Probability Theory

Tutorial 7

1. Maria has three baby children: A (Ana), B (Ben), and C (Carl). She puts them to sleep at night. Because she sings beautifully, all three fall asleep as soon as she finishes the song. The number of consecutive hours each baby sleeps is a discrete random variable that takes the values 6 and 8 with probability 1/2 each. We assume that these three random variables are independent of each other. Find the distribution of the number of consecutive hours the mother sleeps, if she falls asleep at the same moment as her three children (she is very tired) and wakes up when the first of her three children wakes up.

2. A fair six-sided die is rolled. Let $S = \{1, 2, 3, 4, 5, 6\}$ be the sample space. Define two random variables X and Y based on the outcome of the roll:

$$X = \begin{cases} 1 & \text{if the outcome is less than 4} \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad Y = \begin{cases} 1 & \text{if the outcome is 1} \\ 0 & \text{otherwise} \end{cases}$$

   (a) Calculate the joint probability mass function of (X, Y).
   (b) Calculate the marginal probability mass functions of X and Y.
   (c) Let $A = \{Y = 1\}$, $B = \{X = 1\}$ and $C = \{X = 0\}$. Calculate $P(A|B)$ and $P(A|C)$.

3. The alcohol percentage in a one-liter bottle of a certain beverage is a continuous random variable X with density:

$$f_X(x) = c(1 - x^2)\mathbf{1}_{(0,1)}(x)$$

   where $\mathbf{1}_{(0,1)}(x)$ is the indicator function for the interval $(0, 1)$.

   (a) What value must c take for f_X to truly be a probability density function?
   (b) Find the cumulative distribution function (CDF) of X.
   (c) If the alcohol content of the bottles is to be divided into two equally probable categories ("low" and "high"), what is the threshold value that defines the median
   (d) If a case of these beverages contains ten one-liter bottles, assuming the bottles are independent, calculate the probability that in a case there are at least 2 bottles with an alcohol percentage lower than 50%.

Problem 1

Maria has three baby children: A (Ana), B (Ben), and C (Carl). She puts them to sleep at night. Because she sings beautifully, all three fall asleep as soon as she finishes the song. The number of consecutive hours each baby sleeps is a discrete random variable that takes the values 6 and 8 with probability 1/2 each. We assume that these three random variables are independent of each other. Find the distribution of the number of consecutive hours the mother sleeps, if she falls asleep at the same moment as her three children (she is very tired) and wakes up when the first of her three children wakes up.

Solution

Let A, B, and C be the number of consecutive hours Ana, Ben, and Carl sleep, respectively. The PMF for each child is identical and independent:

$$P(A = a) = P(B = b) = P(C = c) = \begin{cases} 1/2 & \text{if } a, b, c \in \{6, 8\} \\ 0 & \text{otherwise} \end{cases}$$

Let M be the number of consecutive hours the mother sleeps. Since the mother wakes up when the first child wakes up, M is the minimum of the three sleeping times:

$$M = \min(A, B, C)$$

Since A, B, C can only take values 6 or 8, M can only take the values 6 or 8.

Calculating $P(M = 8)$ The mother sleeps for 8 hours if and only if all three children sleep for 8 hours.

$$M = 8 \iff A = 8 \text{ and } B = 8 \text{ and } C = 8$$

Due to the independence of A, B, and C:

$$\begin{aligned} P(M = 8) &= P(A = 8, B = 8, C = 8) \\ &= P(A = 8)P(B = 8)P(C = 8) \\ &= \left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\left(\frac{1}{2}\right) \\ &= \frac{1}{8} \end{aligned}$$

Calculating $P(M = 6)$ Since M is a discrete variable that only takes values 6 or 8, $P(M = 6)$ is the complement of $P(M = 8)$:

$$P(M = 6) = 1 - P(M = 8) = 1 - \frac{1}{8} = \frac{7}{8}$$

The distribution of the number of consecutive hours the mother sleeps is:

$$P(M = m) = \begin{cases} 7/8 & \text{if } m = 6 \\ 1/8 & \text{if } m = 8 \\ 0 & \text{otherwise} \end{cases}$$

Probability Ana sleeps 8 hours: $P(A = 8) = \frac{1}{2}$ Probability the Mother sleeps 8 hours: $P(M = 8) = \frac{1}{8}$. Since $1/2 > 1/8$, the probability that Ana sleeps 8 consecutive hours is significantly higher than the probability that the mother sleeps 8 consecutive hours. The mother's sleep time is severely restricted by the minimum sleep time among her three children.

