Course: Machine Learning - Foundations
Week 11 Questions

GRADED QUESTIONS

1. (1 point) The number of hours Messi spends each day practicing in ground is modelled by the continuous random variable X, with p.d.f. $f(x)$ defined by

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

Find the probability that Messi will practice between 2 and 5 hours in ground on a randomly selected day.

Answer: 0.80

We know that $\int_{-\infty}^{\infty} f(x)dx = 1$

Solving above equation taking required $f(x)$, value of a can be calculated. i.e $a = \dfrac{6}{125}$

Then calculate $P(2 \le X \le 5) = \int_{2}^{5} f(x)dx$

2. (1 point) Let X be a continuous random variable with PDF

$$f_X(x) = \begin{cases} ax & \text{for } 0 < x < 2 \\ a(4-x) & \text{for } 2 \le x \le 4 \\ 0 & \text{otherwise} \end{cases}$$

Calculate $P(1 \le x \le 3)$

Answer: 0.75

We know that $\int_{-\infty}^{\infty} f(x)dx = 1$

Solving above equation taking required $f(x)$, value of a can be calculated. i.e $a = \dfrac{1}{4}$

Then calculate $P(1 \le X \le 2) = \int_{1}^{2} f(x)dx$

3. (1 point) The probability density function of X is given by

$$f_X(x) = \begin{cases} x & \text{for } 0 < x < 1 \\ 2-x & \text{for } 1 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

Calculate $E(X)$

Answer: 1

Solution:
We know that $\int_{-\infty}^{\infty} f(x)dx = 1$

$E(X) = \int_0^2 xf(x)dx$

4. (1 point) The distribution of the lengths of a cricket bat is uniform between 80 cm and 100 cm. There is no cricket bat outside this range. The mean and variance of the lengths of the the cricket ball is a and b. Calculate $a + b$

Answer: 127.33

$V(X) = \dfrac{(h-l)^2}{12}$

$E(X) = \dfrac{(h+l)}{2}$

5. (points) Suppose that random variable X is uniformly distributed between 0 and 10. Then find $P(X + \dfrac{10}{X} \geq 7)$. (Write answer upto two decimal places)

Answer: 0.7

Solve this quadratic equation, $X + \dfrac{10}{X} \geq 7$
get the values of X for this $X \geq 0$.

$X \in [0, 2] \cup [5, 10]$

So the total area equals $0.2 + 0.5 = 0.7$

6. (1 point) **(Multiple Select)** Which of the following option is/are correct?
   A. For a standard normal variate, the value of Standard Deviation is 1.
   B. Normal Distribution is also known as Gaussian distribution.
   C. In Normal distribution, the highest value of ordinate occurs at mean.
   D. The shape of the normal curve depends on its standard deviation.

Answer: A, B, C, D

Option A: Standard normal variate(distributions) have a mean of 0 and variance of 1. SD is the squared root of variance.

Option B: Normal distribution is indeed known as the Gaussian distribution.

Option C: The normal distribution resembles a bell curve. The maximum concentration is around the mean value/middle portion.

Option D: The spread of the distribution will change based on the SD. Hence, the shape is dependent on the standard deviation.

Let X and Y be continuous random variables with joint density

$$f_{XY}(x,y) \begin{cases} cxy & \text{for } 0 < x < 2,\, 1 < y < 3 \\ 0 & \text{otherwise} \end{cases}$$

From the above information answer questions from 7-13

7. (points) Calculate the value of c

Answer: $\frac{1}{8}$

Solution:
We know that $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)dydx = 1$

$\int_0^2\int_1^3 cxydydx = 1$
c can be calculated from above equation.

8. (points) Calculate $P(0 < X < 1, 1 < Y < 2)$

Answer: $\frac{3}{32}$

Solution:

$P(0 < X < 1, 1 < Y < 2) = \int_0^1\int_1^2 \frac{1}{8}xydydx$

9. (points) Calculate $P(0 < X < 1, Y > 2)$

Answer: $\frac{5}{32}$

Solution:

$P(0 < X < 1, Y > 2) = \int_0^1 \int_2^3 \frac{1}{8} xydydx$

10. (points) Calculate $P((X + Y) < 3)$

Answer: 0.25

Solution:

$= \int_0^2 \int_1^{3-x} \frac{1}{8} xydydx$

11. (points) Calculate $F_X(1)$

Answer: 0.25

Solution:

Here, $F_X(x)$ means marginal cumulative distribution i.e $P(X \leq x)$

It can be calculated as follows i.e to integrate x from 0 to $x = 1$ and for all values of y.

