Homework 1

Chengbo Gu

Problem 1

(a) Let A denotes the event that none has high blood lead level in a randomly selected group of 16 children.

Then $P(A) = (1-\frac{1}{8})^{16} = 0.1181$

(b) Let B denotes the event that one child has high blood lead level in a randomly selected group of 16 children.

Then $P(B) = C_{16}^{1}\left(\frac{1}{8}\right)\times\left(\frac{7}{8}\right)^{15} = 0.2699$

(c) Let C denotes the event that two children have high blood lead level in a randomly selected group of 16 children.

Then $P(C) = C_{16}^{2}\left(\frac{1}{8}\right)^{2}\times\left(\frac{7}{8}\right)^{14} = 0.2891$

(d) Let D denotes the event that three or more have high blood lead level in a randomly selected group of 16 children.

Then $P(D) = 1 - P(A) - P(B) - P(C) = 0.3229$

Problem 2

Let X denotes the height of corn plants, then $X \sim N(145, 22)$

(a) $P(135 \le X \le 155) = P(\frac{135-145}{22} \le \frac{X-145}{22} \le \frac{155-145}{22}) = P(-\frac{5}{11} \le Z \le \frac{5}{11}) = P(Z \le \frac{5}{11}) - P(Z \le -\frac{5}{11})$

Because $pnorm(5/11) = 0.6753$ and $pnorm(-5/11) = 0.3247$,

$P(135 \le X \le 155) = 0.6753 - 0.3247 = 0.3506$

(b) According to sampling distribution, $\bar{X}_{16} \sim N(145, \frac{22}{\sqrt{16}}) \Rightarrow \bar{X}_{16} \sim N(145, 5.5)$ for a random sample of 16 plants.

$$P(135 \le \bar{X}_{16} \le 155) = P(\frac{135-145}{5.5} \le \frac{\bar{X}_{16}-145}{5.5} \le \frac{155-145}{5.5}) = P(-\frac{20}{11} \le Z \le \frac{20}{11}) = P(Z \le \frac{20}{11}) - P(Z \le -\frac{20}{11})$$

Because $pnorm(20/11) = 0.9655$ and $pnorm(-20/11) = 0.0345$,

$$P(135 \le \bar{X}_{16} \le 155) = 0.9655 - 0.0345 = 0.9310$$

(c) According to sampling distribution, $\bar{X}_{32} \sim N(145, \frac{22}{\sqrt{32}}) \Rightarrow \bar{X}_{32} \sim N(145, \frac{11}{2\sqrt{2}})$ for a random sample of 32 plants.

$$P(135 \le \bar{X}_{32} \le 155) = P(\frac{135-145}{22/\sqrt{32}} \le \frac{\bar{X}_{32}-145}{22/\sqrt{32}} \le \frac{155-145}{22/\sqrt{32}}) = P(Z \le 2.5713) - P(Z \le -2.5713)$$

$$P(135 \le \bar{X}_{32} \le 155) = pnorm(2.5713) - pnorm(-2.5713) = 0.9899$$

Problem 3

Let A denotes the event that individuals preferred to use right hand to write.

Let B denotes the event that individuals preferred to use right foot to kick a ball.

Then $P(B) = \frac{2012+121}{2391} = 0.8921$

$$P(B \mid A) = \frac{2012}{2012+142} = 0.9341$$

Since $P(B) \neq P(B \mid A)$, hand and foot preferences are not independent.

Exercise 2.1

Let G denotes a child is a girl.

Let B denotes a child is a boy.

Let A denotes that at least one child is boy.

Then we have $P(GG) = P(BB) = P(GB) = P(BG) = \frac{1}{4}$

(a) The question actually asks what is $P(GB \mid A) + P(BG \mid A)$.

Using Bayes rule, we have

$$P(GB \mid A) + P(BG \mid A) = 1 - P(BB \mid A) = 1 - \frac{P(A \mid BB) \times P(BB)}{P(A)} = 1 - \frac{1/4}{3/4} = \frac{2}{3}$$

(b) Let b denotes the event that I happen to see one of his children and it is a boy.

The question actually asks what is $P(GB \mid b) + P(BG \mid b)$.

Using Bayes rule, we have

$$P(GB \mid b) + P(BG \mid b) = \frac{P(b \mid GB) \times P(GB)}{P(b)} + \frac{P(b \mid BG) \times P(BG)}{P(b)} = \frac{1/2 \times 1/4}{1/2} + \frac{1/2 \times 1/4}{1/2} = \frac{1}{2}$$

Exercise 2.2

Let B denotes the event that one has the crime blood type.

