Lecture 12

Covariance
Correlation Coefficient

Index

- Two random variables
- Expected value of a function
- Covariance
- Correlation Coefficient
- Questions

Question

Each front tire on a particular type of vehicle is supposed to be filled to a pressure of 26 psi. Suppose the actual air pressure in each tire is a random variable X for the right tire and Y for the left tire with the joint pdf

$$f(x, y) = \begin{cases} K(x^2 + y^2) & 20 \le x \le 30, 20 \le y \le 30 \\ 0 & \text{otherwise} \end{cases}$$

a. What is the value of K?
b. What is the probability that both tires are underfilled?
c. Determine the (marginal) distribution of air pressure in the right tire alone.
d. Are X and Y independent rv’s?

$$f(x,y)=\begin{cases} K\left(x^2+y^2\right) & 20\le x\le 30 \\ 0 & \text{Otherwise}\end{cases}$$

$$\int_x\int_y f(x,y)\,dydx=1$$

$$\int_x\left(\int_y f(x,y)\,dy\right)dx=1$$

$$K\int_x\left(\int_y\left(x^2+y^2\right)dy\right)dx=1$$

$$K\int_x\left(\left(x^2y+\frac{y^3}{3}\right)_{20}^{30}\right)dx=1$$

$$K\int_x\left(\left(x^2\left(30-20\right)+\frac{30^3-20^3}{3}\right)\right)dx=1$$

$$K\int_x\left(10x^2+\frac{19000}{3}\right)dx=1$$

$$K\left(10\frac{x^3}{3}+\frac{19000}{3}x\right)_{20}^{30}=1$$

$$K\left(10\frac{30^3-20^3}{3}+\frac{19000}{3}(30-20)\right)=1$$

$$K\left(10\frac{19000}{3}+\frac{19000}{3}(10)\right)=1$$

$$\frac{380000}{3}K=1$$

$$K=\frac{3}{380000}$$

$$P\left(X\le 26,Y\le 26\right)=\int_{20}^{26}\int_{20}^{26}K\left(x^2+y^2\right)dydx$$

$$=K\int_x\left(\int_y\left(x^2+y^2\right)dy\right)dx$$

$$=K\int_x\left(\left(x^2y+\frac{y^3}{3}\right)_{20}^{26}\right)dx$$

$$=K\int_x\left(\left(x^2\left(26-20\right)+\frac{26^3-20^3}{3}\right)\right)dx$$

$$=K\int_x\left(6x^2+\frac{9576}{3}\right)dx$$

$$=K\left(6\frac{x^3}{3}+\frac{9576}{3}x\right)_{20}^{26}$$

$$=K\left(6\frac{26^3-20^3}{3}+\frac{9576}{3}(26-20)\right)$$

$$=K\left(2\left(9576\right)+\frac{19000}{3}(6)\right)$$

$$=\left(38304\right)K$$

$$=\left(38304\right)\frac{3}{380000}$$

$$=0.3024$$

$$
\begin{aligned}
f_x(x) &= \int_y f(x,y)\,dy \\
&= K\int_y \left(x^2+y^2\right)dy \\
&= K\left(x^2y+\frac{y^3}{3}\right)_{20}^{30} \\
&= K\left(x^2(30-20)+\frac{30^3-20^3}{3}\right) \\
&= K\left(10x^2+\frac{19000}{3}\right) \\
&= \frac{3}{38000}\left(10x^2+\frac{19000}{3}\right) \\
&= \frac{1}{20}+\frac{3x^2}{38000}
\end{aligned}
$$

$$
\begin{aligned}
f_y(y) &= \int_x f(x,y)\,dx \\
&= K\int_x \left(x^2+y^2\right)dx \\
&= K\left(xy^2+\frac{x^3}{3}\right)_{20}^{30} \\
&= K\left(y^2(30-20)+\frac{30^3-20^3}{3}\right) \\
&= K\left(10y^2+\frac{19000}{3}\right) \\
&= \frac{3}{38000}\left(10y^2+\frac{19000}{3}\right) \\
&= \frac{1}{20}+\frac{3y^2}{38000}
\end{aligned}
$$

