Mathematical Statistics

MAS 713

Chapter 3.4

This lecture

3.4. Random variables

- 3.4.1 Jointly distributed Random Variables

Additional reading : Chapter 4 in the textbook

Jointly distributed Random Variables

Joint distribution function

Often, probability statements concerning two random variables, say X and Y, defined on the same sample space are of interest :

$$\omega \to (X(\omega), Y(\omega)))$$

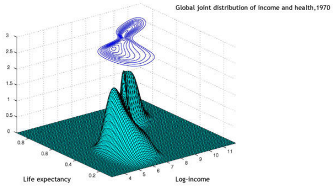
$\leadsto$ these two variables are most certainly related

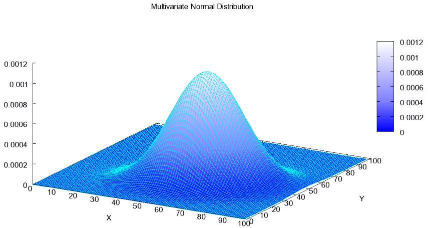
$\leadsto$ they should be jointly analysed, in order to understand the degree of relationship between them

Joint distribution function

For instance, we may simultaneously measure:

- the weight and hardness of a rock,
- the pressure and temperature of a gas,
- thickness and compression strength of a piece of glass, etc.





Joint distribution function

Coin Flips

Consider the flip of two fair coins.
Let A and B be discrete random variables associated with the outcomes first and second coin flips respectively.

If a coin displays "heads" then associated random variable is 1, and is 0 otherwise.

- The **joint probability mass** function of A and B defines probabilities for each pair of outcomes.
- All possible outcomes are

$$S_X = \{(A=0, B=0), (A=0, B=1), (A=1, B=0), (A=1, B=1)\}$$

- As each outcome is equally likely the joint pmf is

$$P(A, B) = \frac{1}{4}, \qquad A, B \in \{0, 1\}$$

Joint distribution function

Definition

The joint cumulative distribution function of X and Y is given by

$$F_{XY}(x, y) = \mathbb{P}(X \leq x, Y \leq y) \qquad \forall (x, y) \in \mathbb{R} \times \mathbb{R}$$

Note : $(X \leq x, Y \leq y)$ is the usual notation for $(X \leq x) \cap (Y \leq y)$

Joint distribution : discrete case

- If X and Y are **both** discrete, the joint probability mass function is defined by

$$p_{XY}(x,y) = \mathbb{P}(X = x, Y = y)$$

- The marginal pmf of X and Y can be obtained by

$$p_X(x) = \sum_{y \in S_Y} p_{XY}(x,y) \quad \text{and} \quad p_Y(y) = \sum_{x \in S_X} p_{XY}(x,y)$$

- The conditional pmf of X given Y can be obtained by

$$p_{X|Y}(x|y) := \mathbb{P}(X = x \mid Y = y) = \frac{p_{XY}(x,y)}{p_Y(y)}, \quad \text{if } p_Y(y) > 0.$$

Joint distribution : continuous case

Definition

X and Y are said to be jointly continuous if there exists a function $f_{XY}(x, y) : \mathbb{R} \times \mathbb{R} \to \mathbb{R}^+$ such that for any sets A and B of real numbers

$$\mathbb{P}(X \in A, Y \in B) = \int_A \int_B f_{XY}(x, y) dy\, dx$$

- The function $f_{XY}(x, y)$ is the joint probability density of X and Y

Joint distribution : continuous case

- The marginal densities follows from

$$\int_A f_X(x)dx = \mathbb{P}(X \in A) = \mathbb{P}(X \in A, Y \in S_Y) = \int_A \int_{S_Y} f_{XY}(x,y)dy\,dx$$

Thus,

$$f_X(x) = \int_{S_Y} f_{XY}(x,y)dy \quad \text{and} \quad f_Y(y) = \int_{S_X} f_{XY}(x,y)dx$$

- The marginal density of X given Y can be derived similarly and is

$$f_{X|Y}(x|y) := \begin{cases} \frac{f_{XY}(x,y)}{f_Y(y)} & \text{if } f_Y(y) > 0 \\ 0 & \text{if } f_Y(y) = 0 \end{cases}$$

Expectation of a function of two random variables

Expectation of a function of two random variables

For any function $g : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, the expectation of $g(X, Y)$ is given by:

Discrete case:

$$\mathbb{E}(g(X, Y)) = \sum_{x \in S_X} \sum_{y \in S_Y} g(x, y) p_{XY}(x, y)$$

Continuous case:

$$\mathbb{E}(g(X, Y)) = \int_{S_X} \int_{S_Y} g(x, y) f_{XY}(x, y) dy\, dx.$$

Example: For instance, in the continuous case,

$$\begin{aligned}
\mathbb{E}(aX+bY) &= \int_{S_X}\int_{S_Y}(ax+by)f_{XY}(x,y)dy\,dx \\
&= \int_{S_X}\int_{S_Y} ax\,f_{XY}(x,y)dy\,dx + \int_{S_X}\int_{S_Y} by\,f_{XY}(x,y)dy\,dx \\
&= a\int_{S_X} x\int_{S_Y} f_{XY}(x,y)dy\,dx + b\int_{S_Y} y\int_{S_X} f_{XY}(x,y)dx\,dy \\
&= a\int_{S_X} xf_X(x)dx + b\int_{S_Y} yf_Y(y)dy \\
&= a\mathbb{E}(X) + b\mathbb{E}(Y)
\end{aligned}$$

Example

What is the expected sum obtained when two fair dice are rolled?

Let X be the sum and X_i the value shown on the ith die. Then, $X = X_1 + X_2$, and

$$\mathbb{E}(X) = \mathbb{E}(X_1) + \mathbb{E}(X_2) = 2\times 3.5 = 7$$

Independent random variables

Independent random variables

Recall: Definition independence of random variables

The random variables X, Y are said to be independent **if and only if** for all $(x, y) \in \mathbb{R} \times \mathbb{R}$,

$$\mathbb{P}(X \leq x, Y \leq y) = \mathbb{P}(X \leq x) \times \mathbb{P}(Y \leq y)$$

Characterisation : X, Y are independent $\iff$ for any $(x, y) \in \mathbb{R} \times \mathbb{R}$,

$$F_{XY}(x, y) = F_X(x) \times F_Y(y),$$

which reduces to

$$p_{XY}(x, y) = p_X(x) \times p_Y(y) \qquad \text{in the **discrete** case}$$

or

$$f_{XY}(x, y) = f_X(x) \times f_Y(y) \qquad \text{in the **continuous** case}$$

Independent random variables

Recall: Property

If X and Y are independent, then for any functions h and g,

$$\mathbb{E}(h(X)g(Y)) = \mathbb{E}(h(X)) \times \mathbb{E}(g(Y))$$

Proof (in the continuous case) :

$$\begin{aligned}
\mathbb{E}(h(X)g(Y)) &= \iint_{S_X \times S_Y} h(x)g(y)f_{XY}(x,y)dy\,dx \\
&= \int_{S_X}\int_{S_Y} h(x)g(y)f_X(x)f_Y(y)dy\,dx \\
&= \int_{S_X} h(x)f_X(x)dx \times \int_{S_Y} g(y)f_Y(y)dy \\
&= \mathbb{E}(h(X)) \times \mathbb{E}(g(Y))
\end{aligned}$$

Independence via Characteristic function

Theorem: Characterization of Independence

Let X_1, X_2 be two random variables. Then the following are equivalent.

1) X_1, X_2 are independent

2) For **all** $(t_1, t_2) \in \mathbb{R}^2$ it holds that

$$\varphi_{(X_1,X_2)}(t_1, t_2) := \mathbb{E}\left[e^{i(t_1 X_1 + t_2 X_2)}\right] = \mathbb{E}\left[e^{it_1 X_1}\right] \mathbb{E}\left[e^{it_2 X_2}\right] =: \varphi_{X_1}(t_1)\, \varphi_{X_2}(t_2)$$

Careful:

$\mathbb{E}\left[e^{it(X_1+X_2)}\right] = \mathbb{E}\left[e^{itX_1}\right] \mathbb{E}\left[e^{itX_2}\right] \; \forall t$ NOT sufficient for independence

Covariance of two random variables

Covariance of two random variables

Definition

The covariance of two random variables X and Y is defined by

$$\mathbb{C}\text{ov}(X, Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y)))$$

Properties (proofs are left as an exercise) :