$F_X(x) = \int_1^3 \int_0^x \frac{1}{8} xydydx$

12. (points) Calculate $F_Y(2)$

Answer: $\frac{3}{8}$

Solution:

Here, $F_Y(y)$ means marginal cumulative distribution i.e $P(Y \leq y)$

It can be calculated as follows i.e to integrate y from 1 to $y = 2$ and for all values of x.

$F_X(x) = \int_0^2 \int_1^y cxydydx$

13. (points) Calculate $F_{X,Y}(1, 4)$

Answer: 0.25

$F_{X,Y} = F_X(x) \times F_Y(y)$

14. (1 point) Suppose a random variable X is best described by a uniform probability distribution with range 1 to 5. Find the value of a such that $P(X \leq a) = 0.5$

Answer: 3

Solution: $P(X \leq 3) = 0.5$, From the area of Uniform distribution curve.

15. (1 point) If X is an exponential random variable with rate parameter λ then which of the following statement(s) is(are) correct.

a) $E[X] = \frac{1}{\lambda}$

b) $Var[X] = \frac{1}{\lambda^2}$

c) $P(X > x + k | X > k) = P(X > x)$ for $k, x \geq 0$.

d) $P(X > x + k | X > k) = P(X > k)$ for $k, x \geq 0$.

Answer: A, B, C

Solution :
Options (a) and (b) are correct for Exponential distribution.

$$P(X > x + k | X > k) = \frac{P((X > x + k) \cap (X > k)}{P(X > k)}$$

$$\Rightarrow \frac{P(X > x + k)}{P(X > k)} = \frac{e^{-\lambda \times (x+k)}}{e^{-\lambda k}}$$

$$\Rightarrow e^{-\lambda x}$$

Hence, Option C is also correct and option (d) is incorrect.

16. (1 point) **(Multiple Select)** For three events, A, B, and C, with $P(C) > 0$, Which of the following is/are correct?

A. $P(A^c|C)$= 1 - $P(A|C)$

B. $P(\phi|C) = 0$

C. $P(A|C) \leq 1$

D. if $A \subset B$ then $P(A|C) \leq P(B|C)$

Answer: A, B, C, D

Option A: Using standard probability properties. If we have an event E, then:

$$P(E) = 1 - P(E^C)$$

Option B: The option asks for the probability of getting a null set given that an event C has already occurred. It is also given to us that the probability of the occurrence of the event C is not zero. Hence, $P(\phi|C) = 0$

Option C: The probability of an event given to another with non-zero probability will always be less than or equal to 1 because the total probability can only be 1.

Option D: The probability of getting a bigger set is more than a smaller set. A is a smaller set than B, hence $P(A|C) \leq P(B|C)$

17. (2 points) **(Multiple Select)** Let the random experiment be tossing an unbiased coin two times. Let A be the event that the first toss results in a head, B be the event that the second toss results in a tail and C be the event that on both the tosses, the coin landed on the same side. Choose the correct statements from the following:

    A. A and C are independent events.

    B. A and B are independent events.

    C. B and C are independent events.

    D. A, B, and C are independent events.

Answer: A, B, C

Solution:

$A = \{HT, HH\}$

$B = \{HT, TT\}$

$C = \{TT, HH\}$

$P(A) = \dfrac{1}{2}$

$P(B) = \dfrac{1}{2}$

$P(C) = \dfrac{1}{2}$

$P(A \cap B) = \{HT\}$

$P(C \cap B) = \{TT\}$

$P(A \cap C) = \{HH\}$

$P(A \cap B) = P(A) \times P(B)$ Hence, option B is correct

$P(A \cap C) = P(A) \times P(C)$ Hence, option A is correct

$P(C \cap B) = P(C) \times P(B)$ Hence, option C is correct

18. (2 points) **(Multiple Select)** If $A_1, A_2, A_3, , A_n$ are non empty disjoint sets and subsets of sample space S, and a set A_{n+1} is also a subset of S, then which of the following statements are true?