Answer

$$f_x(x)\,f_y(y) = \left(\frac{1}{20} + \frac{3x^2}{38000}\right)\left(\frac{1}{20} + \frac{3y^2}{38000}\right)$$

$$= \frac{\left(3x^2 + 1900\right)\left(3y^2 + 1900\right)}{1444000000}$$

$$\neq f(x, y)$$

Expected Value of a function

Let X and Y be jointly distributed rv's with pmf $p(x, y)$ or pdf $f(x, y)$ according to whether the variables are discrete or continuous. Then the expected value of a function $h(X, Y)$, denoted by $E[h(X, Y)]$ or $\mu_{h(X, Y)}$, is given by

$$E[h(X, Y)] = \begin{cases} \displaystyle\sum_x \sum_y h(x, y) \cdot p(x, y) & \text{if } X \text{ and } Y \text{ are discrete} \\ \displaystyle\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y) \cdot f(x, y)\, dx\, dy & \text{if } X \text{ and } Y \text{ are continuous} \end{cases}$$

Covariance

The **covariance** between two rv's X and Y is

$$\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

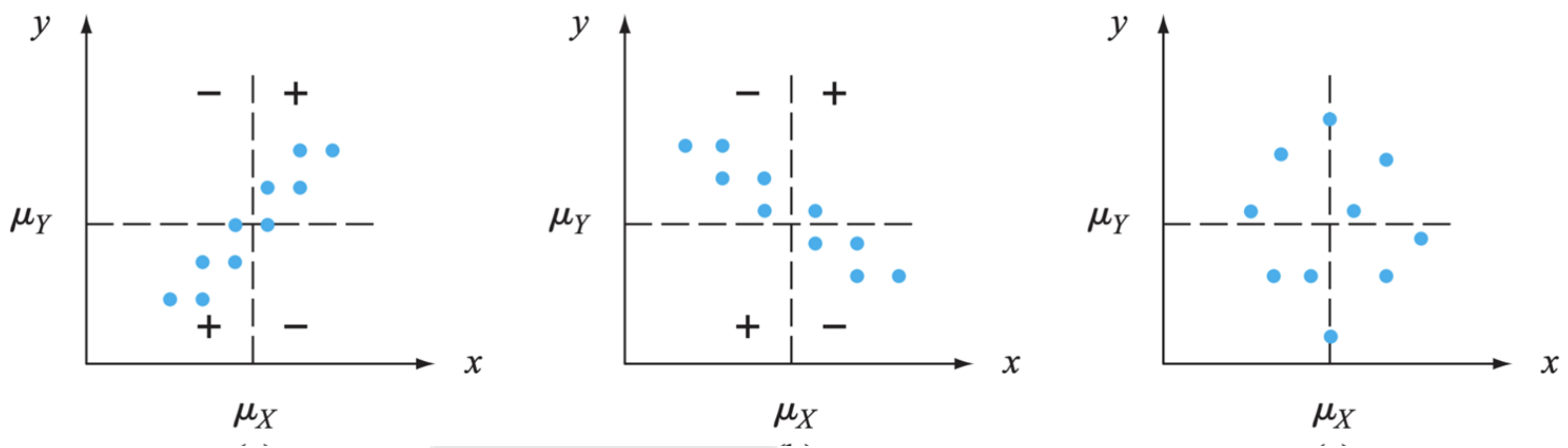
$$= \begin{cases} \sum_x \sum_y (x - \mu_X)(y - \mu_Y)p(x, y) & X, Y \text{ discrete} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y)f(x, y)\, dx\, dy & X, Y \text{ continuous} \end{cases}$$

That is, since $X - \mu_X$ and $Y - \mu_Y$ are the deviations of the two variables from their respective mean values, the covariance is the expected product of deviations. Note that $\text{Cov}(X, X) = E[(X - \mu_X)^2] = V(X)$.