In this example, we see that what changes the distribution of the minimum is, of course, that the minimum is systematically smaller than the rest of the random variables. That is why, if we compare the number of hours Maria sleeps, we see that the probability of her sleeping 8 consecutive hours is much lower than the probability of Ana doing so. The same happens in the rest of the examples we saw, but this one is very simple, so we can clearly see what happens. In fact, we can make the following construction.

Let us make the following construction of random variables based on a uniform variable $U \sim U(0,1)$. Let:

$$
\begin{aligned}
X_1 &= 6 \cdot I_{(0,1/2)}(U) + 8 \cdot I_{[1/2,1)}(U)\\
X_2 &= 6 \cdot I_{(0,1/4)\cup[1/2,3/4)}(U) + 8 \cdot I_{[1/4,1/2)\cup[3/4,1)}(U)\\
X_3 &= 6 \cdot I_{(0,1/8)\cup[2/8,3/8)\cup[4/8,5/8)\cup[6/8,7/8)}(U) + 8 \cdot I_{[1/8,2/8)\cup[3/8,4/8)\cup[5/8,6/8)\cup[7/8,1)}(U)
\end{aligned}
$$

1. Verify that X_1, X_2, X_3 are random variables that have the same distribution as the sleeping hours from the previous Example.

2. Prove that they are independent random variables. (Hint: Try to perform the minimum amount of calculations possible to demonstrate independence).

3. Graph them on the same coordinate axis, and graph their minimum, $Y = \min\{X_1, X_2, X_3\}$, the number of consecutive sleeping hours of the mother.

We can see that that for each X_i, $P(X_i = 6) = 1/2$ and $P(X_i = 8) = 1/2$. Since $U \sim U(0,1)$, the probability of U falling into any set $S \subset (0,1)$ is equal to the length (Lebesgue measure) of S, denoted $|S|$. For example

- $P(X_1 = 6) = P(U \in (0,1/2)) = |(0,1/2)| = 1/2$.
- $P(X_1 = 8) = P(U \in [1/2,1)) = |[1/2,1)| = 1/2$.

Since X_i are discrete variables taking values in $\{6,8\}$, the variables are independent if and only if for all combinations $(x_1, x_2, x_3) \in \{6,8\}^3$

$$P(X_1 = x_1, X_2 = x_2, X_3 = x_3) = P(X_1 = x_1)P(X_2 = x_2)P(X_3 = x_3)$$

Given the symmetry, it is sufficient to prove independence for the event $X_i = 8$. Let $S_i = \{u \in (0,1) : X_i(u) = 8\}$. We left this as an exercise.

For these variables, the minimum function $Y(u) = \min\{X_1(u), X_2(u), X_3(u)\}$ will be equal to 8 only when $X_1(u) = 8$ AND $X_2(u) = 8$ AND $X_3(u) = 8$. This occurs on the interval $S_1 \cap S_2 \cap S_3 = [7/8, 1)$.

$$Y(u) = \begin{cases} 8 & \text{if } u \in [7/8, 1) \\ 6 & \text{if } u \in (0, 7/8) \end{cases}$$

This confirms that $P(Y = 8) = 1/8$, consistent with the calculation in the previous problem.

Problem 2

A fair six-sided die is rolled. Let $S = \{1,2,3,4,5,6\}$ be the sample space. Define two random variables X and Y based on the outcome of the roll:

$$X = \begin{cases} 1 & \text{if the outcome is less than 4} \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad Y = \begin{cases} 1 & \text{if the outcome is 1} \\ 0 & \text{otherwise} \end{cases}$$

1. Calculate the joint probability mass function of (X, Y).

2. Calculate the marginal probability mass functions of X and Y.

3. Let $A = \{Y = 1\}$, $B = \{X = 1\}$ and $C = \{X = 0\}$. Calculate $P(A|B)$ and $P(A|C)$.

Solution

Variable Definitions and Events

The possible outcomes and their probabilities are $P(\text{outcome} = k) = 1/6$ for $k \in \{1, \ldots, 6\}$.