Let G denotes the event that one is guilty.

Both prosecutor and defender want to infer $P(G \mid B)$ given some evidence.

Using Bayes rule, we have

$$P(G \mid B) = \frac{P(G) \times P(B \mid G)}{P(B)} = \frac{P(G) \times P(B \mid G)}{P(G) \times P(B \mid G) + P(\bar{G}) \times P(B \mid \bar{G})}$$

(a) The prosecutor only has the evidence that $P(B \mid \bar{G}) = \frac{1}{100}$ and nothing else.

Thus, there is no way for us to infer $P(G)$, $P(B)$ and $P(G \mid B)$.

So simply using $1 - P(B \mid \bar{G})$ to represent $P(G \mid B)$ is wrong.

(b) The defender has the evidence that $P(G) = \frac{1}{800000}$ and $P(B) = \frac{8000}{800000} = \frac{1}{100}$

Thus,

$$P(G \mid B) = \frac{P(G) \times P(B \mid G)}{P(B)} = \frac{1/800000 \times 1}{1/100} = \frac{1}{8000}$$

Strictly speaking, the defender was right until the last statement " thus has no relevance".

The posterior probability $P(G \mid B) = \frac{1}{8000}$ is 100 times as large as the priori probability $P(G) = \frac{1}{800000}$, thus it is not convincing to say one has no relevance just depending on $P(G \mid B)$ is a small number.

Exercise 2.4

Let A denotes the event that one has this disease.
Let B denotes the event that the test is positive.

Then we have $P(B \mid A) = P(\bar{B} \mid \bar{A}) = 0.99$ and $P(A) = 1/10000$.

We want to infer $P(A \mid B)$.

Using Bayes rule,

$$P(A \mid B) = \frac{P(A) \times P(B \mid A)}{P(B)} = \frac{P(A) \times P(B \mid A)}{P(A) \times P(B \mid A) + P(\bar{A}) \times P(B \mid \bar{A})} = \frac{1/10000 \times 0.99}{1/10000 \times 0.99 + 9999/10000 \times 0.01} = 0.98\%$$

Exercise 2.5

Let A denotes the event that the prize was behind door 1.
Let B denotes the event that the prize was behind door 2.
Let C denotes the event that the prize was behind door 3.
Let c denotes the event that I choose door 1 and the host open door 3.

Then we have $P(A) = P(B) = P(C) = \frac{1}{3}$, $P(c \mid A) = \frac{1}{2}$, $P(c \mid B) = 1$ and $P(c \mid C) = 0$.

Thus,

$$P(A \mid c) = \frac{P(A) \times P(c \mid A)}{P(c)} = \frac{P(A) \times P(c \mid A)}{P(A) \times P(c \mid A) + P(B) \times P(c \mid B) + P(C) \times P(c \mid C)}$$

$$= \frac{1/3 \times 1/2}{1/3 \times 1/2 + 1/3 \times 1 + 1/3 \times 0} = \frac{1}{3}$$

$$P(B \mid c) = 1 - P(A \mid c) = \frac{2}{3}$$

So, it is better to switch to door 2.

Exercise 2.12

$$I(X,Y) = \sum_x \sum_y P(x,y) \log \frac{P(x,y)}{p(x)p(y)} = \sum_x \sum_y P(x,y) \log \frac{P(x,y)}{p(x)} - \sum_x \sum_y P(x,y) \log P(y)$$

$$= \sum_x P(x) \left(\sum_y P(y \mid x) \log P(y \mid x) \right) - \sum_y \log P(y) \left(\sum_x P(x,y) \right)$$

$$= -\sum_x P(x) H(Y \mid X = x) - \sum_y \log P(y) P(y)$$

$$= -H(Y \mid X) + H(Y) = H(Y) - H(Y \mid X)$$

Similarly,

$$I(X,Y) = \sum_x \sum_y P(x,y) \log \frac{P(x,y)}{p(x)p(y)} = \sum_x \sum_y P(x,y) \log \frac{P(x,y)}{p(y)} - \sum_x \sum_y P(x,y) \log P(x)$$

$$= \sum_y P(y) \left(\sum_x P(x|y) \log P(x|y) \right) - \sum_x \log P(x) \left(\sum_y P(x,y) \right)$$

$$= -\sum_y P(y) H(X|Y=y) - \sum_x \log P(x) P(x)$$

$$= -H(X|Y) + H(X) = H(X) - H(X|Y)$$