Independent random variables

- Definition. Random variables X_1, ..., X_k are independent means

  $$P(X_1 \in A_1 \text{ and } X_2 \in A_2 \text{ and } \cdots \text{ and } X_k \in A_k) = P(X_1 \in A_1)P(X_2 \in A_2) \cdots P(X_k \in A_k)$$

  for all $A_1, A_2, \ldots, A_k \in \mathcal{B}(R)$.

- We say X_1, ..., X_k are dependent if X_1, ..., X_k are not independent.

- Fact 1 X_1, ..., X_k are independent if and only if

  $$P(X_1 \leq x_1 \text{ and } X_2 \leq x_2 \text{ and } \cdots \text{ and } X_k \leq x_k) = P(X_1 \leq x_1)P(X_2 \leq x_2) \cdots P(X_k \leq x_k)$$

  for all $x_1, \ldots, x_k \in R$.

- Fact 2 Suppose that $(X_1, \ldots, X_k)$ is a discrete random vector, X_1, ..., X_k are independent if and only if

  $$p_{X_1,\ldots,X_k}(x_1, \ldots, x_k) = p_{X_1}(x_1) \cdots p_{X_k}(x_k)$$

  for all $x_1, \ldots, x_k \in R$.

- Fact 3 Suppose that $(X_1, \ldots, X_k)$ is has a joint PDF $f_{X_1,\ldots,X_k}$. Define

  $$g(x_1, \ldots, x_k) = f_{X_1}(x_1) \cdots f_{X_k}(x_k)$$

  for all $x_1, \ldots, x_k \in (-\infty, \infty)$, where f_{X_1}, ..., f_{X_k} are PDFs of X_1, ..., X_k respectively. Then g can be a PDF of $(X_1, \ldots, X_k)$ if and only if X_1, ..., X_k are independent.

- Remark. Suppose that X, Y are random variables and (X, Y) has joint PDF. Let f_X be a PDF of X and f_Y is a PDF of Y. Let $R_Y = \{y : f_Y(y) > 0\}$. If X and Y are independent, then $\{f_X : y \in R_Y\}$ is a version of the conditional PDF of X given Y.

- The following result can be used together with Fact 3 to show that random variables are not independent.

  Fact 4 Suppose that g and $f_{X_1,\ldots,X_k}$ are both PDFs of $(X_1, \ldots, X_k)$, and g and $f_{X_1,\ldots,X_k}$ are continuous on an open set A, then $g(x_1, \ldots, x_k) = f_{X_1,\ldots,X_k}(x_1, \ldots, x_k)$ for every $(x_1, \ldots, x_k) \in A$.

- Example 1. Suppose that (X, Y) has a joint PDF $f_{X,Y}$, where

  $$f_{X,Y}(x, y) = \begin{cases} 1.5x + 0.5y & \text{if } (x, y) \in (0, 1) \times (0, 1); \\ 0 & \text{otherwise.} \end{cases}$$

  Show that X and Y are not independent.

Sol. Compute $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$ and $f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx$, then we have

$$f_X(x) = \begin{cases} 1.5x + 0.25 & \text{if } x \in (0,1); \\ 0 & \text{otherwise} \end{cases}$$

and

$$f_Y(y) = \begin{cases} 0.75 + 0.5y & \text{if } y \in (0,1); \\ 0 & \text{otherwise} \end{cases}$$

Let $g(x,y) = f_X(x)f_Y(y)$ for $(x,y) \in R^2$. Then g and $f_{X,Y}$ are both continuous on $(0,1) \times (0,1)$. If g is a PDF of (X,Y), then by Fact 4, $g = f_{X,Y}$ on $(0,1) \times (0,1)$. It is clear that on $(0,1) \times (0,1)$,

$$g(x,y) = (1.5x + 0.25)(0.75 + 0.5y) \neq 1.5x + 0.5y = f_{X,Y}(x,y)$$

for some $(x,y) \in (0,1) \times (0,1)$, so g cannot be a PDF of (X,Y), which implies that X and Y are not independent by Fact 3.

