EE5907/EE5027 Week 1: Probability Review Solutions

Exercise 2.6

(a) According to Bayes Rule,

$$\vec{P}(H|e_1, e_2) = \frac{P(H, e_1, e_2)}{P(e_1, e_2)} = \frac{P(e_1, e_2|H)P(H)}{P(e_1, e_2)} \tag{1}$$

thus (ii) is sufficient for calculation

(b) Given $E_1 \perp E_2|H$, $P(e_1, e_2|H) = P(e_1|H)P(e_2|H)$
From (a), we have

$$\vec{P}(H|e_1, e_2) = \frac{P(e_1, e_2|H)P(H)}{P(e_1, e_2)} \tag{2}$$

From $E_1 \perp E_2|H$, we have

$$\vec{P}(H|e_1, e_2) = \frac{P(e_1|H)P(e_2|H)P(H)}{P(e_1, e_2)} \tag{3}$$

Eq. (3) corresponds to terms in (i). In addition, we can calculate $P(e_1, e_2)$ by $\sum_H (P(e_1, e_2|H)P(H))$, so (iii) is also sufficient.
To conclude, (i),(ii),(iii) are all sufficient.

Exercise 2.7

Proof by counter example:

(I) Let X_1 and X_2 be outcomes of independent coin toss (1 means head, 0 means tails). $X_3 = X_1 \oplus X_2$, where $\oplus$ is XOR operator. $p(X_3|X_1, X_2) \neq p(X_3)$ since X_1 and X_2 determines X_3 deterministically, so X_1, X_2, X_3 are not mutually independent. However, $p(X_3|X_1) = p(X_3)$, $p(X_3|X_2) = p(X_3)$, so X_1, X_2, X_3 are pairwise independent.

(II) Consider a tetrahedron die where three of the faces are colored red, green, and blue respectively. On the fourth face, include all the three colors. Roll the dice and define following events:

X_1 : "red appeared on the face the dice landed on"

X_2 : "green appeared on the face the dice landed on"

X_3 : "blue appeared on the face the dice landed on"

Therefore

$$P(X_i, X_j) = \frac{1}{4} = P(X_i)P(X_j)$$

$$\begin{aligned} P(X_1, X_2, X_3) &= \frac{1}{4} \\ &\neq P(X_1)P(X_2)P(X_3) = \frac{1}{8} \end{aligned}$$

Therefore X_1, X_2, X_3 are pairwise independent, but not mutually independent.

Exercise 2.8

Proof

($\Rightarrow$) Given $X \perp Y|Z$, we have $p(x,y|z) = p(x|z)p(y|z)$. Let $g(x,z) = p(x|z)$ and $h(y,z) = p(y|z)$, then $p(x,y|z) = g(x,z)h(y,z)$.

($\Leftarrow$) Suppose $p(x,y|z) = g(x,z)h(y,z)$. Integrate both sides over x (or summation if x is discrete)

$$\begin{aligned} \int p(x,y|z)dx &= \int g(x,z)dx \times h(y,z) \\ \implies p(y|z) &= G(z)h(y,z), \end{aligned} \tag{4}$$

where $G(z) = \int g(x,z)dx$.

Integrate both sides over y (or summation if y is discrete)

$$\begin{aligned} \int p(x,y|z)dy &= g(x,z) \times \int h(y,z)dy \\ \implies p(x|z) &= g(x,z)H(z), \end{aligned} \tag{5}$$

where $H(z) = \int h(y,z)dy$

Finally, let's integrate with respect to both x and y:

$$1 = \int \int p(x, y|z) dx dy = \int \int g(x, z) h(y, z) dx dy \tag{6}$$

$$= \int g(x, z) dx \int h(y, z) dy = G(z) H(z) \tag{7}$$

Therefore

$$p(x, y|z) = g(x, z) h(y, z) = \frac{p(x|z)}{G(z)} \frac{p(y|z)}{H(z)} \text{ using Eq. (4) and Eq. (5)}$$

$$= p(x|z) p(y|z) \text{ using Eq. (7)} \tag{8}$$

Exercise 2.10

According to the "change of variable formula"

$$p_y(y) = p_x(x) \left| \frac{dx}{dy} \right|$$

In this case, $y = \frac{1}{x} \quad \Rightarrow \quad \frac{dy}{dx} = -\frac{1}{x^2} \quad \Rightarrow \quad \frac{dx}{dy} = -x^2$

$$p_y(y) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-xb} \cdot | - x^2| \tag{9}$$

