# Covariance & Corrlation

1. Covariance: $Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] = E[XY] - \mu_X \cdot \mu_Y$.

$Var(X+Y) = Var(X) + Var(Y) + 2\,Cov(X,Y)$.

1). $Cov(X,Y) = Cov(Y,X)$.

2). When $Cov(X,Y) > 0$, then $X, Y$ tend to increase together

When $Cov(X,Y) < 0$, then $X, Y$ tend to increase oppositely.

3). If $Y = X$. $Cov(X,X) = E[(X-\mu_X)^2] = Var(X) \geq 0$.

If $Y = -X$. $Cov(X,-X) = E[-(X-\mu_X)^2] = -Var(X) \leq 0$.

4). If $Y = cX$. $Var(Y) = c^2 Var(X)$, which $Sd(X)Sd(Y) = |c| \cdot Var(X)$.

$Cov(X,Y) = Cov(X,cX) = c\,Cov(X,X) = c\,Var(X)$.

If $c \geq 0$, $Cov(X,Y) = Sd(X)\,Sd(Y)$.

If $c < 0$, $Cov(X,Y) = -Sd(X)\,Sd(Y)$.

$-\|u\|\cdot\|v\| \leq u \cdot v \leq \|u\|\cdot\|v\|$.

5). $-Sd(X)\,Sd(Y) \leq Cov(X,Y) \leq Sd(X)\,Sd(Y)$.

$\Leftrightarrow -\sqrt{Var(X)Var(Y)} \leq Cov(X,Y) \leq \sqrt{Var(X)Var(Y)}$.

6). $Cov(aX+bY, Z) = a\,Cov(X,Z) + b\,Cov(Y,Z)$. (bilinear).

$E(XY) = E(X) \cdot E(Y)$

7). If $X, Y$ are independent, $Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]$

converse is false.

$= E(X-\mu_X) \cdot E(Y-\mu_Y)$.

$= (E[X]-\mu_X)(E[Y]-\mu_Y)$.

$= (\mu_X-\mu_X)(\mu_Y-\mu_Y) = 0$.

when $Corr(X,Y) = 0$, it's uncorrelated

2. Correlation. $Corr(X,Y) = \dfrac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$, $-1 \leq Corr(X,Y) \leq 1$.

$= \dfrac{Cov(X,Y)}{Sd(X) \cdot Sd(Y)}$, only defined when $Var(X), Var(Y) > 0$.

non-constant.

1). If $Z = cY, c > 0$, $Corr(X,Y) = Corr(X,Z)$.

$c < 0$, $Corr(X,Y) = -Corr(X,Z)$.

2). $Corr(X,X) = \dfrac{Cov(X,X)}{Sd(X) \cdot Sd(X)} = \dfrac{Var(X)}{Var(X)} = 1$.

3. Example: $p_{X,Y}(5,1) = p_{X,Y}(5,9) = p_{X,Y}(7,3) = p_{X,Y}(7,7) = \frac{1}{4}$, otherwise

0. Solve $Cov(X,Y)$.

$\mu_X = E[X] = \sum_{x \in \mathbb{R}} x \cdot p_X(x) = \sum_{x,y \in \mathbb{R}} x \cdot p_{X,Y}(x,y) = 5\times\frac{1}{4} + 5\times\frac{1}{4} + 7\times\frac{1}{4} + 7\times\frac{1}{4}$

$= 6.$

$\mu_Y = E[Y] = \sum_{y \in \mathbb{R}} y \cdot p_Y(y) = \sum_{x,y \in \mathbb{R}} y \cdot p_{X,Y}(x,y) = 1\times\frac{1}{4} + 9\times\frac{1}{4} + 3\times\frac{1}{4} + 7\times\frac{1}{4}$

$= 5.$

$E[XY] = \sum_{x,y \in \mathbb{R}} x \cdot y \cdot p_{X,Y}(x,y) = 5\times1\times\frac{1}{4} + 5\times9\times\frac{1}{4} + 7\times3\times\frac{1}{4} + 7\times7\times\frac{1}{4} = 30.$

$Cov(X,Y) = E[XY] - \mu_X \cdot \mu_Y = 30 - 30 = 0.$

Also, since $Cov(X,Y) = 0$, gives $Corr(X,Y) = 0$ and.

However, $p_X(5) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2} > 0$, $p_Y(3) = \frac{1}{4} > 0$. but $p_{X,Y}(5,3) = 0$

$\neq p_X(5) \cdot p_Y(3)$.

$\therefore$ $X, Y$ are not independent.