Probability Theory

Covariance

Given two random variables, X and Y, the **covariance** between them is a measure of how they depend on one another. A __________________________ covariance means that large values of X are typically accompanied by large values of Y, while a ______________________ covariance means that large values of X are typically accompanied my small values of Y. If the covariance is 0, then X and Y do not have a linear relationship.

<u>Def:</u> The covariance of X and Y is given by

$$Cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] = \sum_y \sum_x (x - \mu_X)(y - \mu_Y).$$

Similar to variance, we have a short-cut version also:

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

We know from before that $E[X + Y] = E[X] + E[Y]$, but in general $Var(X + Y) \neq Var(X) + Var(Y)$. Observe that

$$\begin{aligned}Var(X+Y) &= E[(X+Y)^2] - E[X+Y]^2 = E[X^2+2XY+Y^2] - (\mu_X+\mu_Y)^2 \\ &= E[X^2] - \mu_X^2 + 2E[XY] - 2\mu_X\mu_Y + E[Y^2] - \mu_Y^2 \\ &= Var(X) + Var(Y) + 2Cov(X,Y).\end{aligned}$$

If X and Y are independent, then $E[XY] = E[X] \cdot E[Y]$. Thus if X and Y are independent, then $Cov(X,Y) =$ __________, and $Var(X+Y) =$ ______________________.

The converse is not true in general! X and Y can be dependent and still have zero covariance.

Def: The **correlation** between X and Y is the strength of the linear association between them given by

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{Cov(X,Y)}{\sigma_X\sigma_Y}.$$

The correlation is always between -1 (a perfect negative linear association) and 1 (a perfect positive linear association).

Properties of covariance and variance:

1. $Cov(X,X) = Var(X)$.
2. $Cov(X,Y) = Cov(Y,X)$.
3. $Cov(aX,bY) = abCov(X,Y)$.
4. $Var(X+a) = Var(X)$.

5. $Var(aX + bY) = a^2Var(X) + b^2Var(Y) + 2abCov(X, Y)$.

6. (bilinearity) $Cov(X, aY + bZ) = aCov(X, Y) + bCov(X, Z)$.

<u>**Ex:**</u> X and Y have joint probability distribution given by:

X \ Y	1	2	3
1	0.40	0.12	0.08
2	0.30	0.06	0.04

Find the covariance and the correlation between X and Y.

<u>**Ex:**</u> The profit for a new product is given by $Z = 3X - Y - 5$ where X and Y are independent, and $Var(X) = 1$, $Var(Y) = 2$. Calculate $Var(Z)$.

Ex: The profit for a new product is given by $Z = 5X - 2Y - 5$ where X and Y are such that $Var(X) = 1$, $Var(Y) = 2$, and $Cov(X, Y) = -1$. Calculate $Var(Z)$.

Ex: X and Y have joint probability distribution given by:

$P(X = 1, Y = 1) = 0.10$

$P(X = 1, Y = 2) = 0.15$

$P(X = 2, Y = 1) = 0.18$

$P(X = 3, Y = 1) = 0.22$

$P(X = 3, Y = 2) = 0.05$

$P(X = 3, Y = 3) = 0.30$

Find the covariance between X and Y.

Ex: X and Y have joint probability distribution given by:

$Y \backslash X$	-1	0	1
0	0.1	0.1	0.2
1	0.1	0.3	0.2

Find the covariance between X and Y.

Ex: X and Y have joint probability distribution given by:

		x			
		2	3	4	5
	0	0.05	0.05	0.15	0.05
y	1	0.40	0	0	0
	2	0.05	0.15	0.10	0

For this distribution, $E[X] = 2.85$ and $E[Y] = 1$. Find $Cov(X, Y)$.

Ex: Let X and Y be random variables with $Var(X) = 4$, $Var(Y) = 9$, and $Var(X-Y) = 16$. Determine $Cov(X, Y)$.

Ex: An insurance policy pays a total medical benefit consisting of two parts for each claim. Let X be the part paid to the surgeon, and let Y be the part paid to the hospital. Suppose $Var(X) = 5,000$, $Var(Y) = 10,000$, and $Var(X+Y) = 17,000$. If the company that issues the policay increases the X amount by a flat amount of 100 per claim and the Y amount by 10% per claim, find the new variance for the total benefit.