EP2200 Home assignment I – Probability theory

December 28, 2015

1. Let A and B be two events such that:

$$P(A) = 0.4, P(B) = 0.7, P(A \cup B) = 0.9$$

(a) Find $P(A \cap B)$.
(b) Find $P(A^c \cap B)$.
(c) Find $P(A - B)$.
(d) Find $P(A^c - B)$.
(e) Find $P(A^c \cup B)$.
(f) Find $P(A \cap (B \cup A^c))$.

Solution:

We have

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Thus,

$$0.9 = 0.7 + 0.4 - P(A \cap B)$$

which results in:

$$P(A \cap B) = 0.2.$$

(b)

$$\begin{aligned} P(A^c \cap B) &= P(B - A) \\ &= P(B) - P(B \cap A) \\ &= 0.7 - 0.2 \\ &= 0.5 \end{aligned}$$

(c)

$$P(A - B) = 0.2$$

(d)

$$\begin{aligned} P(A^c - B) &= P(A^c \cap B^c) \\ &= P((A \cup B)^c) \\ &= 1 - P(A \cup B) \\ &= 1 - 0.9 \\ &= 0.1 \end{aligned}$$

(e)

$$\begin{aligned} P(A^c \cup B) &= P(A^c) + P(B) - P(A^c \cap B) \\ &= 1 - P(A) + P(B) - P(B - A) \\ &= 1 - 0.4 + 0.7 - 0.5 \\ &= 0.8 \end{aligned}$$

(f)

$$\begin{aligned} P(A \cap (B \cup A^c)) &= P((A \cap B) \cup (A \cap A^c)) \quad \text{(distributive law)} \\ &= P((A \cap B) \cup \varnothing) \quad (\text{since} \quad A \cap A^c = \varnothing) \\ &= P(A \cap B) \\ &= 0.2 \end{aligned}$$

2. Suppose that of all the customers at a coffee shop:

   -70% purchase a cup of coffee.
   -40% purchase a piece of cake.
   -20% purchase both a cup of coffee and a piece of cake.
   Given that a randomly chosen customer has purchased a piece of cake, what is the probability that he/she has also purchased a cup of coffee?

   Solution:

   We know

$$\begin{aligned} P(A) &= 0.7 \\ P(B) &= 0.4 \\ P(A \cap B) &= 0.2 \end{aligned}$$

   Therefore:

$$\begin{aligned} P(A|B) &= \frac{P(A \cap B)}{P(B)} \\ &= \frac{0.2}{0.4} \\ &= \frac{1}{2} \end{aligned}$$

3. One way to design a spam filter is to look at the words in an email. In particular, some words are more frequent in spam emails. Suppose that we have the following information:

   (a) 50% of emails are spam.
   (b) 1% of spam emails contain the word “refinance”.
   (c) 0.001% of non-spam emails contain the word “refinance”.

   Suppose that an email is checked and found out to contain the word refinance. What is the probability that the email is a spam?

Solution:

Let S be the event that an email is a spam and let R be the event that the email contains the word "refinance". Then,

$$\begin{aligned} P(S) &= \frac{1}{2} \\ P(R|S) &= \frac{1}{100} \\ P(R|S^c) &= \frac{1}{100000} \end{aligned}$$

Then,

$$\begin{aligned} P(S|R) &= \frac{P(R|S)P(S)}{P(R)} \\ &= \frac{P(R|S)P(S)}{P(R|S)P(S) + P(R|S^c)P(S^c)} \\ &= \frac{\frac{1}{100} \times \frac{1}{2}}{\frac{1}{100} \times \frac{1}{2} + \frac{1}{100000} \times \frac{1}{2}} \\ &\approx 0.999 \end{aligned}$$

4. Let X and Y be two independent discrete random variables with the following PMFs:

$$P_X(k) = \begin{cases} \frac{1}{4} & \text{for } k = 1 \\ \frac{1}{8} & \text{for } k = 2 \\ \frac{1}{8} & \text{for } k = 3 \\ \frac{1}{2} & \text{for } k = 4 \\ 0 & \text{otherwise} \end{cases}$$

and

$$P_Y(k) = \begin{cases} \frac{1}{6} & \text{for } k = 1 \\ \frac{1}{6} & \text{for } k = 2 \\ \frac{1}{3} & \text{for } k = 3 \\ \frac{1}{3} & \text{for } k = 4 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find $P(X \leq 2 \text{ and } Y \leq 2)$.

