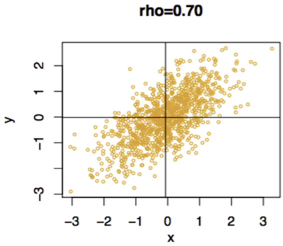
TECHNISCHE
UNIVERSITÄT
WIEN

Covariance and correlation



Please send your comments to: T. Levajković, tijana.levajkovic@tuwien.ac.at

Covariance

- Covariance of two random variables X and Y is given by

$$\mathbb{C}ov(X,Y) = \mathbb{E}\left((X - \mathbb{E}(X)) \cdot (Y - \mathbb{E}(Y))\right).$$

  - Variance is a special case of covariance. For $X = Y$ we get
$$\mathbb{C}ov(X,X) = \mathbb{V}ar(X).$$

  - Properties
    - $\mathbb{C}ov(X,Y) = \mathbb{E}(X \cdot Y) - \mathbb{E}(X) \cdot \mathbb{E}(Y)$
    - $\mathbb{C}ov(aX + b, cY + d) = ac\,\mathbb{C}ov(X,Y)$ for the constants a, b, c and d.
    - $\mathbb{C}ov(X_1 + X_2, Y) = \mathbb{C}ov(X_1, Y) + \mathbb{C}ov(X_2, Y)$

      Particularly, for $Y = Y_1 + Y_2$ we obtain
      $\mathbb{C}ov(X_1 + X_2, Y_1 + Y_2) = \mathbb{C}ov(X_1, Y_1) + \mathbb{C}ov(X_1, Y_2) + \mathbb{C}ov(X_2, Y_1) + \mathbb{C}ov(X_2, Y_2)$
    - $\mathbb{V}ar(X + Y) = \mathbb{V}ar(X) + \mathbb{V}ar(Y) + 2\mathbb{C}ov(X,Y)$ for all X and Y
      * Especially if X and Y are independent, then $\mathbb{V}ar(X + Y) = \mathbb{V}ar(X) + \mathbb{V}ar(Y)$
    - If X and Y are independent, then $\mathbb{C}ov(X,Y) = 0$.
      * The converse is not true! If covariance $\mathbb{C}ov(X,Y)$ is zero, random variables X and Y might not be independent.

Example

- Consider random variable $X \sim U(-1,1)$ and $Y = X^2$.
Their covariance is zero. But, as Y is a function of X, these random variables are not independent.

  - Namely, since $X \sim U(-1,1)$ we know that $\mathbb{E}(X) = 0$ and $\mathbb{V}ar\,(X) = \mathbb{E}(X^2) = \frac{1}{3}$. Also,

$$\mathbb{E}(XY) = \mathbb{E}(X^3) = \int_{-1}^{1} x^3 \cdot \frac{1}{2}\,dx = 0.$$

  Therefore,

$$\mathbb{C}ov(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X) \cdot \mathbb{E}(Y) = 0.$$

Reasoning: The covariance measures the linear dependance.

Two random variables might have quadratic relationship and their relationship would be **not** detected by the covariance calculation.

HW Confirm this result in R.
For computing the covariance use the function cov.

Correlation coefficient

- The correlation coefficient between X and Y is defined by

$$\rho(X,Y) = \frac{\mathbb{C}ov(X,Y)}{\sqrt{\mathbb{V}ar(X)} \cdot \sqrt{\mathbb{V}ar(Y)}}.$$

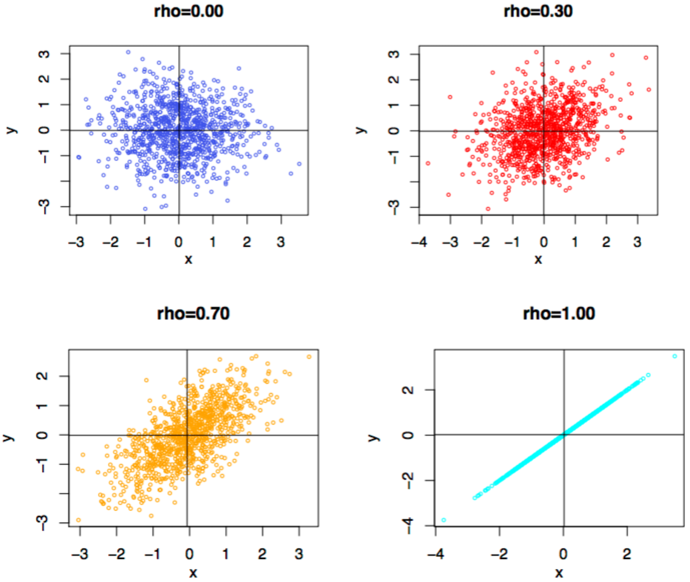
- Properties
  - $\rho(X,Y)$ is the covariance of the standardized versions of X and Y, i.e.

