MA2261 - DLI, Linear Statistical Models, Year 2022-2023

Solutions of exercises for feedback class 1

(Note: the exercise number refers to the workbook)

EXERCISE 1.1

i) $P(A \cap B)/P(B)$.

ii) $P(A|B) = P(A \cap B)/P(B) = P(B|A)P(A)/P(B)$ by definition of $P(B|A)$

iii) We define the following events

$$\begin{aligned} t_+ &= \{\text{Test positive}\} \\ t_- &= \{\text{Test negative}\} \\ d_+ &= \{\text{someone have the disease}\} \\ d_- &= \{\text{someone don't have the disease}\} \end{aligned}$$

Then from the given information,

$$P(t_+|d_+) = 0.95, P(t_-|d_-) = 0.9, P(d_+) = 0.0025$$

Therefore

$$\begin{aligned} P(d_+|t_+) =& \frac{P(t_+|d_+)P(d_+)}{P(t_+|d_+)P(d_+) + P(t_+|d_-)P(d_-)} \\ =& \frac{0.95 \times 0.0025}{0.95 \times 0.0025 + (1 - 0.9) \times 0.9975} \approx 0.0233 \end{aligned}$$

EXERCISE 1.8

i) Two random variables X and Y are independent if $P(X \leq x \text{ and } Y \leq y) = P(X \leq x)P(Y \leq y)$ for all x and y.

ii) $f_{X,Y}(x, y) = f_X(x)f_Y(y)$.

iii) $f_X(x) = \int_x^1 15x^2 y dy = 15x^2 \left[\frac{1}{2}y^2\right]_x^1 = \frac{15x^2(1-x^2)}{2}$,

$f_Y(y) = \int_0^y 15x^2 y dx = y \left[5x^3\right]_0^y = 5y^4$.

iv) No, $f_{X,Y}(x,y) \neq f_X(x)f_Y(y)$.

v) $E[Y|X=x] = \int y \frac{f_{X,Y}(x,y)}{f_X(x)} dy = \int_x^1 \frac{15x^2y^2}{15x^2(1-x^2)/2} dy = \frac{2}{1-x^2}\int_x^1 y^2 dy = \frac{2(1-x^3)}{3(1-x^2)}$.

EXERCISE 1.10

Recall the Central Limit Theorem:
Let $S = X_1 + \cdots + X_n$ where X_i, $i = 1, \ldots, n$ are n independently and identically distributed random variables with $\mu = E[X_i]$ and $\sigma^2 = \mathrm{var}\,(X_i)$, then

$$T_n = \frac{(S - n\mu)}{\sigma\sqrt{n}}$$

is approximately distributed as $N(0,1)$ with $\lim_{n\to\infty} P[T_n \leq t] = \Phi(t)$.

From exercise data we calculate

$$\begin{aligned}\mu &= E[X_i] = \frac{1}{4}(1+2+3+4) = 2.5\\ \sigma^2 &= \mathrm{var}\,(X_i) = E[X_i^2] - 2.5^2 = 1.25\end{aligned}$$

Thus by the CLT above,

$$T_{500} = \frac{S - 500 \times 2.5}{\sqrt{500 \times 1.25}} \sim N(0,1)$$

Therefore,

$$\begin{aligned}P(1275 \leq S \leq 1300) &= P\left(\frac{1275 - 1250}{\sqrt{625}} \leq T_{500} \leq \frac{1300 - 1250}{\sqrt{625}}\right)\\ &= P(1 \leq T_{500} \leq 2)\\ &= \Phi(2) - \Phi(1) = 0.97725 - 0.84134 = 0.13591\end{aligned}$$

EXERCISE 1.11

Let X be the number of customers who make a buy during a day. Then $X \sim \mathrm{Bin}(400, 0.1)$. In this case, $np = 400 \cdot 0.1 = 40$, $np(1-p) = 400 \cdot 0.1 \cdot 0.9 = 36$. Since this number is bigger than 10, we can expect that the normal approximation works well. Hence,

$$\begin{aligned}P(X \geq 30) &= 1 - P(X \leq 29) \approx 1 - \Phi\left(\frac{29 + 0.5 - 40}{\sqrt{36}}\right)\\ &= 1 - \Phi(-1.75) = \Phi(1.75) = 0.95994\end{aligned}$$