Possibility and Statistics

Zihang Wang - zihang.wang@epita.fr

02/01/2023

PartI

Are the following functions f (defined on $\mathbf{R}$) PDFs? If they are, what is their expected value?

Q1

$f(x) = \cos(x)$ if $x \in (0, \frac{\pi}{2})$ and 0 elsewhere

Solution: Rewrite the fuction as:

$$f(x), x \in \mathbf{R} = \begin{cases} \cos(x) & x \in (0, \frac{\pi}{2}) \\ 0 & \end{cases}$$

It is a PDF because $f(x) \geq 0$ when $x \in X$ and $\displaystyle\int f(x)dx = 1$

$$\begin{aligned} E[X] &= \int x \cdot f(x)dx \\ &= \int_0^{\frac{\pi}{2}} xcos(x)dx + 0 \\ &= cosx + xsinx \Rightarrow F(x)\Big|_0^{\frac{\pi}{2}} = \frac{\pi}{2} - 1 \end{aligned}$$

Q2

$g(x) = sin(x) + 1, \forall x \in \mathbf{R}$

Solution: $\displaystyle\int g(x) = x - cosx, x \in \mathbf{R}$, the integral can not be 1 so it is not a PDF

Q3

$h(x) = \dfrac{1}{b-a}, x \in (a,b)$ (with $a, b \in \mathbf{R}$ and $a < b$) and 0 everywhere else

Solution: Rewrite the fuction as:

$$h(x), x \in \mathbf{R} = \begin{cases} \dfrac{1}{b-a}, x \in (a,b) \\ 0 \end{cases}$$

It is a PDF because a, b are constants and $a < b$, therefore $h(x) > 0$ and $\displaystyle\int h(x) = 1$

$$\begin{aligned} E[X] &= \int x \cdot h(x) dx \\ &= \int_a^b \frac{x}{b-a} dx + 0 \\ &= \frac{x^2}{2(b-a)} \Rightarrow F(x)\Big|_a^b = \frac{a+b}{2} \end{aligned}$$

PartII

In a country XYZ, the height of a 18-year-old man has a normal distribution with a mean 170cm and a standard deviation of 4cm.

Q1

What is the percentage of 18-year-old men being more than 1m74?
Solution:

$$\begin{aligned} 1.74 &= 1.70 + 0.04 \\ &= \mu + \sigma \Rightarrow \\ \mathbb{P}(h > 1.74) &= \frac{1 - 68.27\%}{2} = 15.87\% \end{aligned}$$

Q2

Among 18-year-old men being more than 1m80, what proportion is more than 1m82?

Solution:

$$\begin{aligned}
&Z_1 = (1.8 - 1.7)/0.04 = 2.5 \Rightarrow \mathbb{P}(h > 1.8) = 1 - 0.9938 = 0.0062\\
&Z_2 = (1.82 - 1.7)/0.04 = 3 \Rightarrow \mathbb{P}(h > 1.82) = 1 - 0.9987 = 0.0013 \Rightarrow\\
&\frac{\mathbb{P}(h > 1.82)}{\mathbb{P}(h > 1.8)} = 20.97\%
\end{aligned}$$

Q3

Give 2 different ways to find the height of a 18-year-old man to have a 50% chance to be taller than a random other 18-year-old man.

Solution:
S1:Z-score:

$$\begin{aligned}
&\mathbb{P} = 0.5 \Rightarrow\\
&Z = (h - \mu)/\sigma = 0 \Rightarrow\\
&h = 1.7
\end{aligned}$$

S2:Property of normal distribution:

$$\begin{aligned}
&\text{To find the 50\% portion that means to find the medium}\\
&\text{Medium} = \mu = 1.7 \Rightarrow\\
&h = 1.7
\end{aligned}$$

PartIII

A department store's breakdown service has teams who respond to customer calls. For a variety of reasons, call-outs are sometimes delayed. It is assumed that the calls occur independently of each other, and that for each call, the probability of a delay is 0.25.A customer calls the service department 4 times. We denote by X the random variable taking as its values the number of times this customer has experienced a delay.

