HOMEWORK 1

Full name: Nguyễn Thị Minh Ngọc
Student ID: 11219280
Class: DSEB 63

<u>Ex1:</u>

a. The marginal distribution:

- $p(X = x_1) = 0.01 + 0.05 + 0.1 = 0.16$
- $p(X = x_2) = 0.02 + 0.1 + 0.05 = 0.17$
- $p(X = x_3) = 0.03 + 0.05 + 0.03 = 0.11$
- $p(X = x_4) = 0.1 + 0.07 + 0.05 = 0.22$
- $p(X = x_5) = 0.1 + 0.2 + 0.04 = 0.34$

- $p(Y = y_1) = 0.01 + 0.02 + 0.03 + 0.1 + 0.1 = 0.26$
- $p(Y = y_2) = 0.05 + 0.1 + 0.05 + 0.07 + 0.2 = 0.47$
- $p(Y = y_3) = 0.1 + 0.05 + 0.03 + 0.05 + 0.04 = 0.27$

b. The conditional distribution

- $p(X = x_1|Y = y_1) = \frac{0.01}{0.01+0.02+0.03+0.1+0.1} = \frac{1}{26} \approx 0.038$
- $p(X = x_2|Y = y_1) = \frac{0.02}{0.01+0.02+0.03+0.1+0.1} = \frac{1}{13} \approx 0.077$
- $p(X = x_3|Y = y_1) = \frac{0.03}{0.01+0.02+0.03+0.1+0.1} = \frac{3}{26} \approx 0.115$
- $p(X = x_4|Y = y_1) = \frac{0.1}{0.01+0.02+0.03+0.1+0.1} = \frac{5}{13} \approx 0.385$
- $p(X = x_5|Y = y_1) = \frac{0.1}{0.01+0.02+0.03+0.1+0.1} = \frac{5}{13} \approx 0.385$

- $p(Y = y_1|X = x_3) = \frac{0.03}{0.03+0.05+0.03} = \frac{3}{11} \approx 0.273$
- $p(Y = y_2|X = x_3) = \frac{0.05}{0.03+0.05+0.03} = \frac{5}{11} \approx 0.455$
- $p(Y = y_3|X = x_3) = \frac{0.03}{0.03+0.05+0.03} = \frac{3}{11} \approx 0.273$

<u>Ex2:</u>

* <u>Case 1</u>: X and Y are discrete random variables

$$E_x[X|Y] = \sum_{i=1}^{n} x_i P(X = x_i|Y)$$

$$
\begin{aligned}
E_y[E_x[X|Y]] &= \sum_{j=1}^{m} E_x\big[X|Y = y_j\big]P\big(X|Y = y_j\big) \\
&= \sum_{j=1}^{m} \Big(\sum_{i=1}^{n} x_i P\big(X = x_i|Y = y_j\big)\, P\big(X|Y = y_j\big)\Big) \\
&= \sum_{j=1}^{m} \sum_{i=1}^{n} x_i P\big(X = x_i|Y = y_j\big)\, P\big(X|Y = y_j\big) \\
&= \sum_{i=1}^{n} \sum_{j=1}^{m} x_i P\big(X = x_i|Y = y_j\big)\, P\big(X|Y = y_j\big) \\
&= \sum_{i=1}^{n} \sum_{j=1}^{m} x_i P\big(X = x_i, Y = y_j\big) \\
&= \sum_{i=1}^{n} x_i P(X = x_i) \\
&= E(X)
\end{aligned}
$$

Hence, $E(X) = E_y[E_x[X|Y]]$

* <u>Case 2</u>: X and Y are continuous random variables

$$
E_x[X|Y] = \int_{-\infty}^{+\infty} x f_{x|y}(x, y) dx
$$

$$
\begin{aligned}
E_y[E_x[X|Y]] &= \int_{-\infty}^{+\infty} (E_x[X|Y]) f_y(y) dy \\
&= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x f_{x|y}(x, y) dx) f_y(y) dy \\
&= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x (f_{x|y}(x, y)\, f_y(y)) dx dy \\
&= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x f_{x,y}(x, y)\, dx dy \\
&= \int_{-\infty}^{+\infty} x \int_{-\infty}^{+\infty} f_{x,y}(x, y)\, dy dx \\
&= \int_{-\infty}^{+\infty} x f_x(x)\, dx \\
&= E(X)
\end{aligned}
$$