$$\rho(X,Y) = \mathbb{C}ov\left(\frac{X-\mathbb{E}(X)}{\sqrt{\mathbb{V}ar\,(X)}}, \frac{Y-\mathbb{E}(Y)}{\sqrt{\mathbb{V}ar\,(Y)}}\right)$$

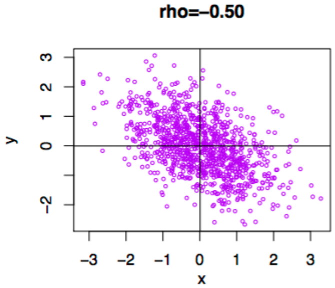
  - $-1 \leqslant \rho(X,Y) \leqslant 1$

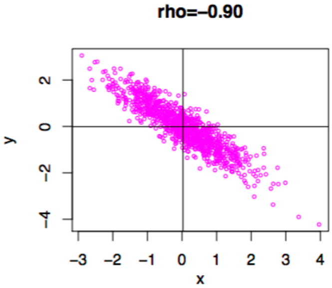
$$\begin{aligned} \rho(X,Y) &= 1 \qquad \Longleftrightarrow \qquad Y = aX + b \text{ with } a > 0 \\ \rho(X,Y) &= -1 \qquad \Longleftrightarrow \qquad Y = aX + b \text{ with } a < 0 \end{aligned}$$

Correlation coefficient



Correlation coefficient





Examples

(1) We flip a fair coin three times. Let X be the number of heads in the first two throws and let Y be the number of heads in the last two throws. Calculate $\mathbb{C}ov(X, Y)$ and $\rho(X, Y)$.

Answer:

- $\Omega = \{HHH, HHT, HTH, THH, TTH, THT, HTT, TTT\}, \ |\Omega| = 8.$
- Joint probability table:

X \ Y	0	1	2	$P(X = i) = p(x_i)$
0	$\frac{1}{8}$	$\frac{1}{8}$	0	$\frac{1}{4}$
1	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$	$\frac{1}{2}$
2	0	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{4}$
$P(Y = j) = p(y_j)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	1

- From marginal distributions we compute $\mathbb{E}(X) = 1$ and $\mathbb{E}(Y) = 1$.
- $\mathbb{E}(XY) = 1 \cdot \frac{2}{8} + 2 \cdot \frac{1}{8} + 2 \cdot \frac{1}{8} + 4 \cdot \frac{1}{8} = \frac{5}{4}$
- $\mathbb{C}ov(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = \frac{1}{4}$
- $\mathbb{V}ar(X) = \frac{1}{2}, \mathbb{V}ar(Y) = \frac{1}{2}$
- $\rho(X, Y) = \frac{\mathbb{C}ov(X,Y)}{\sigma_X \cdot \sigma_Y} = \frac{1}{2}$

Computatons: The pmf of the two dimensional random variable (X, Y), i.e. the vaues in the joint probability table are obtained in the following way:

$$P(X=0, Y=0) = P(\{TTT\}) = 1/8$$
$$P(X=0, Y=1) = P(\{TTH\}) = 1/8$$
$$P(X=0, Y=2) = 0$$
$$P(X=1, Y=0) = 0$$
$$P(X=1, Y=1) = P(\{HTH, THT\}) = 2/8$$
$$P(X=1, Y=2) = P(\{THH\}) = 1/8$$
$$P(X=2, Y=0) = P(\{HTT\}) = 1/8$$
$$P(X=2, Y=1) = P(\{HHT\}) = 1/8$$
$$P(X=2, Y=2) = P(\{HHH\}) = 1/8$$

$$\mathbb{E}(XY) = \sum_{i,j} x_i \cdot y_j \cdot P(X=i, Y=j) = 1 \cdot 2/8 + 2 \cdot 1/8 + 2 \cdot 1/8 + 4 \cdot 1/8 = 5/4$$
$$\mathbb{E}(X) = \sum_i x_i \cdot P(X=i) = 1 \cdot 1/2 + 2 \cdot 1/4 = 1 \text{ and } \mathbb{E}(Y) = 1$$
$$\mathbb{V}ar(X) = \sum_i {x_i}^2 \cdot P(X=i) - (\mathbb{E}(X))^2 = 1 \cdot 1/2 + 4 \cdot 1/4 - 1 = 1/2 \text{ and } \mathbb{V}ar(Y) = 1/2$$
$$\mathbb{C}ov(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = 5/4 - 1 = 1/4$$
$$\rho(X, Y) = 1/2.$$