The events corresponding to X and Y are:

- $X = 1$: outcome is $\{1, 2, 3\}$. (3 outcomes)
- $X = 0$: outcome is $\{4, 5, 6\}$. (3 outcomes)
- $Y = 1$: outcome is $\{1\}$. (1 outcome)
- $Y = 0$: outcome is $\{2, 3, 4, 5, 6\}$. (5 outcomes)

The range of (X, Y) is the set of possible pairs: $\mathcal{S}_{X,Y} = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$.

We calculate the probability for each pair:

1. $\mathbf{p_{X,Y}(1, 1) = P(X = 1, Y = 1)}$: Outcome is < 4 (i.e., $\{1, 2, 3\}$) AND outcome is 1 (i.e., $\{1\}$).
$$P(X = 1, Y = 1) = P(\text{outcome} = 1) = \mathbf{1/6}$$
2. $\mathbf{p_{X,Y}(1, 0) = P(X = 1, Y = 0)}$: Outcome is < 4 (i.e., $\{1, 2, 3\}$) AND outcome is $\neq 1$ (i.e., $\{2, 3, 4, 5, 6\}$).
$$P(X = 1, Y = 0) = P(\text{outcome} \in \{2, 3\}) = 1/6 + 1/6 = \mathbf{2/6}$$
3. $\mathbf{p_{X,Y}(0, 1) = P(X = 0, Y = 1)}$: Outcome is ≥ 4 (i.e., $\{4, 5, 6\}$) AND outcome is 1 (i.e., $\{1\}$).
$$P(X = 0, Y = 1) = P(\emptyset) = \mathbf{0}$$
4. $\mathbf{p_{X,Y}(0, 0) = P(X = 0, Y = 0)}$: Outcome is ≥ 4 (i.e., $\{4, 5, 6\}$) AND outcome is $\neq 1$ (i.e., $\{2, 3, 4, 5, 6\}$).
$$P(X = 0, Y = 0) = P(\text{outcome} \in \{4, 5, 6\}) = 1/6 + 1/6 + 1/6 = \mathbf{3/6}$$

The range is $\mathcal{S}_{X,Y} = \{(0, 0), (1, 0), (1, 1)\}$. Note that $P(0, 1) = 0$, so $(0, 1)$ is not a possible value pair, though it is sometimes included in the domain. A strict range would exclude it.

$$p_X(x) = \sum_y p_{X,Y}(x, y)$$

- $\mathbf{p_X(1)}$: $P(X = 1) = p_{X,Y}(1, 0) + p_{X,Y}(1, 1) = 2/6 + 1/6 = \mathbf{3/6} = \mathbf{1/2}$
- $\mathbf{p_X(0)}$: $P(X = 0) = p_{X,Y}(0, 0) + p_{X,Y}(0, 1) = 3/6 + 0 = \mathbf{3/6} = \mathbf{1/2}$

$$p_Y(y) = \sum_x p_{X,Y}(x, y)$$

- $\mathbf{p_Y(1)}$: $P(Y = 1) = p_{X,Y}(0, 1) + p_{X,Y}(1, 1) = 0 + 1/6 = \mathbf{1/6}$
- $\mathbf{p_Y(0)}$: $P(Y = 0) = p_{X,Y}(0, 0) + p_{X,Y}(1, 0) = 3/6 + 2/6 = \mathbf{5/6}$

We use the definition $P(A|B) = \frac{P(A \cap B)}{P(B)}$.

1. **Calculate** $P(A|B) = P(Y = 1|X = 1)$:
   - $A \cap B = \{Y = 1 \cap X = 1\}$. From section 1(a), $P(A \cap B) = P(X = 1, Y = 1) = 1/6$.