Q1

Determine the probability distribution of X, its expectation and variance.

Solution:
It is a binomial distribution because there are two possible outcomes in total and they are

independent. Therefore, it can be denoted as:

$$P(X = x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$
$$E(X) = np$$
$$\sigma^2 = V(X) = np(1-p)$$

Q2

Calculate the probability of the event: "The customer experienced at least one delay".

Solution: We need to determine the zero delay at one day firstly. According to Q1, we get:

$$\begin{aligned} P(X = 0) &= \frac{4!}{0!(4-0)!} 0.25^0 (1-0.25)^{4-0} \\ &= 1 \times 1 \times 0.75^4 \\ &= 0.31640625 \Rightarrow \\ P(X >= 1) &= 1 - P(X = 0) \approx 0.68 \end{aligned}$$

PartIIII

In a small country with a population of 10 million people, 100 000 have an illness ABC. There is a test for that disease with the following numbers :

- if a person is sick, the test is positive with a probability of 98%

- if a person is not sick, the test is positive with a probability of 6%

Q1

What is the probability to be sick for a person whose test results are positive? What do you think of that probability?

Solution:
In conclusion, we know:

$$\mathbb{P}(sick) = \mathbb{P}(s) = 100000/10000000 = 0.01$$
$$\mathbb{P}(health) = \mathbb{P}(h) = 1 - 0.01 = 0.99$$
$$\mathbb{P}(+|s) = 0.98$$
$$\mathbb{P}(+|h) = 0.06$$

To find $\mathbb{P}(s|+)$:

$$
\begin{aligned}
\mathbb{P}(+) &= \mathbb{P}(+|s) \cdot \mathbb{P}(s) + \mathbb{P}(+|h) \cdot \mathbb{P}(h) \\
&= 0.98 \times 0.01 + 0.06 \times 0.99 \\
&= 0.0692
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{P}(s|+) &= \frac{\mathbb{P}(+|s) \cdot \mathbb{P}(s)}{\mathbb{P}(+)} \\
&= \frac{0.98 \times 0.01}{0.0692} \\
&\approx 0.14 = 14\%
\end{aligned}
$$

This result shows the possibility of the real patients among the ”patients” we consider is low. To address this problem, the testers are supposed to take multiple tests for their patients.

Q2

Let’s say the test results of a person are positive and she decides to take it another time : what is the probability of being healthy if this second test results are also positive? What is your interpretation of the value of that probability?

Solution:

To find $\mathbb{P}(h| +_1 \cap +_2)$:

$$\begin{aligned}\mathbb{P}(+_1 \cap +_2|s) &= \mathbb{P}(+_1|s)\mathbb{P}(+_2|s) \\ &= 0.98^2 \\ &= 0.9604\end{aligned}$$

$$\begin{aligned}\mathbb{P}(+_1 \cap +_2|h) &= \mathbb{P}(+_1|h)\mathbb{P}(+_2|h) \\ &= 0.06^2 \\ &= 0.0036\end{aligned}$$

$$\begin{aligned}\mathbb{P}(+_1 \cap +_2) &= \mathbb{P}(+_1 \cap +_2|s)\mathbb{P}(s) + \mathbb{P}(+_1 \cap +_2|h)\mathbb{P}(h) \\ &= 0.9604 + 0.0036 \\ &= 0.964\end{aligned}$$

$$\begin{aligned}\mathbb{P}(h| +_1 \cap +_2) &= \frac{\mathbb{P}(h \cap (+_1 \cap +_2))}{\mathbb{P}(+_1 \cap +_2)} \\ &= \frac{\mathbb{P}(+_1 \cap +_2|h)\mathbb{P}(h)}{\mathbb{P}(+_1 \cap +_2)} \\ &= \frac{0.0036 \times 0.99}{0.964} \\ &\approx 0.00396 = 0.396\%\end{aligned}$$

The result determines that 2 or more test could precisely detect the real patients. If someone was detected sick in the first test, his best choice is to take another test instead of crying.