- Second way: we use the properties of covariance
  - Let X_i be the result of ith flip, $i = 1, 2, 3$
  - $X_i \sim ber(\frac{1}{2})$
    - $\mathbb{E}(X_i) = \frac{1}{2}$ and $\mathbb{V}ar(X_i) = \frac{1}{4}$
    - Also, the different tosses are independent. Thus,
$$\mathbb{C}ov(X_1, X_2) = \mathbb{C}ov(X_1, X_3) = \mathbb{C}ov(X_2, X_3) = 0$$
  - Then, $X = X_1 + X_2$ and $Y = X_2 + X_3$
  - We obtain
$$\begin{aligned}\mathbb{C}ov(X, Y) &= \mathbb{C}ov(X_1 + X_2, X_2 + X_3)\\ &= \mathbb{C}ov(X_1, X_2) + \mathbb{C}ov(X_1, X_3) + \mathbb{C}ov(X_2, X_2) + \mathbb{C}ov(X_2, X_3)\\ &= \mathbb{C}ov(X_2, X_2) = \mathbb{V}ar(X_2) = \frac{1}{4}.\end{aligned}$$
  - Correlation: From the pairwise independence we have $\mathbb{C}ov(X_1, X_2) = 0$ and $\mathbb{C}ov(X_2, X_3) = 0$. Thus, $\mathbb{V}ar(X) = \mathbb{V}ar(X_1) + \mathbb{V}ar(X_2) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$ and $\mathbb{V}ar(Y) = \mathbb{V}ar(X_2) + \mathbb{V}ar(X_3) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$. Then,
$$\rho(X, Y) = \frac{\mathbb{C}ov(X, Y)}{\sigma_X \cdot \sigma_Y} = \frac{\frac{1}{4}}{\sqrt{\frac{1}{2}} \cdot \sqrt{\frac{1}{2}}} = \frac{1}{2}.$$

Examples

HW Let X be a random variable that takes values $-2, -1, 0, 1, 2$, each with probability $\frac{1}{5}$. Let $Y = X^2$. Show that $\mathbb{C}ov(X, Y) = 0$, but X and Y are not independent.

HW Let $X_1, \ldots, X_{10}$ be an independent and identically distributed random variables with $\mathcal{N}(3, 12)$. Let

$$Y_1 = \frac{1}{6} \sum_{i=1}^{6} X_i \qquad \text{and} \qquad Y_2 = \frac{1}{4} \sum_{i=7}^{10} X_i$$

(i) What is the distribution of $Y_1 + Y_2$?

(ii) Compute $\mathbb{C}ov(2Y_1 - 5, Y_2 + 4)$ and $\rho(Y_1, Y_2)$.

Examples

HW Let $X \sim bern(p)$.
Let Y_1 be the indicator that $X = 1$ and Y_0 be the indicator that $X = 0$.
Compute $\mathbb{C}ov(Y_0, Y_1) = 0$.

HW Let $Z \sim \mathcal{N}(0, 1)$.
Compute $\mathbb{C}ov(Z^2, Z^3)$.

HW Let $T \sim \exp(\lambda)$ and $S = T + 10$.

(i) Does S have an exponential distribution?
Use the pdf of S to answer this question.

(ii) Compute $\rho(T, S)$.

HW Consider $G \sim U(-3, -1)$, $V \sim U(-1, 1)$ and $W \sim U(1, 3)$ and assume that all are independent. Compute $\mathbb{C}ov(W - V, V - G)$.

Examples

HW Show that if X and Y are identically distributed (but not necessarily independent) then $\mathbb{C}ov(X+Y, X-Y) = 0$.

HW Suppose that the sample space S contains three elements $\{1, 2, 3\}$, with probabilities 0.5, 0.2 and 0.3 respectively. Suppose

$$X(s) = s^2 - 4, \quad \text{for} \quad s \in S.$$

Compute $\mathbb{E}(X)$ and $\mathbb{C}ov(X, |X|)$.

HW Suppose X is Poisson random variable with parameter $\lambda_1 = 1$, Y is an independent Poisson random variable with $\lambda_2 = 2$ and Z is a Poisson random variable with parameter $\lambda_3 = 3$. Assume X, Y and Z are independent. Compute

(a) $P(X+Y+Z=8)$.
(b) $\mathbb{C}ov(X+2Y, 2Y+3Z)$.
(c) $\mathbb{E}(XY)$.