- $P(B) = P(X = 1)$. From section 2, $P(B) = 3/6$.

$$P(Y = 1|X = 1) = \frac{P(X = 1, Y = 1)}{P(X = 1)} = \frac{1/6}{3/6} = \mathbf{1/3}$$

2. **Calculate** $P(A|C) = P(Y = 1|X = 0)$:

- $A \cap C = \{Y = 1 \cap X = 0\}$. From section 1(c), $P(A \cap C) = P(X = 0, Y = 1) = 0$.
- $P(C) = P(X = 0)$. From section 2, $P(C) = 3/6$.

$$P(Y = 1|X = 0) = \frac{P(X = 0, Y = 1)}{P(X = 0)} = \frac{0}{3/6} = \mathbf{0}$$

Problem 3

The alcohol percentage in a one-liter bottle of a certain beverage is a continuous random variable X with density:

$$f_X(x) = c(1 - x^2)\mathbf{1}_{(0,1)}(x)$$

where $\mathbf{1}_{(0,1)}(x)$ is the indicator function for the interval $(0, 1)$.

1. What value must c take for f_X to truly be a probability density function?
2. Find the cumulative distribution function (CDF) of X.
3. If the alcohol content of the bottles is to be divided into two equally probable categories ("low" and "high"), what is the threshold value that defines the median
4. If a case of these beverages contains ten one-liter bottles, assuming the bottles are independent, calculate the probability that in a case there are at least 2 bottles with an alcohol percentage lower than 50%.

Solution

Finding the Constant c

For $f_X(x)$ to be a valid PDF, its integral over its support must equal 1:

$$\int_{-\infty}^{\infty} f_X(x)dx = \int_0^1 c(1 - x^2)dx = 1$$

$$c\left[x - \frac{x^3}{3}\right]_0^1 = 1$$

$$c\left(1 - \frac{1}{3}\right) = 1 \quad \Rightarrow \quad c\left(\frac{2}{3}\right) = 1$$

Result: $c = \mathbf{3/2}$. The PDF is $f_X(x) = \frac{3}{2}(1 - x^2)$ for $0 < x < 1$.

Cumulative Distribution Function (CDF)

For $0 \leq x \leq 1$:

$$F_X(x) = \int_0^x f_X(t)dt = \int_0^x \frac{3}{2}(1 - t^2)dt = \frac{3}{2}\left[t - \frac{t^3}{3}\right]_0^x$$

$$\mathbf{F_X(x) = \frac{3x}{2} - \frac{x^3}{2} \quad \text{for } 0 \leq x \leq 1}$$

And $F_X(x) = 0$ for $x < 0$, $F_X(x) = 1$ for $x > 1$.

Threshold (Median)

The threshold m (median) must satisfy $\P(X \leq m) = F_X(m) = 0.5$.

$$\frac{3m}{2} - \frac{m^3}{2} = \frac{1}{2}$$

The problem simplifies to solving the cubic equation:

$$m^3 - 3m + 1 = 0 \quad \text{for } m \in (0, 1)$$

Binomial Distribution

The criterion is $P_p = ¶(X < 0.50) = F_X(0.5)$.

$$P_p = F_X(1/2) = \frac{3(1/2)}{2} - \frac{(1/2)^3}{2} = \frac{3}{4} - \frac{1/8}{2} = \frac{3}{4} - \frac{1}{16} = \frac{11}{16}$$

Let Z be the number of bottles with $X < 0.50$. Since the bottles are independent, $Z \sim \text{Binomial}(n = 10, p = 11/16)$. The probability of failure is $q = 1 - p = 5/16$.

1) Probability of at least 2 bottles with $X < 0.50$

$$¶(Z \geq 2) = 1 - ¶(Z < 2) = 1 - [¶(Z = 0) + ¶(Z = 1)]$$

$$¶(Z = 0) = \binom{10}{0} p^0 q^{10} = \left(\frac{5}{16}\right)^{10}$$

$$¶(Z = 1) = \binom{10}{1} p^1 q^9 = 10 \left(\frac{11}{16}\right) \left(\frac{5}{16}\right)^9$$

$$¶(Z \geq 2) = 1 - \left[\left(\frac{5}{16}\right)^{10} + 10 \left(\frac{11}{16}\right) \left(\frac{5}{16}\right)^9\right]$$