Below is another way to deduce that X and Y are not independent based on the $f_{X,Y}$, f_X and f_Y computed above. Suppose that X and Y are independent, then

$$P(X \leq s \text{ and } Y \leq t) = P(X \leq s)P(X \leq t) \text{ for } (s,t) \in (0,1) \times (0,1), \tag{1}$$

whhich implies that for $(s,t) \in (0,1) \times (0,1)$,

$$\int_0^s \int_0^t (1.5x + 0.5y)dydx = \int_0^s (1.5x + 0.25)dx \int_0^t (0.5y + 0.75)dy.$$

Take the partial derivative $\partial^2/\partial t \partial s$ at both sides of the above equation and we have

$$(1.5s + 0.5t) = (1.5s + 0.25)(0.5t + 0.75)$$

for all $(s,t) \in (0,1) \times (0,1)$, which is clearly not true. Therefore, (1) does not hold and X and Y are not independent.

- Example 2. Suppose that (X,Y) has a joint PDF $f_{X,Y}$, where $f_{X,Y}(x,y) = g(x)h(y)$ for $(x,y) \in R^2$ for some functions g and h. Show that X and Y are independent.

  The proof is left as an exercise.

- Example 3. Suppose that (X,Y) is a discrete random vector with joint PMF $p_{X,Y}$, where

$$p_{X,Y}(x,y) = \begin{cases} 0.5 & \text{if } (x,y) = (1,2); \\ 0.3 & \text{if } (x,y) = (2,3); \\ 0.2 & \text{if } (x,y) = (1,5); \\ 0 & \text{otherwise.} \end{cases}$$

  Determine whether X and Y are independent.

  Ans. No since $p_{X,Y}(2,5) = 0 \neq p_X(2)p_Y(5) = 0.3 \cdot 0.2$.

- Fact 5 Suppose that X and Y are independent and $E(u(X))$ and $E(v(Y))$ are finite, then
$$E(u(X)v(Y)) = E(u(X))E(v(Y))$$

- Fact 6 Suppose that (X, Y) has joint MGF $M_{X,Y}$, where $M_{X,Y}(t_1, t_2)$ is finite for $|t_1| < \delta$ and $|t_2| < \delta$ for some $\delta > 0$. Then X and Y are independent if and only if
$$M_{X,Y}(t_1, t_2) = M_{X,Y}(t_1, 0)M_{X,Y}(0, t_2)$$
for $|t_1| < \delta$ and $|t_2| < \delta$.

- Example 4. For $\mu \in (-\infty, \infty)$ and $\sigma > 0$, define
$$f_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2)}$$
for $x \in (-\infty, \infty)$. It can be shown that $\int_{-\infty}^{\infty} f_{\mu,\sigma}(x)dx = 1$ (see the remark after Homework problem 35 for example). Suppose that Z is a random variable with PDF $f_{\mu,\sigma}$. Suppose that X and Y are independent random variables with PDFs $f_{1,3}$ and $f_{2,4}$ respectively.

  (a) Find the MGF of Z.
  (b) Find the joint MGF of (X, Y).
  (c) Find the MGF of $X - Y$.
  (d) Find a PDF of $X - Y$.

  A sketch of solution.

  (a) Let M_Z be the MGF of Z, then from the solution to Problem 30(b), we have
  $$M_Z(t) = E(e^{tZ}) = \int_{-\infty}^{\infty} e^{tz} f_{\mu,\sigma}(z)dz = e^{\mu t + 0.5\sigma^2 t^2}$$
  for $t \in (-\infty, \infty)$.
  (b) Let $M_{X,Y}$ be the MGF of (X, Y), then by independence,
  $$M_{X,Y}(t_1, t_2) = M_X(t_1)M_Y(t_2) = e^{t_1 + 4.5t_1^2} e^{2t_2 + 8t_2^2}$$
  for $(t_1, t_2) \in R^2$.
  (c) Let M be the MGF of $X - Y$, then
  $$M(t) = M_{X,Y}(t, -t) = e^{t+4.5t^2} e^{-2t+8(-t)^2} = e^{-t-12.5t^2}$$
  for $t \in (-\infty, \infty)$.
  (d) Since the MGF of $X - Y$ is the same as the MGF of Z when $\mu = -1$ and $\sigma = 5$ and $f_{-1,5}$ is a PDF of Z when $\mu = -1$ and $\sigma = 5$, $f_{-1,5}$ is a PDF of $X - Y$.

- The distribution of Z in Example 4 is called the normal distribution with mean μ and variance σ^2, denoted by $N(\mu, \sigma^2)$.