- $\mathbb{C}\text{ov}(X, Y) = \mathbb{C}\text{ov}(Y, X)$
- $\mathbb{C}\text{ov}(X, X) = \mathbb{V}\text{ar}(X)$
- $\boxed{\mathbb{C}\text{ov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)}$
- $\mathbb{C}\text{ov}(aX + b, cY + d) = ac\,\mathbb{C}\text{ov}(X, Y)$
- $\mathbb{C}\text{ov}(X_1 + X_2, Y_1 + Y_2)$
  $= \mathbb{C}\text{ov}(X_1, Y_1) + \mathbb{C}\text{ov}(X_1, Y_2) + \mathbb{C}\text{ov}(X_2, Y_1) + \mathbb{C}\text{ov}(X_2, Y_2)$

Note : unit of $\mathbb{C}\text{ov}(X, Y) =$ (unit of X) $\times$ (unit of Y)

Covariance: interpretation

- Suppose X and Y are two Bernoulli random variables

$\Longrightarrow XY$ is also a Bernoulli random variable with:

$XY = 1$ if and only if $X = 1$ and $Y = 1$. Thus

$$\mathbb{C}\text{ov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = \mathbb{P}(X = 1, Y = 1) - \mathbb{P}(X = 1)\mathbb{P}(Y = 1)$$

Then,

$$\begin{aligned} \mathbb{C}\text{ov}(X, Y) > 0 &\Leftrightarrow \mathbb{P}(X = 1, Y = 1) > \mathbb{P}(X = 1)\mathbb{P}(Y = 1) \\ &\Leftrightarrow \frac{\mathbb{P}(X = 1, Y = 1)}{\mathbb{P}(X = 1)} > \mathbb{P}(Y = 1) \\ &\Leftrightarrow \mathbb{P}(Y = 1 | X = 1) > \mathbb{P}(Y = 1) \end{aligned}$$

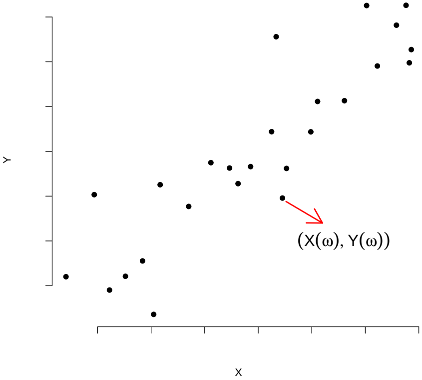
$\rightsquigarrow$ the outcome $X = 1$ makes it more likely that $Y = 1$

$\rightsquigarrow$ Y tends to increase when X does, and vice-versa

This result holds for **any** r.v. X and Y (not only Bernoulli r.v.)

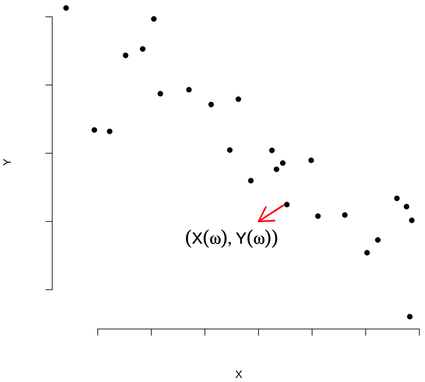
Covariance: interpretation

- $\mathbb{C}\text{ov}(X, Y) > 0 \rightsquigarrow X$ and Y tend to increase or decrease together
- $\mathbb{C}\text{ov}(X, Y) < 0 \rightsquigarrow X$ tends to increase as Y decreases and vice-versa
- $\mathbb{C}\text{ov}(X, Y) = 0 \rightsquigarrow$ no linear association between X and Y (doesn't mean no association at all!)



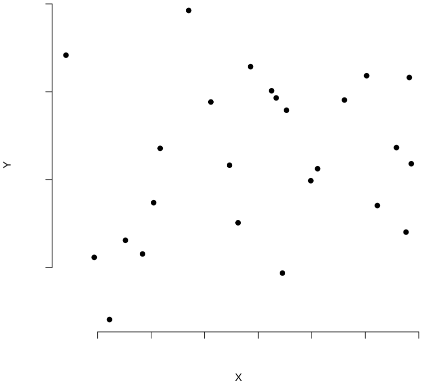
Covariance: interpretation

- $\mathbb{C}\text{ov}(X, Y) > 0 \rightsquigarrow$ X and Y tend to increase or decrease together
- $\mathbb{C}\text{ov}(X, Y) < 0 \rightsquigarrow$ X tends to increase as Y decreases and vice-versa
- $\mathbb{C}\text{ov}(X, Y) = 0 \rightsquigarrow$ no linear association between X and Y (doesn't mean no association at all!)



