Probability and Statistics

Y-DATA School of Data Science

P&P 2

Due: 16.11.2022

PROBLEM 1. Let $X \sim U(1,5)$ (discrete). Define a new random variable $Y = 3^X$.

(1) What is the probability mass function of Y?
(2) Compute the expected value of Y.

(1) The PMF of Y is

$$P_Y(y) = \begin{cases} 1/5 & y = 3 \\ 1/5 & y = 9 \\ 1/5 & y = 27 \\ 1/5 & y = 81 \\ 1/5 & y = 243 \end{cases}$$

(2) Using the formula for expectation of a function of a random variable,

$$\begin{aligned} E(Y) &= \sum_{x \in supp(X)} g(x) P_X(x) \\ &= \sum_{x=1}^{5} 3^x \cdot \frac{1}{5} \\ &= \frac{1}{5} \cdot (3^1 + 3^2 + 3^3 + 3^4 + 3^5) = 72.6 \end{aligned}$$

PROBLEM 2. According to the British secret intelligence service, during the war, the expected number of bombs that fall per day in each quarter of London is 2. It is known that the number of bombs that fall in one day in each quarter is a Poisson random variable.

(1) What is the probability that on some day there will be no bombs at all?
(2) What is the probability that at least 4 bombs will fall on some specific quarter in one day?

(1) It is given that $X \sim Pois(2)$. Thus,

$$P(X = 0) = \frac{e^{-2} 2^0}{0!} = e^{-2}$$

(2)

$$\begin{aligned} P(X \geq 4) &= 1 - P(X < 4) = 1 - P(X \leq 3) \\ &= 1 - \sum_{x=0}^{3} \frac{e^{-2} 2^x}{x!} \\ &= 1 - 0.857 = 0.143 \end{aligned}$$

PROBLEM 3. Let X be some discrete random variable. Show that for any constants $a, b \in \mathbb{R}$,

(1) $E(aX+b) = aE(X)+b$
(2) $Var(aX+b) = a^2Var(X)$

(1)

$$\begin{aligned}E(aX+b) &= \sum_{x\in supp(X)} (ax+b)P_X(x) = \sum_{x\in supp(X)} (axP_X(x)+bP_X(x)) \\ &= \sum_{x\in supp(X)} axP_X(x) + \sum_{x\in supp(X)} bP_X(x) \\ &= a\sum_{x\in supp(X)} xP_X(x) + b\sum_{x\in supp(X)} P_X(x) = aE(X)+b\end{aligned}$$

where the last equality holds from the definition of expectation and because

$$\sum_{x\in supp(X)} P_X(x) = 1$$

(2)

$$\begin{aligned}Var(aX+b) &= E(aX+b-E(aX+b))^2 \\ &= E(a(X-E(X))+b-b))^2 \\ &= Ea^2(X-E(X))^2 = a^2E(X-E(X))^2 = a^2Var(X)\end{aligned}$$

where the last two equalities follow from linearity of expectation and the definition of variance, respectively.

PROBLEM 4. Give a simple example of some random variable X and some function g to show that in general $Eg(X) \neq g(EX)$.

We can simply take the scenario form the first problem. Take $X \sim U(1,5)$ and $g(X) = 3^X$. Then,

$$E(X) = \frac{1}{5}(1+2+3+4+5) = \frac{15}{5} = 3$$

and

$$E(g(X)) = 72.6 \neq 3^3 = 3^{E(X)} = g(E(X))$$

PROBLEM 5. In a multiple-choice exam, there are 10 questions. Each has 4 possible answers (only one correct). Alice didn't prepare for the exam, so she guessed all her answers. Let X denote the number of her correct answers.

(1) What is the distribution of X? Write its PMF.
(2) To pass the test, a student should get 55%. What is the probability that Alice passed the test?

(1) Alice's guesses are independent with probability of success $1/4$. Thus, $X \sim Bin(10,1/4)$. The PMF is given by

$$P(X=x) = \binom{10}{x}(\frac{1}{4})^x(\frac{3}{4})^{10-x}, x = 0,1,2,...,10$$

(2) In order to pass the exam, Alice must get at least 6 answers correct (she can't get exactly 55%). The probability of that is,

$$\begin{aligned}
P(X \geq 6) &= \sum_{k=6}^{10} P(X=k) \\
&= \sum_{k=6}^{10} \binom{10}{k} (\frac{1}{4})^k (\frac{3}{4})^{10-k} \\
&= \binom{10}{6}(\frac{1}{4})^6(\frac{3}{4})^4 + \binom{10}{7}(\frac{1}{4})^7(\frac{3}{4})^3 + \binom{10}{8}(\frac{1}{4})^8(\frac{3}{4})^2 + \binom{10}{9}(\frac{1}{4})^9(\frac{3}{4})^1 + \binom{10}{10}(\frac{1}{4})^{10}(\frac{3}{4})^0 \\
&= 0.02
\end{aligned}$$

PROBLEM 6. Let X be a random variable with the following density function,

$$f_X(x) = \begin{cases} ax & \text{if } 0 < x < 1 \\ a & \text{if } 1 \leq x < 2 \\ a(3-x) & \text{if } 2 \leq x < 3 \\ 0 & \text{o.w.} \end{cases}$$

(1) Find the constant a for which f_X is a density function.
(2) Compute the expectation and variance of X.
(3) Find the cumulative distribution function of X.