Substitute x by $1/y$

$$p_y(y) = \frac{b^a}{\Gamma(a)} y^{-a+1} e^{-\frac{b}{y}} \cdot y^{-2} = \frac{b^a}{\Gamma(a)} y^{-(a+1)} e^{-\frac{b}{y}} \tag{10}$$

Since $IG(x|\text{shape} = a, \text{scale} = b) = \frac{b^a}{\Gamma(a)} x^{-(a+1)} e^{-\frac{b}{x}}$, y is $IG(a, b)$

Exercise 2.12

According to definition

$$
\begin{aligned}
I(X;Y) \triangleq KL(p(X,Y)||p(X)p(Y)) &= \sum_x \sum_y p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \\
&= \sum_x \sum_y p(x,y) \log \frac{p(y|x)}{p(y)} \\
&= \sum_x \sum_y p(x,y) \log p(y|x) - \sum_x \sum_y p(x,y) \log p(y) \\
&= \sum_x \sum_y p(x)p(y|x) \log p(y|x) - \sum_y p(y) \log p(y) \\
&= -\sum_x p(x) H(Y|X=x) + H(Y) \\
&= -H(Y|X) + H(Y)
\end{aligned}
$$

By symmetry of the above derivation, $I(X;Y) = H(X) - H(X|Y)$.

Exercise 2.16

According to definition of Beta distribution

$$
\mathrm{Beta}(x|a,b) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1} \;\; 0 \le x \le 1, \tag{11}
$$

where

$$
B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = \int_0^1 x^{a-1}(1-x)^{b-1} dx \tag{12}
$$

Mean:

$$
\begin{aligned}
E[X] &= \int_0^1 x \times \frac{x^{a-1}(1-x)^{b-1}}{B(a,b)}\, dx \\
&= \frac{1}{B(a,b)} \int_0^1 x^a (1-x)^{b-1}\, dx \\
&= \frac{B(a+1,b)}{B(a,b)} \qquad \text{from Eq. (12)} \\
&= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+b+1)} \\
&= \frac{\Gamma(a+b)\Gamma(a+1)}{\Gamma(a)\Gamma(a+b+1)} \\
&= \frac{\Gamma(a+b)\Gamma(a)a}{\Gamma(a)\Gamma(a+b)(a+b)} \\
&= \frac{a}{a+b},
\end{aligned}
$$

where we have used the property that $\Gamma(t+1) = t\Gamma(t)$.

To compute variance, we first compute

$$E[X^2] = \frac{B(a+2,b)}{B(a,b)} = \frac{\Gamma(a+b)\Gamma(a+2)\Gamma(b)}{\Gamma(a)\Gamma(b)\Gamma(a+b+2)} = \frac{a(a+1)}{(a+b)(a+b+1)}$$

Then

$$\begin{aligned}
\text{variance} &= E[X^2] - E^2[X] \\
&= \frac{a(a+1)}{(a+b)(a+b+1)} - \left(\frac{a}{a+b}\right)^2 \\
&= \frac{a(a+1)(a+b) - a^2(a+b+1)}{(a+b)^2(a+b+1)} \\
&= \frac{ab}{(a+b)^2(a+b+1)}
\end{aligned}$$

To obtain mode, we want:

$$\begin{aligned}
\operatorname*{argmax}_x \text{Beta}(x|a,b) &= \operatorname*{argmax}_x \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1} \\
&= \operatorname*{argmax}_x \log \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1} \\
&= \operatorname*{argmax}_x (a-1)\log x + (b-1)\log(1-x)
\end{aligned}$$

Differentiating with respect to x and set to 0, we get:

$$\begin{aligned}
&\frac{a-1}{x} - \frac{b-1}{1-x} = 0 \\
\implies &(a-1)(1-x) = (b-1)x \\
\implies &x = \frac{a-1}{a+b-2}
\end{aligned}$$

There are several cases here:

- When $a > 1$ and $b > 1$, the distribution is concave, and so mode is $\frac{a-1}{a+b-2}$.
- When $a = b = 1$, we have a uniform distribution, so the mode is any value between 0 and 1.
- When $a = b$ and both are less than 1, then we get a convex distribution symmetric about 0.5, so the modes are at 0 and 1.
- When $a > b$ and $b \leq 1$, then the distribution is convex and mode is 1.
- When $b > a$ and $a \leq 1$, then the distribution is convex and mode is 0.