Covariance

For a strong positive relationship that is when X increases then Y also increases, Cov(X, Y) would be quite positive.
For a strong negative relationship that is when X increases then Y decreases, Cov(X, Y) would be quite negative.
If X and Y are not strongly related, covariance will be near 0

Covariance



Question

Given the joint pmf, calculate Cov(X,Y)

		y		
	p(x,y)	0	1	2
x	0	0.4	0.1	0.1
	1	0.1	0.2	0.1

$$Cov(X,Y) = \sum\sum (x-\mu_x)(y-\mu_y)\,p(x,y)$$

Step 1: Figure out μ_x

$$\mu_x = \sum x\, p_X(x)$$

$$= 0.\,p_x(0) + 1.\,p_x(1)$$

$$= 0 + 1(0.4)$$

$$= 0.4$$

Step 2 : Figure out μ_y

$$\mu_y = \Sigma y\, p_Y(y)$$

$$= 0\, p_Y(0) + 1\, p_Y(1) + 2\, p_Y(2)$$

$$= 1(0.3) + 2(0.2)$$

$$= 0.3 + 0.4 = 0.7$$

Step 3 Find covariance

$$\sum\sum (x-\mu_x)(y-\mu_y)\, p(x,y)$$

$$(0-0.4)(0-0.7)(0.4) + (0-0.4)(1-0.7)(0.1)$$
$$+ (0-0.4)(2-0.7)(0.1) + (1-0.4)(0-0.7)(0.1)$$
$$+ (1-0.4)(1-0.7)(0.2) + (1-0.4)(2-0.7)(0.1)$$

$$= (-0.4)\times(-0.7)\times 0.4 + (-0.4)(0.3)(0.1)$$
$$+ (-0.4)(1.3)(0.1) + (0.6)(-0.7)(0.1)$$
$$+ (0.6)(0.3)(0.2) + (0.6)(1.3)(0.1)$$

$$= 0.112 - 0.012 - 0.052 - 0.042 + 0.036 + 0.078$$

$$= 0.12$$

Cov(X,Y) shortcut formula

$$
\begin{aligned}
\mathrm{Cov}(X,Y) &= E[(X-u_X)(Y-u_Y)] \\
&= E[XY + u_X u_Y - X u_Y - Y u_X] \\
&= E(XY) + E(u_X u_Y) - E(X u_Y) - E(Y u_X) \\
&= E(XY) + u_X u_Y - u_y E(X) - u_x E(Y) \\
&= E(XY) + u_X u_Y - u_y u_x - u_x u_y \\
&= E(XY) - u_X u_Y
\end{aligned}
$$

$$\mathrm{Cov}(X, Y) = E(XY) - \mu_X \cdot \mu_Y$$

$p(x,y)$		y = 0	1	2
x	0	0.4	0.1	0.1
	1	0.1	0.2	0.1

$$E(XY) = \sum xy\, p(x,y)$$

$$= (0)(0)(0.4) + (0)(1)(0.1) + (0)(2)(0.1) + (1)(0)(0.1)$$

$$+ (1)(1)(0.2) + (1)(2)(0.1)$$

$$= 0.2 + 0.2 = 0.4$$

$$\mu_x = 0.4 \qquad \mu_y = 0.7$$

$$E(XY) - \mu_x \mu_y = 0.4 - (0.4)(0.7)$$

$$= 0.4 - 0.28 = 0.12$$

Correlation Coefficient

The **correlation coefficient** of X and Y, denoted by Corr(X, Y), $\rho_{X,Y}$, or just ρ, is defined by

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y}$$

$$-1 \leq \mathrm{Corr}(X, Y) \leq 1.$$

Correlation Coefficient

1. If X and Y are independent, then $\rho = 0$, but $\rho = 0$ does not imply independence.
2. $\rho = 1$ or -1 iff $Y = aX + b$ for some numbers a and b with $a \neq 0$.

- ρ is a measure of the degree of linear relationship between X and Y.
- ρ = 0 does not imply that X and Y are independent, but only that there is a complete absence of a linear relationship.
- When ρ = 0, X and Y are said to be uncorrelated. Two variables could be uncorrelated yet highly dependent because there is a strong nonlinear relationship, so be careful not to conclude too much from knowing that ρ = 0.

Covariance vs Correlation

Covariance is nothing but a measure of correlation.	Correlation refers to the scaled form of covariance.
Covariance indicates the direction of the linear relationship between variables.	Correlation on the other hand measures both the strength and direction of the linear relationship between two variables.
Covariance can vary between -∞ and +∞	Correlation ranges between -1 and +1

References

Probability and statistics for engineers RA Johnson, I Miller, JE Freund - 2000 - 117.239.47.98

Statistics for business & economics DR Anderson, DJ Sweeney, TA Williams, JD Camm

Probability and statistics for engineering and science J Deovre