(1) We integrate over $\mathbb{R}$ and equate to one:

$$\begin{aligned}
1 &= \int_{-\infty}^{\infty} f_X(x)dx = \int_0^1 axdx + \int_1^2 adx + \int_2^3 a(3-x)dx = \\
&= a\left[\frac{x^2}{2}\bigg|_0^1 + x|_1^2 + \left(3x - \frac{x^2}{2}\right)\bigg|_2^3\right] = \\
&= a\left(\frac{1}{2} + 2 - 1 + 9 - \frac{9}{2} - 6 + \frac{4}{2}\right) = 2a \\
a &= \frac{1}{2}
\end{aligned}$$

(2) For the expectation we have,

$$\begin{aligned}
E(X) &= \int_0^1 x \cdot \frac{1}{2}xdx + \int_1^2 x \cdot \frac{1}{2}dx + \int_2^3 x \cdot \frac{1}{2}(3-x)dx = \\
&= \frac{x^3}{6}\bigg|_0^1 + \frac{x^2}{4}\bigg|_1^2 + \left(\frac{3x^2}{4} - \frac{x^3}{6}\right)\bigg|_2^3 = \\
&= \frac{1}{6} + 1 - \frac{1}{4} + \frac{27}{4} - \frac{27}{6} - 3 + \frac{8}{6} = \frac{3}{2}
\end{aligned}$$

For the variance, we need to calculate the second moment as well,

$$
\begin{aligned}
E\left(X^2\right) &= \int_0^1 x^2 \cdot \frac{1}{2} x dx + \int_1^2 x^2 \cdot \frac{1}{2} dx + \int_2^3 x^2 \cdot \frac{1}{2}(3-x) dx = \\
&= \left.\frac{x^4}{8}\right|_0^1 + \left.\frac{x^3}{6}\right|_1^2 + \left.\left(\frac{3x^3}{6} - \frac{x^4}{8}\right)\right|_2^3 = \\
&= \frac{1}{8} + \frac{8}{6} - \frac{1}{6} + \frac{81}{6} - \frac{81}{8} - 4 + 2 = \frac{8}{3}
\end{aligned}
$$

Therefore,

$$
Var(X) = E(X^2) - E^2(X) = \frac{8}{3} - 1.5^2 = \frac{5}{12}
$$

(3) We know that

$$
F_X(x) = \int_{-\infty}^{x} f_X(t) dt
$$

We need to integrate the PDF over the relevant intervals.
For $0 < x < 1$,

$$
F_X(x) = \int_0^x 1/2 t dt = 1/2 \frac{t^2}{2}|_0^x = \frac{1}{4} x^2
$$

For $1 \le x < 2$,

$$
F_X(x) = 1/2 \frac{t^2}{2}|_0^1 + \int_1^x 1/2 dt = \frac{1}{4} + 0.5(x-1)
$$

For $2 \le x < 3$,

$$
\begin{aligned}
F_X(x) = & \left.\frac{1}{4}x^2\right|_0^1 + 0.5(x-1)|_1^2 + \int_2^x 0.5(3-t) dt \\
= & \quad \frac{3}{4} + 0.5\left.\left(3t - \frac{t^2}{2}\right)\right|_2^x \\
= & \quad \frac{3}{4} + 0.5\left(3x - \frac{x^2}{2} - 4\right) \\
= & \quad \frac{3}{4} - \frac{1}{4}\left(x^2 - 6x + 8\right) \\
= & \quad \frac{3}{4} - \frac{1}{4}(x-4)(x-2)
\end{aligned}
$$

Overall, the CDF is given by,

$$
F_X(x) = \begin{cases} 0 & x \le 0 \\ \frac{1}{4}x^2 & 0 < x < 1 \\ \frac{1}{2}x - \frac{1}{4} & 1 \le x < 2 \\ \frac{3}{4} - \frac{1}{4}(x-4)(x-2) & 2 \le x < 3 \\ 1 & x \ge 3 \end{cases}
$$

PROBLEM 7. Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables with distribution $Bin(48, 1/4)$. Let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Use the central limit theorem to calculate

$$P(\bar{X}_{144} > 12.75)$$

First, note that $\mu = E(X_1) = 48 \cdot \frac{1}{4} = 12$ and $\sigma^2 = Var(X_1) = 48 \cdot \frac{1}{4} \cdot \frac{3}{4} = 9$. Therefore, $\sigma = SD(X_1) = 3$. In this case, $n = 144$.
We are ready to compute the probability:

$$\begin{aligned}
P(\bar{X}_{144} > 12.75) &= P(\frac{\bar{X}_{144} - \mu}{\sigma}\sqrt{n} > \frac{12.75 - \mu}{\sigma}\sqrt{n}) \\
&= P(\frac{\bar{X}_{144} - 12}{3} 12 > \frac{12.75 - 12}{3} 12) \\
&\approx P(Z > \frac{12.75 - 12}{3} 12) \\
&= P(Z > 3) = 1 - P(Z < 3) = 1 - \Phi(3) = 1 - 0.9987 = 0.0013
\end{aligned}$$

where the approximation to $Z \sim N(0,1)$ is true due to the central limit theorem.