(b) Find $P(X > 2 \text{ or } Y > 2)$.

(c) Find $P(X > 2|Y > 2)$.

(d) Find $P(X < Y)$.

Solution:

(a) X and Y are two independent random variables. So:

$$\begin{aligned}
P(X \leq 2 \text{ and } Y \leq 2) &= P(X \leq 2) \cdot P(Y \leq 2) \\
&= \big(P_X(1) + P_X(2)\big) \cdot \big(P_Y(1) + P_Y(2)\big) \\
&= (\frac{1}{4} + \frac{1}{8})(\frac{1}{6} + \frac{1}{6}) \\
&= \frac{3}{8} \cdot \frac{1}{3} = \frac{1}{8}
\end{aligned}$$

(b) Using $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ and the fact that X and Y are two independent random variables:

$$\begin{aligned}
P(X > 2 \text{ or } Y > 2) &= P(X > 2) + P(Y > 2) - P(X > 2 \text{ and } Y > 2) \\
&= (\frac{1}{8} + \frac{1}{2}) + (\frac{1}{3} + \frac{1}{3}) - (\frac{1}{8} + \frac{1}{2}) \cdot (\frac{1}{3} + \frac{1}{3}) \\
&= \frac{5}{8} + \frac{2}{3} - \frac{5}{8} \cdot \frac{2}{3} = \frac{21}{24} = \frac{7}{8}
\end{aligned}$$

(c) Since X and Y are two independent random variables:

$$\begin{aligned}
P(X > 2 | Y > 2) &= P(X > 2) \\
&= \frac{1}{8} + \frac{1}{2} = \frac{5}{8}
\end{aligned}$$

(d) Using the law of total probability:

$$\begin{aligned}
P(X < Y) &= \sum_{k=1}^{4} P(X < Y | Y = k) \cdot P(Y = k) \\
&= P(X < 1 | Y = 1) \cdot P(Y = 1) + P(X < 2 | Y = 2) \cdot P(Y = 2) \\
&+ P(X < 3 | Y = 3) \cdot P(Y = 3) + P(X < 4 | Y = 4) \cdot P(Y = 4) \\
&= 0 \cdot \frac{1}{6} + \frac{1}{4} \cdot \frac{1}{6} + (\frac{1}{4} + \frac{1}{8}) \cdot \frac{1}{3} + (\frac{1}{4} + \frac{1}{8} + \frac{1}{8}) \cdot \frac{1}{3} \\
&= \frac{1}{24} + \frac{1}{8} + \frac{1}{6} = \frac{8}{24} = \frac{1}{3}
\end{aligned}$$

5. Suppose that $Y = -2X + 3$. If we know $E[Y] = 1$ and $E[Y^2] = 9$, find $E[X]$ and $\text{Var}(X)$.

Solution:

$$Y = -2X + 3$$

$$E[Y] = -2E[X] + 3 \quad \text{linearity of expectation}$$

$$1 = -2E[X] + 3 \quad \rightarrow \qquad E[X] = 1$$

$$var(Y) = 4 \times var(X) = E[Y^2] - E[Y]^2 = 9 - 1 = 8$$
$$\rightarrow \quad var(X) = 2$$

6. Let X be a continuous random variable with the following PDF

$$f_X(x) = \begin{cases} ce^{-4x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

where c is a positive constant.

(a) Find c.
(b) Find the CDF of X, $F_X(x)$.
(c) Find $P(2 < X < 5)$.
(d) Find $E[X]$.

Solution:

(a)

To find c

$$1 = \int_{-\infty}^{\infty} f_X(u)du = \int_0^{\infty} ce^{-4u}du$$
$$= \frac{c}{4}\Big[-e^{-4x}\Big]_0^{\infty} = \frac{c}{4}$$

Thus, we must have $c = 4$.