A. The sets $A_1 \cap A_{n+1}, A_2 \cap A_{n+1}, A_3 \cap A_{n+1}, , A_n \cap A_{n+1}$ are disjoint.

B. If A_{n+1}, A_n are disjoint then $A_1, A_2, , A_{n-1}$ are disjoint with A_{n+1}.

C. The sets $A_1, A_2, A_3, , A_n, \phi$ are disjoint.

D. The sets $A_1, A_2, A_3, , A_n, S$ are disjoint.

Answer: A,C

Option A: Consider two sets $A_i \cap A_{n+1}$ and $A_j \cap A_{n+1}$, where $i \neq j$. The intersection of these two sets is

$$\begin{aligned}(A_i \cap A_{n+1}) \cap (A_j \cap A_{n+1}) &= (A_i \cap A_j) \cap A_{n+1} \\ &= \phi \cap A_{n+1} \quad \text{(since } A_i, A_j \text{ are disjoint sets)} \\ &= \phi\end{aligned}$$

Hence, $A_i \cap A_{n+1}$ and $A_j \cap A_{n+1}$ are disjoint sets for all $i \neq j$. Therefore, the sets $A_1 \cap A_{n+1}, A_2 \cap A_{n+1}, A_3 \cap A_{n+1}, , A_n \cap A_{n+1}$ are disjoint.

Option B: Again, if A_{n+1} is disjoint with A_n, it doesn't mean that it'll be disjoint with other n-1 sets as well. Take this example:

you have 3 sets, A_1, A_2, A_3, you can have $A_4 = A_1 + A_2$ Now, A_4 is disjoint with A_3 but not with the other two.

Option C: $A_1, ...A_n$ are disjoint with each other(given). Every n set will also be disjoint with ϕ (Intersection will give empty set)

Option D: Every n set is a subset of S. So, it'll result in something when taken an intersection with S. Hence, not disjoint.

19. (3 points) A triangular spinner having three outcomes can lands on one of the numbers 0, 1 and 2 with probabilities shown in table.

Outcome	0	1	2
Probability	0.7	0.2	0.1

Table 1: Table 10.2: Probability distribution

The spinner is spun twice. The total of the numbers on which it lands is denoted by X. The the probability distribution of X is.

A.

x	2	3	4	5	6
$P(X = x)$	$\frac{49}{100}$	$\frac{28}{100}$	$\frac{1}{100}$	$\frac{4}{100}$	$\frac{18}{100}$

B.

x	2	3	4	5	6
$P(X = x)$	$\frac{28}{100}$	$\frac{49}{100}$	$\frac{18}{100}$	$\frac{1}{100}$	$\frac{4}{100}$

C.

x	0	1	2	3	4
$P(X=x)$	$\frac{49}{100}$	$\frac{28}{100}$	$\frac{18}{100}$	$\frac{4}{100}$	$\frac{1}{100}$

D.

x	2	3	4	5	6
$P(X=x)$	$\frac{28}{100}$	$\frac{49}{100}$	$\frac{18}{100}$	$\frac{4}{100}$	$\frac{1}{100}$

Answer: C

The maximum sum you can get is 4 (2+2).

$$P(X=4) = P(2 \text{ and } 2) = P(2) * P(2) = 0.1 * 0.1 = \frac{1}{100}$$

$$P(X=3) = P(2 \text{ and } 1) \text{ or } P(1 \text{ and } 2) = (P(2) * P(1)) + (P(1) * P(2))$$
$$\implies P(X=3) = (0.1 * 0.2) + (0.2 * 0.1) = \frac{4}{100}$$

Similarly, We'll have the other probabilities like:

$$P(X=2) = \frac{18}{100}$$
$$P(X=1) = P(X=2) = \frac{28}{100}$$
$$P(X=0) = \frac{49}{100}$$

20. (1 point) When throwing a fair die, what is the variance of the number of throws needed to get a 1?

Answer: 30

Solution:

$$= Var(X) = \frac{1-p}{p^2}$$

$$= \frac{1-\frac{1}{6}}{\frac{1}{6}^2}$$

$$= 30$$

21. (1 point) Joint pmf of two random variables X and Y are given in Table

$x \backslash y$	1	2	3	$f_X(x)$
1	0.05	0	a_1	0.15
2	0.1	0.2	a_3	a_2
3	a_4	0.2	a_5	0.45
$f_Y(y)$	0.3	0.4	a_6	

Find the value of $f_{Y|X=3}(1)$ i.e ($P(Y = 1|X = 3)$)

Answer: 0.22

Solution:

$\sum f_{XY}(x, y) = 1$ .... (i)

$f_X(x) = \sum\limits_{y \in R_y} f_{XY}(x, y)$ ....(ii)

$f_Y(y) = \sum\limits_{x \in R_X} f_{XY}(x, y)$ ....(iii)

Hence, $a_1 = 0.10$, $a_2 = 0.40$, $a_3 = 0.1$, $a_4 = 0.15$, $a_5 = 0.1$, $a_6 = 0.3$

$$f_{Y|X=3}(1) = \frac{f_{XY}(1,3)}{f_X(3)} = \frac{0.1}{0.45} = 0.22$$

22. (1 point) **(Multiple Select)** Which of the following options is/are correct?

