3Y03–3J04 PROBABILITY & STATISTICS FOR (CIVIL) ENGINEERING

C01 – Lecture 16

Last time: TWO RANDOM VARIABLES

Joint probability distribution function $f_{XY}(x,y)$ with

$$1 = \iint f_{XY}(x,y)\,dx\,dy \quad \text{and} \quad P((X,Y)\in R) = \iint_R f_{XY}(x,y)\,dx\,dy.$$

Marginal Distributions: $f_X(x) = \int f_{XY}(x,y)\,dy$ & $f_Y(y) = \int f_{XY}(x,y)\,dx$.

X, Y independent: $f_{XY}(x,y) = f_X(x)\cdot f_Y(y)$.

5.2 Covariance and Correlation

How X and Y vary together.

The covariance of X and Y is given by:

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

If $\sigma_{XY} > 0$, greater values of X correspond to greater values of Y (or smaller values to smaller values)

e.g. economic growth and stock market rise at same time

If $\sigma_{XY} < 0$, greater values of X correspond to smaller values of Y (or smaller to greater)

e.g. production increases, price goes down.

So how to find $\sigma_{XY} = E((X-\mu_X)(Y-\mu_Y))$

In general: if $h(X,Y)$ is a function of X & Y, then $E(h(X,Y)) = \iint h(x,y) f_{XY}(x,y)\,dx\,dy$.

So in particular $\sigma_{XY} = \iint (x-\mu_X)(y-\mu_Y) f_{XY}(x,y)\,dx\,dy$

$$\vdots$$

$$= \iint xy\, f_{XY}(x,y)\,dx\,dy - \mu_X \mu_Y$$

$$\sigma_{XY} = E(XY) - E(X)E(Y).$$

Notice If X, Y independent, $f_{XY}(x,y) = f_X(x) f_Y(y)$

$$E(XY) = \iint xy\, f_{XY}(x,y)\,dx\,dy = \int x f_X(x)\,dx \int y f_Y(y)\,dy$$
$$= E(X)E(Y).$$

So $\sigma_{XY} = 0$.

More strongly we have Corellation: the extent to which two variables move together by

Standardizing the measure of interdepence:

$$\mathrm{Corr}(X,Y) = \rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \in [-1, 1]$$

dimensionless (regardless of dimensions of X and Y).

$\rightarrow$ close to 1 "positively correlated"
$\rightarrow$ close to -1 "negatively correlated" $\Big\}$ $\leftarrow$ only reach ± 1 if $Y = aX + b$ for $a \gtrless 0$
$\rightarrow$ 0 - not correlated

<u>Notice</u> Since X, Y independent
$\Rightarrow \sigma_{XY} = 0$
We also have $\rho_{XY} = 0$

(But $\rho_{XY} = 0 \not\Rightarrow X, Y$ independent.)

<u>Exercise</u> Suppose X and Y are jointly distributed with joint pdf $f_{XY}(x,y) = \begin{cases} 6 & \text{if } x^2 < y < x \\ 0 & \text{otherwise.} \end{cases}$

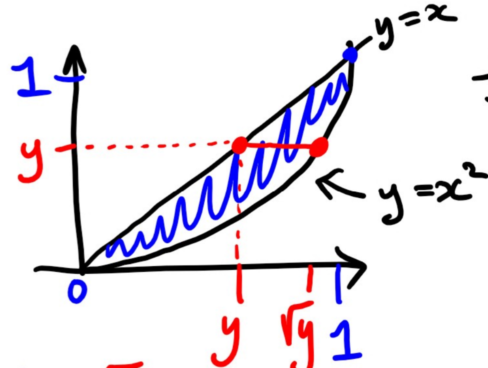
Find σ_{XY} and ρ_{XY}.

<u>Solution</u> $\sigma_{XY} = E(XY) - E(X)E(Y)$

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \quad \text{and} \quad \sigma_X = \sqrt{E(X^2) - E(X)^2}$$
$$\sigma_Y = \sqrt{E(Y^2) - E(Y)^2}$$

We need to find $E(X)$, $E(Y)$, $E(X^2)$, $E(Y^2)$, $E(XY)$.

Domain of f_{XY} :



$$f_{XY}(x,y) = \begin{cases} 6, & x^2 < y < x \\ 0 & \text{otherwise} \end{cases}$$

$$E(X) = \int_0^1 \int_y^{\sqrt{y}} x \cdot 6 \, dx \, dy = \int_0^1 \left[3x^2\right]_y^{\sqrt{y}} dy = \int_0^1 3y - 3y^2 \, dy$$

$$= \left[\frac{3y^2}{2} - y^3\right]_0^1 = \frac{3}{2} - 1 = \frac{1}{2}.$$

$$E(Y) = \int_0^1 \int_y^{\sqrt{y}} y \cdot 6 \, dx \, dy = \int_0^1 \left[6yx\right]_y^{\sqrt{y}} dy = \int_0^1 6y^{3/2} - 6y^2 \, dy$$

$$= \left[\frac{12}{5} y^{5/2} - 2y^3\right]_0^1 = \frac{12}{5} - 2 = \frac{2}{5}.$$

$$E(X^2) = \int_0^1 \int_y^{\sqrt{y}} x^2 \cdot 6 \, dx \, dy = \dots = \frac{3}{10}$$

$$E(Y^2) = \int_0^1 \int_y^{\sqrt{y}} y^2 \cdot 6 \, dx \, dy = \dots = \frac{3}{14}$$

$$E(XY) = \int_0^1 \int_y^{\sqrt{y}} xy \cdot 6 \, dx \, dy = \dots = \frac{1}{4}.$$

So maybe there's a positive relationship, but the size of this number is very hard to interpret as it depends on context

$$\text{So } \sigma_{XY} = E(XY) - E(X)E(Y)$$
$$= 1/4 - (1/2)(2/5) = 1/20 = 0.05.$$

And $\rho_{XY} = \dfrac{\sigma_{XY}}{\sigma_X \sigma_Y}$

$$\sigma_X = \sqrt{\left(\tfrac{3}{10}\right) - \left(\tfrac{1}{2}\right)^2} = \sqrt{\tfrac{1}{20}} = 1/\sqrt{20}$$

$$\sigma_Y = \sqrt{\tfrac{3}{14} - \left(\tfrac{2}{5}\right)^2} = \sqrt{\tfrac{19}{350}}$$

$$\rho_{XY} = \frac{1/20}{\left(1/\sqrt{20}\right)\left(\sqrt{\frac{19}{350}}\right)} = \sqrt{\frac{35}{38}} = \underline{\underline{0.96}}.$$

Very strongly positively correlated.

(Makes sense - the region is 'almost' linear.)

This number we can interpret within the standardized framework that $\rho_{XY} \in [-1, 1]$.

We do not really need to worry in this course about > 2 r.v.s except for knowing in principle that all the ideas above extend AND

5.4 Linear Functions of Random Variables

We can make linear combinations of r.v.s.

$$Y = c_1 X_1 + \ldots + c_k X_k \quad \text{for constants } c_i$$

In this situation,

$$E(Y) = c_1 E(X_1) + \ldots + c_k E(X_k)$$

and $$V(Y) = c_1^2 V(X_1) + \dots + c_k^2 V(X_k) + 2\sum_{i<j} \mathrm{Cov}(X_i, X_j)$$

$c_i c_j$ SORRY

If X_i are all independent then $\mathrm{Cov}(X_i, X_j) = 0$ for all $i \neq j$

So $$V(Y) = c_1^2 V(X_1) + \dots + c_k^2 V(X_k).$$