(b)

To find the CDF of X, we use $F_X(x) = \int_{-\infty}^{x} f_X(u)du$, so for $x < 0$, we obtain $F_X(x) = 0$. For $x \geq 0$, we have

$$F_X(x) = \int_0^x 4e^{-4u}du = -\Big[e^{-4x}\Big]_0^x = 1 - e^{-4x}.$$

Thus,

$$F_X(x) = \begin{cases} 1 - e^{-4x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

(c)

We can find $P(2 < X < 5)$ using either the CDF or the PDF. If we use the CDF, we have

$$P(2 < X < 5) = F_X(5) - F_X(2) = \left[1 - e^{-20}\right] - \left[1 - e^{-8}\right] = e^{-8} - e^{-20}.$$

Equivalently, we can use the PDF. We have

$$P(2 < X < 5) = \int_2^5 f_X(t)dt =$$

$$\int_2^5 4e^{-4t}dt = e^{-8} - e^{-20}.$$

(d)

As we saw, the PDF of X is given by

$$f_X(x) = \begin{cases} 4e^{-4x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

so to find its expected value, we can write:

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} x f_X(x)dx \\ &= \int_0^{\infty} x(4e^{-4x})dx \\ &= \Big[-xe^{-4x}\Big]_0^{\infty} + \int_0^{\infty} e^{-4x}dx \\ &= 0 + \Big[-\frac{1}{4}e^{-4x}\Big]_0^{\infty} = \frac{1}{4} \end{aligned}$$

7. Prove the following useful properties of random variables X and Y:

$$E[cX] = cE[X]$$

$$E[X + Y] = E[X] + E[Y]$$

$$\mathrm{Var}[cX] = c^2\mathrm{Var}[X]$$

If X and Y are independent, then

$$\mathrm{Var}[X + Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$$

Consider the mixture distribution given by pdf $p(x) = a_1p_1(x) + a_2p_2(x)$, $a_1 + a_2 = 1$. Show that

$$E[X] = a_1 E_{p_1}[X] + a_2 E_{p_2}[X]$$

$$E[X^2] = a_1 E_{p_1}[X^2] + a_2 E_{p_2}[X^2]$$

Solution:

We give solutions for discrete random variables. Proofs for continuous random variables are similar, with integrals instead of sums. Also note, that instead of two random variables, similar proofs are possible for sums of arbitrary number of random variables, and for mixture distributions with more terms.

$$E[cX] = \sum_x p_x cx = c \sum_x p_x x = cE[X]$$

$$\begin{aligned}
E[X+Y] &= \sum_{x,y} (x+y) p_{x,y} \\
&= \sum_{x,y} x p_{x,y} + \sum_{x,y} y p_{x,y} \\
&= \sum_x x \sum_y p_{x,y} + \sum_y y \sum_x p_{x,y} \\
&= \sum_x x p_x + \sum_y y p_y = E[X] + E[Y].
\end{aligned}$$

$$\begin{aligned}
\mathrm{Var}[cX] &= E[(cX^2)] - E[cX]^2 \\
&= c^2 E[X^2] - c^2 E[X]^2 = c^2 \mathrm{Var}[X]
\end{aligned}$$

If X and Y are independent, then

$$\begin{aligned}
\mathrm{Var}[X+Y] &= E[((X+Y) - E[X+Y])^2] \\
&= E[(X - E[X] + Y - E[Y])^2] \\
&= E[(X - E[X])^2 + (Y - E[Y])^2 + 2(X - E[X])(Y - E[Y])] \\
&= E[(X - E[X])^2] + E[(Y - E[Y])^2] + 2E[(X - E[X])(Y - E[Y])] \\
&= \mathrm{Var}[X] + \mathrm{Var}[Y],
\end{aligned}$$

where the last step holds only if X and Y are independent.

Finally, consider the mixture distribution given by pdf $p(x) = a_1 p_1(x) + a_2 p_2(x)$, $a_1 + a_2 = 1$.

$$
\begin{aligned}
E[X] &= \sum_x (a_1 p_1(x) + a_2 p_2(x))x \\
&= \sum_x a_1 p_1(x)x + \sum_x a_2 p_2(x))x \\
&= a_1 E_{p_1}[X] + a_2 E_{p_2}[X]
\end{aligned}
$$

The proof for the second moments is similar.