Hence, $E(X) = E_y[E_x[X|Y]]$

<u>Ex3:</u>

* <u>Case 1</u>: X and Y are discrete random variables

$$
\begin{aligned}
V_x &= \sum_{i=1}^{n} (x_i - \mu)^2 P(X = x_i) \\
&= \sum_{i=1}^{n} (x_i - E_x(X))^2 P(X = x_i) \\
&= \sum_{i=1}^{n} ({x_i}^2 - 2x_i E_x(X) + E_x(X)^2) P(X = x_i) \\
&= \sum_{i=1}^{n} {x_i}^2 P(X = x_i) - \sum_{i=1}^{n} 2x_i E_x(X) P(X = x_i) + \sum_{i=1}^{n} E_x(X)^2 P(X = x_i) \\
&= E_x(X^2) - 2E_x(X) \sum_{i=1}^{n} x_i P(X = x_i) + E_x(X)^2 \sum_{i=1}^{n} P(X = x_i) \\
&= E_x(X^2) - 2E_x(X)E_x(X) + E_x(X)^2 \\
&= E_x(X^2) - E_x(X)^2
\end{aligned}
$$

Hence, $V_x(X) = E_x(X^2) - E_x(X)^2$

* <u>Case 2</u>: X and Y are continuous random variables

$$
\begin{aligned}
V_x &= \int_{-\infty}^{+\infty} (x - \mu)^2 f(x)\, dx \\
&= \int_{-\infty}^{+\infty} (x^2 - 2x\mu + \mu^2) f(x)\, dx \\
&= \int_{-\infty}^{+\infty} x^2 f(x)\, dx - \int_{-\infty}^{+\infty} 2x\mu f(x)\, dx + \int_{-\infty}^{+\infty} \mu^2 f(x)\, dx \\
&= E_x(X^2) - (2\mu \int_{-\infty}^{+\infty} x f(x)\, dx) + (\mu^2 \int_{-\infty}^{+\infty} f(x)\, dx) \\
&= E_x(X^2) - 2\mu^2 + \mu^2 \\
&= E_x(X^2) - E_x(X)^2
\end{aligned}
$$

Hence, $V_x(X) = E_x(X^2) - E_x(X)^2$

<u>Ex4:</u>

X: Cancerous Tumors → X^c: Benign Tumors
Y: Positive Mammogram Result → Y^c: Negative Mammogram Result

$P(x) = 0.01 \rightarrow P(X^c) = 0.99$
$P(Y|X) = 0.8 \rightarrow P(Y^c|X) = 0.2$
$P(Y^c|X^c) = 0.9 \rightarrow P(Y|X^c) = 0.1$

$P(X|Y) = ?$

According to Baye's Rule:

$$\begin{aligned} P(X|Y) &= \frac{P(Y|X)P(X)}{P(Y)} \\ &= \frac{P(Y|X)P(X)}{P(Y,X)+P(Y,X^c)} \\ &= \frac{P(Y|X)P(X)}{P(Y|X)P(X)+P(Y|X^c)P(X^c)} \\ &= \frac{0.8*0.01}{0.8*0.01+0.1*0.99} = \frac{8}{107} \approx 0.075 \approx 7.5\% \end{aligned}$$

The probability of cancer of this women after receiving positive mammogram result is about 7.5%. Hence, I do not agree with the estimated result of those physicians.

<u>Ex5:</u>

X: Wrong first pick → X^c: Right first pick
Y: Win when switching → Y^c: Lose when switching

We consider the probability of win if I switch under two cases.

- <u>Case 1</u>: My initial pick was wrong
  - The probability that my first pick was wrong is $P(X) = \frac{3}{4}$
  - So, in case that my first pick was wrong, the right door will be one of two unopened door, then the probability that I will win when switching is $P(Y|X) = \frac{1}{2}$

→ The probability of win if my first pick was wrong and I choose to switch when being asked is:

$$\mathrm{P(X,Y) = P(Y|X) * P(X)} = \frac{1}{2} * \frac{3}{4} = \frac{3}{8}$$

- <u>Case 2</u>: My initial pick was right
  - The probability that my first pick was right is $P(X) = \frac{1}{4}$
  - So, in case that my first pick was right, I have no chance to win when switching, then the probability that I will win when switching is $P(Y|X^c) = 0$

→ In case of right first pick, the probability of win if I choose to switch is:

$$\mathrm{P}(X^c, \mathrm{Y}) = \mathrm{P}(\mathrm{Y}|X^c) * \mathrm{P}(X^c) = 0 * \frac{1}{4} = 0$$

Hence, the probability of win when switching is:

$$P(Y) = P(Y,X) + P(Y,X^c) = \frac{3}{8} + 0 = \frac{3}{8}$$