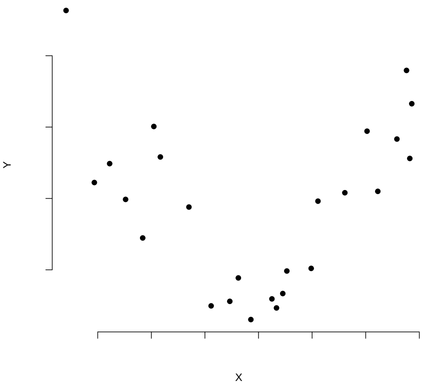
Covariance: interpretation

- $\mathbb{C}\text{ov}(X, Y) > 0 \rightsquigarrow$ X and Y tend to increase or decrease together
- $\mathbb{C}\text{ov}(X, Y) < 0 \rightsquigarrow$ X tends to increase as Y decreases and vice-versa
- $\mathbb{C}\text{ov}(X, Y) = 0 \rightsquigarrow$ no linear association between X and Y (doesn't mean no association at all!)



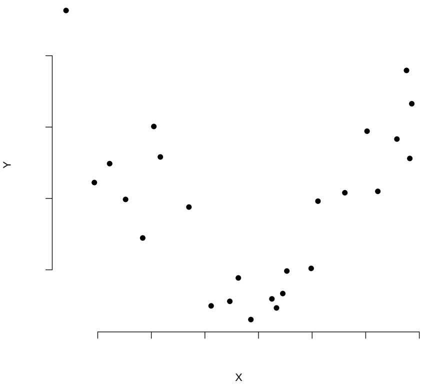
Covariance: interpretation

- $\mathbb{C}\text{ov}(X, Y) > 0 \rightsquigarrow$ X and Y tend to increase or decrease together
- $\mathbb{C}\text{ov}(X, Y) < 0 \rightsquigarrow$ X tends to increase as Y decreases and vice-versa
- $\mathbb{C}\text{ov}(X, Y) = 0 \rightsquigarrow$ no linear association between X and Y (doesn't mean no association at all!)



Covariance: interpretation

- $\mathbb{C}\text{ov}(X,Y) > 0 \rightsquigarrow$ X and Y tend to increase or decrease together
- $\mathbb{C}\text{ov}(X,Y) < 0 \rightsquigarrow$ X tends to increase as Y decreases and vice-versa
- $\mathbb{C}\text{ov}(X,Y) = 0 \rightsquigarrow$ no linear association between X and Y (doesn't mean no association at all!)



Fact

$$X \text{ and } Y \text{ independent } \Rightarrow \mathbb{C}\text{ov}(X,Y) = 0$$

$\not\Leftarrow$ (i.e. converse **NOT** true)

Covariance: examples

Counterexample

Let X be a r. v. with $\mathbb{E}[X] = \mathbb{E}[X^3] = 0$ (for example X centered Gaussian), and let $Y := X^2$. Find $\mathbb{C}\text{ov}(X, Y)$

- We have $\mathbb{C}\text{ov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = \mathbb{E}(X^3) - \mathbb{E}(X)\mathbb{E}(X^2) = 0$. So,

$$\mathbb{C}\text{ov}(X, Y) = 0$$

However there is a direct functional dependence between X and Y

Variance of a sum of random variables

- From the properties of the covariance, it follows :

$$\begin{aligned}\mathbb{V}\text{ar}\,(aX+bY) &= \mathbb{C}\text{ov}\,(aX+bY, aX+bY)\\ &= \mathbb{C}\text{ov}(aX,aX)+\mathbb{C}\text{ov}(aX,bY)\\ &\qquad +\mathbb{C}\text{ov}(bY,aX)+\mathbb{C}\text{ov}(bY,bY)\\ &= \mathbb{V}\text{ar}(aX)+\mathbb{V}\text{ar}(bY)+2\,\mathbb{C}\text{ov}(aX,bY)\\ &= a^2\,\mathbb{V}\text{ar}(X)+b^2\,\mathbb{V}\text{ar}(Y)+2ab\,\mathbb{C}\text{ov}(X,Y)\end{aligned}$$