A. If $Cov[X, Y] = 0$, then X and Y are independent random variables.

B. $Cov[X, X] = Var(X)$

C. If X and Y are two independent random variables and $Z = X + Y$ then $f_Z(z) = \sum_x f_X(x) \times f_Y(z - x)$

D. If X and Y are two independent random variables and $Z = X + Y$ then $f_Z(z) = \sum_y f_X(x) \times f_Y(z - x)$

Answer: B, C

Solution:

Option B

$Cov[X, X]$ is the covariance between X and X i.e $Var(X)$

Option C is correct from its definition.

23. (1 point) **(Multiple Select)**A discrete random variables X has the cumulative distribution function is defined as follows.

$$F_X(x) = \begin{cases} \dfrac{x^3+k}{40}, & \text{for } x = 1, 2, 3 \end{cases}$$

Which of the following options is/are correct for $F(x)$ as given?

A. $k = 17$

B. $Var(X) = \dfrac{259}{320}$

C. $k = 13$

D. $Var(X) = \dfrac{249}{310}$

Answer: B, C

Solution:

For k

$F_X(3) = 1$

$$\frac{x^3+k}{40} = 1$$

Solving above equation to get $k = 13$

To calculate the variance, first calculate the probability distribution of X

We will get

$$P(X = 1) = \frac{14}{40}$$

$$P(X = 2) = \frac{7}{40}$$

$$P(X = 3) = \frac{19}{40}$$

Now easily with $Var(X)$ equation we will get $Var(X) = \frac{259}{320}$

24. (1 point) In a game of Ludo, Player A needs to repeatedly throw an unbiased die till he gets a 6. What is the probability that he needs fewer than 4 throws? (Answer the question correct to two decimal points.)

Solution:

$P(6) = \frac{1}{6}$

As it resembles geometric distribution. Hence,
$\sum_{n=1}^{3} \frac{1}{6} \times (1 - \frac{1}{6})^{n-1} = 0.6$

25. (1 point) **(Multiple Select)** Let X and Y be two random variables with joint PMF $f_{XY}(x, y)$ given in Table 10.3.

$x \backslash y$	0	1	2
0	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{8}$
1	$\frac{1}{8}$	$\frac{1}{6}$	$\frac{1}{6}$

Table 10.3: Joint PMF of X and Y.

Which of the following options is/are correct for $f_{XY}(x, y)$ given in Table 10.1.

A. $P(X = 0, Y \leq 1) = \frac{5}{12}$

B. $P(X = 0, Y \leq 1) = \frac{7}{12}$

C. X and Y are independent.

D. X and Y are dependent.

Answer: A, D

$P(X = 0, Y \leq 1) = P(X = 0 \text{ and } Y = 0)$ or $P(X = 0 \text{ and } Y = 1) = \frac{1}{6} + \frac{1}{4} = \frac{10}{24} = \frac{5}{12}$

$$P(X = 0) = \frac{1}{6} + \frac{1}{4} + \frac{1}{8} = \frac{13}{24}$$

$$P(Y = 0) = \frac{1}{6} + \frac{1}{8} = \frac{7}{24}$$

Now, if they are independent, then the product of the marginal should be equal to the joint probability.

$$P_{XY}(0,0) = \frac{13}{24} * \frac{7}{24} = \frac{91}{576} \approx 0.158 \text{ (If independent)}$$

Also, the tables says that $P_{XY}(0,0) = \frac{1}{6} = 0.1667$

Because the two are not equal, we can conclude that they are not independent.

26. (1 point) A discrete random variables X has the probability function as given in table 10.4.

x	1	2	3	4	5	6
$P(X)$	a	a	a	b	b	0.3

Table 2: Table 10.4: Probability distribution

If $E(X) = 4.2$, then evaluate $a + b$

Answer: 0.3

$\sum P(X = x) = 1$

$3a + 2b = 0.7$

$E(X) = \sum P(X = x_i) \times x_i$

$6a + 9b = 2.4$

Solving both equations, we get $a = 0.1$ and $b = 0.2$

27. (1 point) A discrete random variable X has the probability function as follows.

$$P(X = x) = \begin{cases} k \times (1-x)^2, & \text{for } x = 1, 2, 3 \\ 0, & \text{otherwise} \end{cases}$$

Evaluate $E(X)$

Answer: 2.8

Solution:

$\sum P(X = x) = 1$

$k + 4k = 1$

$k = 0.2$

$E(X) = \sum P(X = x_i) \times x_i$

$0.2 \times 2 + 0.8 \times 3$

$0.4 + 2.4 = 2.8$