May 26

Covariance: Given 2 r.v. X, Y the covariance between X and Y is defined as $\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$

Properties:

1. $\text{Cov}(X, X) = V(X)$ ← variance of X
2. If c is a constant, then $\text{Cov}(X+c, Y) = \text{Cov}(X, Y)$
3. X, Y, Z are random variables
   $\text{Cov}(X+Y, Z) = \text{Cov}(X, Z) + \text{Cov}(Y, Z)$
4. X, Y are r.v. and $\alpha \in \mathbb{R}$, then $\text{Cov}(\alpha X, Y) = \alpha \, \text{Cov}(X, Y)$
5. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$

Remark (about independence):

1. If $\text{Cov}(X, Y) = 0$, then we say that X and Y are uncorrelated
2. Independent r.v. are uncorrelated; However uncorrelated r.v. are not necessarily independent.

Ex:

1.

$y \backslash x$	-1	0	1	$P_Y(y)$
-1	$1/4$	0	$1/4$	$1/2$
0	0	0	0	0
1	$1/4$	0	$1/4$	$1/2$
$P_X(x)$	$1/2$	0	$1/2$	

$\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$

$\mathbb{E}(XY) = 0 = \mathbb{E}(X) = \mathbb{E}(X)$

X and Y are uncorrelated, but X and Y are not independent

②

$y \backslash x$	-1	0	1	
-1	0	1/4	0	1/4
0	1/4	0	1/4	1/2
1	0	1/4	0	1/2
	1/4	1/2	1/4	

$\mathbb{E}(X) = \mathbb{E}(Y) = 0$

$\mathbb{E}(XY) = 0$

Cov $(X,Y) = 0$, but X and Y are not independent.

③ $f(x,y) = \begin{cases} 6x & \text{if } 0<x<y<1 \\ 0 & \text{otherwise} \end{cases}$

$X \sim \text{Beta}(2,2) \longrightarrow \mathbb{E}(X) = \frac{2}{2+2} = \frac{1}{2}$, $\mathbb{E}(Y) = \frac{3}{1+3} = \frac{3}{4}$

$Y \sim \text{Beta}(3,4)$

$\mathbb{E}(XY) = \int_0^1 \left(\int_0^y xy \underbrace{6x}_{\text{pdf}} \, dx \right) dy = \int_0^1 2y^4 \, dy = \frac{2}{5}$

$\text{Cov}(X,Y) = \frac{2}{5} - \frac{3}{8} = \frac{1}{40}$

$\overrightarrow{u} \cdot \overrightarrow{v} = \|\vec{u}\| \, \|\vec{v}\| \cos(\theta)$

$\text{Cov}(X,Y) = \sqrt{V(X)} \sqrt{V(Y)} \cdot \rho(X,X)$

ρ: correlation coefficient

strength of linear association btw X & Y: $\boxed{-1 \le \rho(X,Y) \le 1}$

Proposition:

① $|\text{Cov}(X,Y)| \le \sqrt{V(X)} \sqrt{V(Y)}$ (Cauchy-Schwarz Inequality)

$|\vec{u} \cdot \vec{v}| \le \|\vec{u}\| \, \|\vec{v}\|$

② The correlation coefficient defined as:

$\rho(X,Y) = \dfrac{\text{Cov}(X,Y)}{\sqrt{V(X)} \sqrt{Y}}$ is such that $-1 \le \rho(X,Y) \le 1$

③ $\rho(X,Y) = 1 \longrightarrow Y = \alpha X + \beta$ where $\alpha > 0$ (Linear perfect positive assoc. btw. X & Y)

$\rho(X,Y) = -1 \rightarrow Y = \alpha X + \beta$ where $\alpha < 0$

Ex: Given that $\text{Cov}(X,Y) = -2$. $\mathbb{E}(X) = 1$, $V(X) = 4$ and $\mathbb{E}(Y) = -3$.

Find the relationship between X and Y.

$\rightarrow \rho(X,Y) = \dfrac{Cov(X,Y)}{\sqrt{V(X)}\sqrt{V(Y)}} = \dfrac{-2}{1\times 2} = -1 \rightarrow Y = \alpha X + \beta \Rightarrow -2 = Cov(X, X)$

insensitive to constant

$= Cov(X, \alpha X + \beta) = Cov(X, \alpha X) = \alpha\, Cov(X,X) = \alpha V(X).$

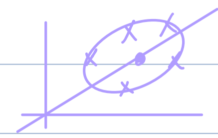
$\alpha = \dfrac{-2}{V(X)} = -2 \; . \quad Y = -2X + \beta$

$-3 = \mathbb{E}(Y) = \mathbb{E}(-2X+\beta) = -2\mathbb{E}(X) + \beta \rightarrow \beta = -3 + 2\mathbb{E}(X) = -1$

relationship: $Y = -2X + 1$ (perfectly linear assoc.)

$(X,Y) \rightarrow (x_i, y_i)$

↳ collect data points plot scatter plot



Application: Variance of linear combination of r.v.: let $X_1, \cdots, X_n$ be n r.v. and $\alpha_1, \cdots, \alpha_n$ $\beta_1, \cdots, \beta_n$ be scalars.

① $Cov\left(\sum_{i=1}^{n} \alpha_i X_i, \sum_{j=1}^{n} \beta_j X_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \beta_j Cov(X_i, X_j).$

In particular $\Rightarrow$

② $V\left(\sum_{i=1}^{n} \alpha_i X_i\right) = V\left(\sum_{i=1}^{n} \alpha_i X_i, \sum_{j=1}^{n} \beta_j X_j\right) = \sum_{i=1}^{n} \alpha_i^2 V(X_i) + 2\sum_{i<j} \alpha_i \alpha_j Cov(X_i, X_j)$

Ex: $V(X) = 2$. $V(X) = 3$. $Cov(X,X) = 1$,

then $V(3X+2Y) = 9V(X) + 4V(Y) = 2\times 6\, Cov(X,Y)$

↳ $Cov(3X+2Y, 3X+2Y) = Cov(3X, 3X+2X) + Cov(2Y, 3X+2X)$

$= Cov(3X,3X) + Cov(3X,2X) + Cov(2X,3X) + Cov(2X,2X)$

$= 18 + 12 + 12 = 42$

(Bi)

Multinomial Distributions

Discovery Ex: An urn contains 3 red balls, 2 blue ball, and 1 green ball. Draw successively with replacement 5 balls from the urn.

X_1 = # of red $\quad X_2$ = # of blue $\quad X_3$ = # of green

$\rightarrow X_1 \sim \text{Binomial}(n=5, p_1 = \frac{1}{2})$ $X_2 \sim \text{Bin}(n=5, p_2 = \frac{1}{3})$ $X_3 \sim \text{Bin}(n=5, p=\frac{1}{6})$

$X_1 + X_2 + X_3 = 5$

↓

What is the joint probability function of X_1, X_2, X_3 :

$P(X_1 = k_1, X_2 = k_2, X_3 = k_3)$ $\boxed{\# k_1 + k_2 + k_3 = 5}$

$$P(X_1 = k_1, X_2 = k_2, X_3 = k_3) = C^{n}_{k_1} \times C^{n-k_1}_{k_2} \times C^{n-k_1 \cdot k_2}_{k_3} \; p_1^{k_1} p_2^{k_2} p_3^{k_3}$$

$$C^{n}_{k_1} C^{n-k_1}_{k_2} C^{n-k_1-k_2}_{k_3} = \frac{n!}{k_1!(n-k_1)!} \frac{(n-k_1)!}{k_2!(n-k_2-k_1)!}$$