- Now, if X and Y are independent random variables,

$$\mathbb{V}\text{ar}\,(aX+bY)=a^2\,\mathbb{V}\text{ar}(X)+b^2\,\mathbb{V}\text{ar}(Y)$$

- For instance, if X and Y are independent,

$$\mathbb{V}\text{ar}(X+Y)=\mathbb{V}\text{ar}(X)+\mathbb{V}\text{ar}(Y)$$

$$\mathbb{V}\text{ar}(X-Y)=\mathbb{V}\text{ar}(X)+\mathbb{V}\text{ar}(Y)$$

Example

We have two scales for measuring small weights in a laboratory. Assume the true weight of an item is 2g. Both scales give readings which have mean 2g and variance 0.05g^2.

Compare using only one scale and using both scales then averaging the two measures in terms of the accuracy.

- The first measure X has $\mathbb{E}(X) = 2$ and $\mathbb{V}\text{ar}(X) = 0.05$.
- The second measure Y, indep. of X, has also $\mathbb{E}(Y) = 2$ and $\mathbb{V}\text{ar}(Y) = 0.05$
- Let $W = \frac{X+Y}{2}$. Then, we have

$$\mathbb{E}(W) = \frac{1}{2}\mathbb{E}(X) + \frac{1}{2}\mathbb{E}(Y) = \frac{2}{2} + \frac{2}{2} = 2 \text{ (g)}$$

and

$$\mathbb{V}\text{ar}(W) = \frac{1}{4}\mathbb{V}\text{ar}(X) + \frac{1}{4}\mathbb{V}\text{ar}(Y) = \frac{1}{4}0.05 + \frac{1}{4}0.05 = 0.025 \text{ (g}^2\text{)}$$

$\rightsquigarrow$ averaging 2 measures reduces the variance by 2

Correlation

Correlation

The covariance of two r.v. is important as an indicator of the relationship between them

However, it heavily depends on units of X and Y (difficult interpretation, not invariant)

$\rightsquigarrow$ the correlation coefficient ρ is often used instead

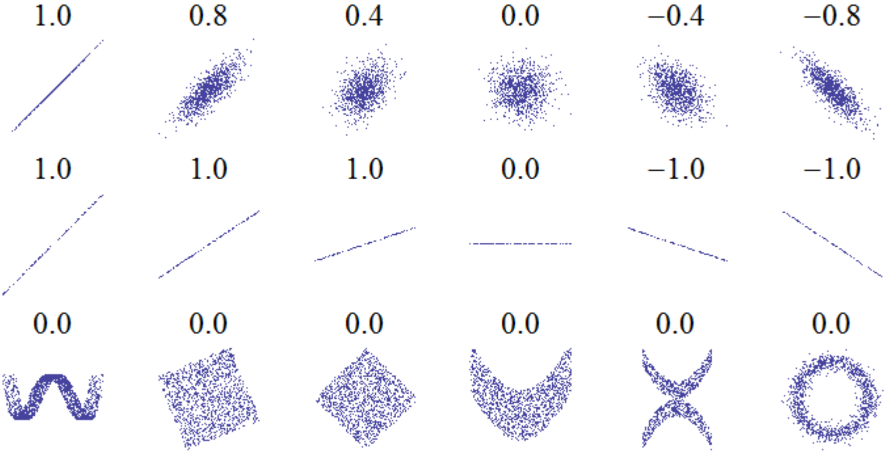
It is the covariance between the standardised versions of X and Y, or, explicitly,

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

Properties :

- ρ is dimensionless (no unit)
- ρ always has a value between -1 and 1 (Cauchy-Schwarz ineq.)
- positive (resp. negative) ρ means positive (resp. negative) linear relationship between X and Y
- the closer $|\rho|$ is to 1, the stronger is the linear relationship

Correlation



Objectives

Now you should be able to :

- use joint pmf and joint pdf to calculate probabilities □
- calculate and interpret covariances and correlations between two random variables □

Put yourself to the test ! $\rightsquigarrow$ Q1 p.192, Q4 p.192, Q10 p.193, Q30 p.195.