the probability that (i) X is less than 5, (ii) X is between 7 and 9?

(i) $F(5) = \frac{1}{2}$, (ii) $F(9) - F(7) = \frac{2}{10} = \frac{1}{5}$

(d) [*Requires integration*] Derive $E(X^2)$ and the variance of X.

$E(X^2) = \int_0^{10} \frac{x^2}{10} dx = 33.33$, $\text{Var}(X) = 8.33$

7. Y is a random variable which is uniformly distributed over the interval $[20, 40]$.

(a) Write the p.d.f. of Y. What is its c.d.f.?

p.d.f.:

$$\begin{aligned} f(y) &= \frac{1}{20} \quad \text{for } x \in [20, 40] \\ &= 0 \quad \text{elsewhere} \end{aligned}$$

c.d.f.:

$$\begin{aligned} F(y) &= 0 \quad \text{for } y < 20 \\ &= \frac{y - 20}{20} \quad \text{for } 20 \le y \le 40 \\ &= 1 \quad \text{for } y > 40 \end{aligned}$$

(b) What is $E(Y)$? What is $E(100 + \frac{Y}{2})$?

$E(Y) = 30$, $E(100 + \frac{Y}{2}) = 115$

(c) Derive the following probabilities: (i) $P(Y > 25)$, (ii) $P(25 \le Y \le 35)$, (iii) $P(25 \le Y \le 35 \mid Y > 25)$.

(i) $1 - F(25) = \frac{3}{4}$, (ii) $F(35) - F(25) = \frac{1}{2}$, (iii) $\frac{2}{3}$

(d) [*Requires integration*] Derive $E(Y^2)$ and $E(5Y^2 + Y^3)$.

$E(Y^2) = \int_{20}^{40} \frac{y^2}{20} dy = 933.33,$

$E(Y^3) = \int_{20}^{40} \frac{y^3}{20} dy = 30000,\ E(5Y^2 + Y^3) = 34,666.65$

(e) What is the 75-th percentile of Y?

$\frac{y-20}{20} = 0.75 \Rightarrow y = 35$

8. The total duration D of baseball games in the MLB in the 2015 season was uniformly distributed between 120 and 260 minutes.

(a) Write the p.d.f. and the c.d.f. of D.

p.d.f.:

$$\begin{aligned} f(d) &= \frac{1}{140} \qquad \text{for } x \in [120, 260] \\ &= 0 \qquad \text{elsewhere} \end{aligned}$$

c.d.f.:

$$\begin{aligned} F(d) &= 0 \qquad \text{for } d < 120 \\ &= \frac{d-120}{140} \qquad \text{for } 120 \le y \le 260 \\ &= 1 \qquad \text{for } y > 260 \end{aligned}$$

(b) What was the mean duration of games in the MLB in the 2015 season?

190

(c) What was the probability that a random game ended within 3 hours?

$F(180) = \frac{60}{140} = 0.43$

(d) What was the probability that a random game ended between 3 and 3:30 hours?

$F(210) - F(180) = \frac{30}{140} = 0.21$

(e) Suppose you are watching a game and it crosses 3 hours. What is the probability that it will end within the next 30 minutes?

$P(210 \le d \le 180 \mid 180 \le d) = \frac{30/140}{80/140} = 0.375$

9. X is a random variable which is normally distributed with mean of 40 and variance of 25.

(a) What is the distribution of (i) $2X$, (ii) $2X + 10$?

(i) Normally distributed with mean = 80, variance = 100, (ii) Normally distributed with mean = 90, variance = 100.

(b) What is the probability that: (i) X is between 35 and 45, (ii) X is between 30 and 40, (iii) X is between 25 and 55?

(i) 0.68, (ii) 0.95 (iii) 0.997

(c) What is the probability that: (i) X is less than 45, (ii) X is between 30 and 45, (iii) X is greater than 55?

(i) 0.84, (ii) 0.815 (iii) 0.0015

(d) Use Excel/Google Sheets to calculate the probability that (i) X is less than 38, (ii) X is greater than 47, (iii) X is between 38 and 50.

(i) 0.345, (ii) 0.081 (iii) 0.63

10. Y is a random variable which is normally distributed with mean of 8 and variance of 16.

(a) What is the distribution of (i) $5Y$, (ii) $\frac{Y-8}{4}$?

(i) Normally distributed with mean = 40, variance = 400, (ii) Normally distributed with mean = 0, variance = 1.

(b) What is the probability that: (i) Y is between 0 and 16, (ii) $5Y$ is less than 20, (iii) $5Y$ is exactly equal to 4?

(i) 0.95, (ii) 0.16 (iii) 0

(c) What is $E(Y^2)$?

$E(Y^2) = \text{Var}(Y) + (EY)^2 = 80$

(d) Use Excel/Google Sheets to calculate the probability that (i) Y is less than 6, (ii) Y is greater than 10, (iii) Y is between 6 and 10.

(i) 0.31, (ii) 0.31 (iii) 0.38

11. GMAT is an exam that is used by many business schools as part of their admission criteria. Scores on the GMAT are approximately normally distributed with a mean of 570 and a standard deviation of 72.

(a) What is the range of scores that 95% of GMAT scores lie within?

Between 426 and 714.

(b) What is the range of scores that 99.7% of GMAT scores lie within?

Between 354 and 786.

(c) Many of the students admitted into the top business schools have GMAT scores of 700 or higher. What is the probability of an individual scoring 700 or higher on the GMAT? *Use Excel/Google Sheets to calculate your answers for this and the subsequent parts.*

0.035

(d) What is the minimum that an individual needs to score to be among the top 10% of GMAT exam-takers?

662

(e) Suppose an individual's score on the GMAT is 680. What is his/her ranking i.e. what fraction of test-takers would have scored higher than him/her?

Top 6%

(f) Suppose you know that your friend scored above 700 on the GMAT, but you don't know her exact score. What is the probability that her score was, in fact, higher than 750?

0.177

12. Suppose the height of trees in a forest is normally distributed with mean 60m and variance of 36.

(a) Within what interval does the height of 95% of the trees in the forest lie?

Between 48m and 72m.

(b) What is the probability that the height of a randomly chosen tree will be less than 70m? *Use Excel/Google Sheets to calculate your answers for this and the subsequent parts.*

0.952

(c) What is the probability that the height of a randomly chosen tree will be more than 75m?

0.006

(d) What height corresponds to the 80-th percentile?

65.05m

(e) What is the height that 75% of the trees in the forest are shorter than?

64.05m

13. Consider the same forest as in the previous question i.e. where the height of trees is normally distributed with mean 60m and variance of 36. Suppose a researcher sampled two random trees from this forest.

(a) What would be the distribution of their total height?

Normally distributed with mean = 120m, variance = 72

(b) What would be the distribution of their average height?

Normally distributed with mean = 60m, variance = 18

(c) Use Excel/Google Sheets to calculate the probability that the average height of the two randomly chosen trees will be more than 65m?

0.12

(d) Suppose instead the researcher sampled n trees from the forest. What would be the distribution of their average height? [*This concept will be important in the module on Statistical Inference later on in the course.*]

Normally distributed with mean = 60m, variance = $\frac{36}{n}$