# §30. Covariance and Correlation

Let $X, Y$ be two random variables with joint pmf (pdf). Let $h(x,y) : \mathbb{R}^2 \to \mathbb{R}$ be a (deterministic) function.

$$E(h(X,Y)) = \sum_{x,y} h(x,y)\, p(x,y)$$

or

$$E\left(h\left(X,Y\right)\right) = \iint h(x,y)\, p(x,y)\, dx\, dy$$

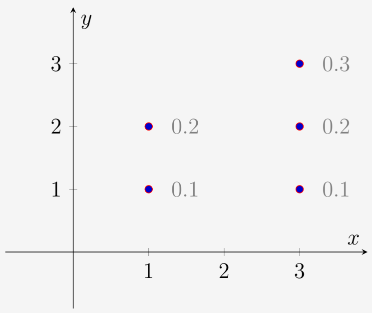
**Definition 1**

The **covariance** between the random variables $X$ and $Y$, is defined as

$$\operatorname{Cov}(X,Y) = E\left[(X - \mu_X)(Y - \mu_Y)\right] = E(XY) - \mu_X \mu_Y$$

**Example 1**

| $Y \backslash X$ | 1 | 3 | $p(y)$ |
|---|---|---|---|
| 3 | 0 | 0.3 | 0.3 |
| 2 | 0.2 | 0.2 | 0.4 |
| 1 | 0.1 | 0.2 | 0.3 |
| $p(x)$ | 0.3 | 0.7 | |




$$\mu_x = 1 \times 0.3 + 3 \times 0.7 = 2.4$$
$$\mu_y = 1 \times 0.3 + 2 \times 0.4 + 3 \times 0.3 = 2$$

$$\begin{aligned}\sigma_{xy} &= (1)(2)(0.2) + (1)(1)(0.1) + (3)(3)(0.3) + (3)(2)(0.2) \\ &\qquad + (3)(1)(0.2) - (2.4)(2) \\ &= 0.2\end{aligned}$$

### Example 2

$X, Y$ with joint pdf $p(x,y) = 2e^{-x-y}, \quad y \geq x \geq 0$. The marginals were completed in Lecture 28.

$$p(x) = 2e^{-2x} \quad , \quad x \geq 0$$
$$p(y) = 2\left(e^{-y} - e^{-2y}\right) \quad , \quad y \geq 0$$

$$\mu_x = \int_0^s 2xe^{-2x}\,dx = 2x\frac{e^{-2x}}{(-2)}\bigg|_0^\infty - 2\int_0^\infty \frac{e^{-2x}}{(-2)}\,dx = \frac{1}{2}$$

$$\mu_y = \int_0^\infty 2y\left(e^{-y} - e^{-2y}\right)dy$$
$$= 2y\left(\frac{e^{-y}}{(-1)} - \frac{e^{-2y}}{(-2)}\right)\bigg|_0^\infty - 2\int_0^\infty \left(\frac{e^{-y}}{(-1)} - \frac{e^{-2y}}{(-2)}\right)dy = \frac{3}{2}$$

$$\operatorname{Cov}(X,Y) = \int_0^\infty dx \int_x^\infty dy\, xy \cdot 2e^{-x-y} - \left(\frac{1}{2}\right)\left(\frac{3}{2}\right)$$
$$= 2\int_0^\infty dx \cdot xe^{-x}\left\{y\frac{e^{-y}}{(-1)}\bigg|_x^\infty - \int_x^\infty \frac{e^{-y}}{(-1)}dy\right\}$$
$$= 2\int_0^\infty dx\, xe^{-x}\left[xe^{-x} + e^{-x}\right] - \frac{3}{4}$$
$$= 2\int_0^\infty x^2e^{-2x}\,dx + 2\int_0^\infty xe^{-2x}\,dx - \frac{3}{4}$$
$$= 2\left(\frac{1}{4}\right) + 2\left(\frac{1}{4}\right) - \frac{3}{4}$$
$$= \frac{1}{4}$$

### Definition 2

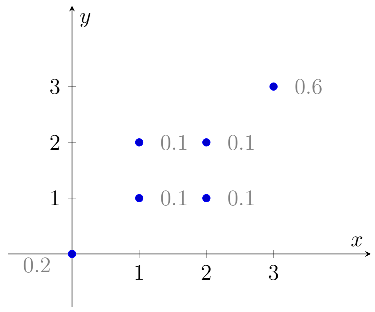
The correlation between random variables $X$ and $Y$ is

$$\rho_{XY} = \frac{\operatorname{Cov}(X,Y)}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y} \leftarrow \text{unitless measure of association}$$

### Theorem 1

$$-1 \leq \rho_{XY} \leq 1$$

### Example 3

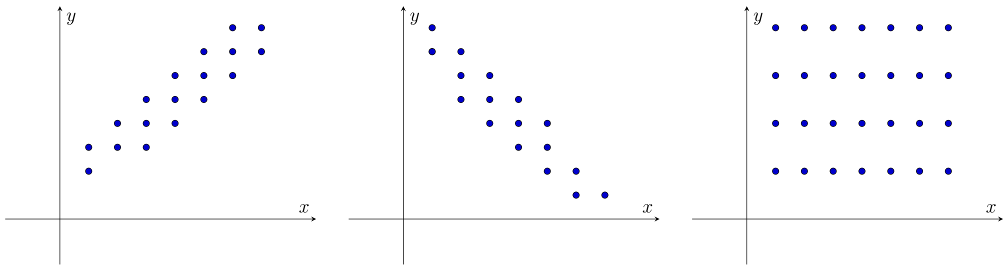



$$E(X) = E(Y) = 1 \times (0.2) + 2 \times (0.2) + 3 \times (0.6) = 2.4$$
$$E(XY) = 1 \times 1 \times (0.1) + 1 \times 2 \times (0.1) + 2 \times 1 \times (0.1) + 2 \times 2 \times (0.1) + 3 \times 3 \times (0.6) = 6.3$$

$$\sigma_{XY} = E(XY) - E(X)E(Y) = 6.3 - (2.4)^2 = 0.54$$
$$\text{Var}(X) = 1^2(0.2) + 2^2(0.2) + 3^2(0.6) - 2.4^2 = 0.64 = \text{Var}(Y)$$

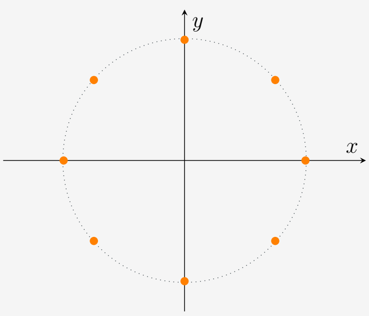
$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{0.54}{\sqrt{0.64^2}} = 0.84375$$

Correlation (and covariance) measure linear relationships between random variables.




Positive Correlation    Negative Correlation    $\rho = 0$

### Example 4




$$X^2 + Y^2 = 1$$

There is a nonlinear relationship between $X$ and $Y$.

### Theorem 2

If $X$ and $Y$ are independent random variables, then

$$\rho_{XY} = 0 \qquad \sigma_{XY} = 0$$