Hint: Recall, If $X \sim Poi(\lambda_1)$, $Y \sim Poi(\lambda_2)$, $Z \sim Poi(\lambda_3)$ are independent, then $X+Y+Z \sim Poi(\lambda_1 + \lambda_2 + \lambda_3)$

Examples

HW Let X and Y be independent uniform random variables on $[0,1]$. Let

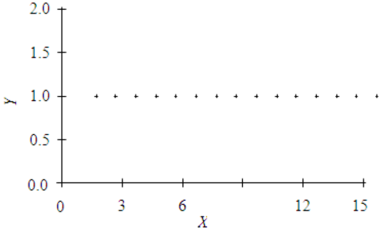
$$U = X + Y \qquad \text{and} \qquad V = \max\{X, Y\}.$$

Compute

(a) $\mathbb{E}(U)$ and $\mathbb{V}ar(U)$
(b) $P(V \geqslant \frac{1}{3})$
(c) the covariance and the correlation coefficient between Y and U.

Multiple-choice questions

(1) A scatterplot of a variable Y versus a variable X produced the scatterplot below. The value of Y for all values of X is exactly 1.0. The correlation coefficient between Y and X is:



a. 1
b. -1
c. 0
d. either 1 or -1

Multiple-choice questions

(2) Let X and Z be independent random variables and both $\mathcal{N}(0, 1)$-distributed. Let

$$Y = \frac{1}{2}\,(X + Z).$$

Then the correlation between X and Y is

a. smaller than the correlation between X and Z.
b. equal to the correlation between X and Z.
c. larger than the correlation between X and Z.
d. not defined.

Multiple-choice questions

(3) Let $a \neq 0$ and b be real constants. Which one of the following statements is true?

a. $\mathbb{V}ar(aX) = \mathbb{C}ov(aX, X)$
b. $\rho(aX, X) = 1$
c. $\mathbb{C}ov(aX + b, X) = a\,\mathbb{C}ov(X, X + b)$
d. $\mathbb{V}ar(aX - b) = a^2\mathbb{V}ar(X) + b.$

(4) Let $Z_1, \ldots, Z_{100}$ be i.i.d. $\mathcal{N}(0, 1)$ random variables. The correlation between $X = \sum_{i=1}^{98} Z_i$ and $Y = \sum_{i=3}^{100} Z_i$ is equal to

a. 0
b. $\frac{96}{98}$
c. $\frac{98}{100}$
d. 1

Multiple-choice questions

(5) Let X_1, X_2 and X_3 be uniform random variables on the interval $(0,1)$ with $\mathbb{C}ov(X_i, X_j) = \frac{1}{24}$ for $i, j = 1,2,3$, $i \neq j$. Then the variance of the sum

$$X_1 + 2X_2 - X_3$$

equals

a. $\frac{1}{6}$
b. $\frac{5}{12}$
c. $\frac{3}{8}$
d. $\frac{3}{4}$

(6) Let $X \sim \mathcal{N}(0,1)$ and $Y \sim \mathcal{N}(1,2)$ be two independent random variables. Compute

$$\mathbb{C}ov\left(3X + 2Y, 5X - 4Y + 7\right).$$

a. -1
b. 31
c. 0
d. -4

Multiple-choice questions

(7) Let X and Y be two independent standard normal random variables and

$$U = 1 + X + XY^2 \qquad \text{and} \qquad V = 1 + X.$$

Compute the covariance $\mathbb{C}ov(U, V)$.

a. 0
b. 2
c. 1
d. 4

(8) Let $Z \sim \mathcal{N}(0, 1)$. Then, the covariance $\mathbb{C}ov(Z^2, Z^3)$

a. does not exist.
b. equals zero.
c. is a positive real number.
d. is a negative real number.

Multiple-choice questions

(9) Let $X \sim \mathcal{N}(0,1)$ and $Y \sim \mathcal{N}(1,2)$ be two random variables such that the correlation $\mathbb{C}orr\,(2X - Y, X + 3) = -0.8$. Then the covariance $\mathbb{C}ov(X,Y)$ is

a. a negative real number less than -1
b. a negative real number bigger than -1
c. a positive real number less than 1
d. a positive real number bigger than 1

(10) Let $X \sim \mathcal{U}(1,2)$ and $Y = \frac{1}{X^2}$. Then,

a. $\mathbb{E}X = \mathbb{E}Y$
b. $\mathbb{V}arX \leqslant \mathbb{V}arY$
c. $\mathbb{C}ov(X,Y) \leqslant \ln 2$
d. $Corr(X,Y) \geqslant 0$

Next week we start with Statistics

Thank you for your attention!