Independent Random Variables

Defn:

$X_1, X_2, \ldots X_n$ are said to be independent random variables if

$$F(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} F_{X_i}(x_i) \quad \forall\, x_1, x_2, \ldots x_n \in \mathbb{R}$$

$$\Updownarrow$$

$$f(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)$$

\# $X_1, X_2, \ldots X_n$ are said to be pairwise independent if for any i, j $i \neq j$, X_i & X_j are independent.

Independence $\Longrightarrow$ Pairwise independent
$\not\Longleftarrow$

\# If $\underline{X_1, X_2, \ldots X_n}$ are independent

If $X_1, X_2, \ldots X_n$ are independent then any sub collection of RVs $X_{i_1}, \ldots, X_{i_k}$ is also independent.

$$\boxed{f(x_{i_1}, \ldots x_{i_k}) = \prod_{j=1}^{k} f_{X_{i_j}}(x_{i_j})}$$

$$\iint f(x_1) \ldots f(x_n)\, dx_1 \ldots dx_n = 1$$

Def^n: A sequence $\{X_n\}_{n=1}^{\infty}$ of random variables is said to be independent if for any n, $X_1, X_2, \ldots X_n$ are independent RVs.

X & Y are said to be identically distributed if

$$F_X(x) = F_Y(x) \quad \forall\, x \in \mathbb{R}$$

$$\Updownarrow$$

$$F_X(x) = F_Y(x) \quad \forall\ x \in \mathbb{R}$$

$$\Updownarrow$$

$$f_X(x) = f_Y(x) \quad \forall\ x \in \mathbb{R}$$

does not mean that $X = Y$.

The random variables which are independent & identically distributed are called as <u>iid</u> Random variables.

X & Y are identically distributed

$$P\{X \in B\} = P\{Y \in B\} \quad \forall\ B \in \mathcal{B}.$$

$$\not\Downarrow$$

$$\boxed{X = Y} \Rightarrow \boxed{X(\omega) = Y(\omega)}$$

$Var(X) \rightarrow$ variability in X w.r.t. $E[X]$.

$Cov(X,Y) \rightarrow$ joint variation between X & Y.

$$Cov(X,Y) = E\left[(X-E[X])(Y-E[Y])\right]$$

$$= E\left[XY - Y \cdot \underline{E[X]} - X \cdot E[Y] + \underline{E[X]E[Y]}\right]$$

$$Cov(X,Y) = \underline{E[XY]} - \underline{E[X]E[Y]}$$

$$Cov(X,X) = Var(X)$$

\# $Cov(aX, Y) = a\, Cov(X,Y)$

\# $Cov(X+Y, Z) = Cov(X,Z) + Cov(Y,Z)$.

\# $Cov\left(\sum_{i=1}^{n} X_i, \underline{Y}\right) = \sum_{i=1}^{n} Cov(X_i, Y)$

\# $Cov\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\right) = \sum_{i=1}^{n} \sum_{j=1}^{m} Cov(X_i, Y_j)$

\# ✓ $\underline{Cov\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{n} X_j\right) = \sum_{i=1}^{n} \sum_{j=1}^{n} Cov(X_i, X_j)}$

$$Var\left(\sum_{i=1}^{n} X_i\right) = \underline{\sum_{i=1}^{n} Var(X_i)} + \underline{\sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} Cov(X_i, X_j)}$$

$\mathrm{Cov}(X_i, X_j) = \mathrm{Cov}(X_j, X_i)$

If X & Y are idependent

$$\mathrm{Cov}(X,Y) = E[XY] - E[X]\cdot E[Y]$$
$$= E[X]E[Y] - E[X]\cdot E[Y] = 0$$

If $X_1, X_2, \ldots X_n$ are independent RVs

$$\mathrm{Cov}(X_i, X_j) = 0 \quad \forall \; i, j, \; i \neq j$$

$$\Rightarrow \quad \mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i)$$

$$\mathrm{Cov}(X,Y) = E\left[\underline{(X - E[X])(Y - E(Y))}\right]$$

$\underline{(X - E[Y])(Y - E[Y]) \geq 0}$ if $\boxed{\begin{array}{l} X \geq E[X], \; Y \geq E[Y] \\ X \leq E[X], \; Y \leq E[Y] \end{array}}$

① — (X, Y) increases together or decreases together

$\underline{(X - E[Y])(Y - E[Y]) \leq 0}$ if $\boxed{\begin{array}{l} X \geq E[X], \; Y \leq E[Y] \\ \text{or} \\ X \leq E[X], \; Y \geq E[Y] \end{array}}$

② — $\boxed{\text{when } X \text{ increases } Y \text{ decreases or } X \text{ dec. the } X \text{ inc.}}$

$E[(X-E[X])(Y-E[Y])] \geq 0$ if ① happens more often.

$\Downarrow$

(X,Y) are positively correlated

$E[(X-E(X))(Y-E(Y))] \leq 0 \Rightarrow$ ② happens more often.

$\Downarrow$

X, Y are negatively correlated.

$\underline{Cov(X,Y)}$ can be positive or negative

Correlation Coefficient:

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sigma_x \cdot \sigma_y}$$

$\sigma_x \rightarrow$ standard deviation of X

$\sigma_y \rightarrow$ std. dev. of Y

Claim: $-1 \leq \rho(X,Y) \leq 1$

Claim: $-1 \leq \rho(X,Y) \leq 1$

Proof:
$$\tilde{X} = X - E[X]$$
$$\tilde{Y} = Y - E[Y]$$

$$E\left[\left(\tilde{X} - \frac{E[\tilde{X}\tilde{Y}]}{E[\tilde{Y}^2]}\tilde{Y}\right)^2\right] \geq 0$$

Complete the proof. (Exercise)

Function of several Random variables

Let $(\Omega, \mathcal{F}, P)$ be a Prob. space, and $X_1, X_2, \dots X_n$ are random variables defined on $(\Omega, \mathcal{F}, P)$.

$g : \mathbb{R}^n \to \mathbb{R}$ is Borel measurable

$g^{-1}(B) \in \mathcal{B}_n$ → Borel σ-field on $\mathbb{R}^n$. $\forall B \in \mathcal{B}$

$\|$

$\{ y \mid g(y) \in B \}$

$$\{ y \mid g(y) \in B \}$$

Then $g(X_1, X_2, \ldots X_n)$ is a Random variable.

[e.g. $g(x_1, \ldots x_n) = \sum_{i=1}^{n} x_i$]

What is the Probability distribution of $g(X_1, X_2, \ldots X_n)$?

Let $Y = g(X_1, \ldots X_n)$

$$P\{Y \le y\} = P\{ g(X_1, \ldots X_n) \le y \}$$

$$= P\{ (X_1, \ldots X_n) \in \underline{g^{-1}((-\infty, y])} \}$$

$$P\{Y \le y\} = \sum_{\{(x_1, \ldots x_n) \mid g(x_1, \ldots x_n) \le y\}} \underline{P\{X_1 = x_1, \ldots X_n = x_n\}}$$

(discrete)

$$= \int \cdots \int_{\{(x_1, \ldots x_n) \mid g(x_1, \ldots x_n) \le y\}} f(x_1, \ldots x_n)\, dx_1, \ldots dx_n$$

(continuous)