CS 107, Probability, Spring 2020
Lecture 36

Michael Poghosyan
mpoghosyan@aua.am

AUA

06 May 2020

Content

- Covariance of r.v.s

Dependency Measures

Dependency Measures: Intro

Recall that when considering the Joint Distribution of r.v.s, we have mostly talked about the Independence. This is because dependency is much harder to describe. Well, we know when r.v.s are dependent - this is when they are not Independent, but we cannot say anything more at this point.

Of course, the best way to describe the dependence between r.v.s is to give one of their complete Joint characteristics- either their Joint CDF or their Joint PMF/PDF. But, in practice, rarely we can find these Joint characteristics. So here, like in the 1D case, 1 r.v. case, we want to give some partial, numerical, characteristics to measure the dependence.
Another thing making the description of Dependency very hard is that dependency can be of any form - say:

Dependency, Example

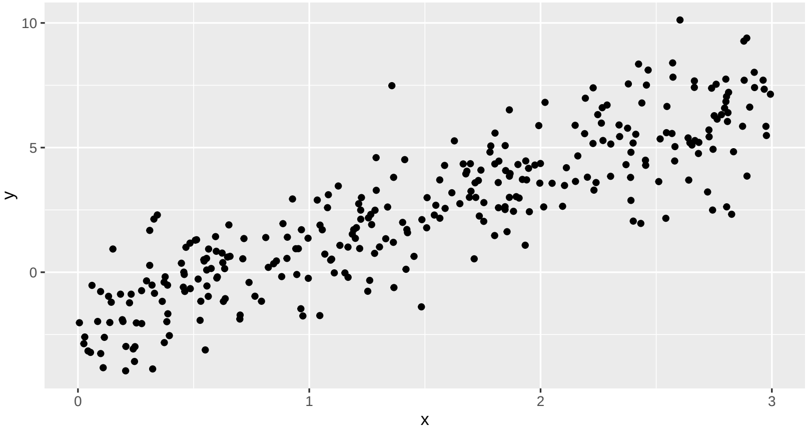


Figure: Data generated from some 2D distribution

Dependency, Example

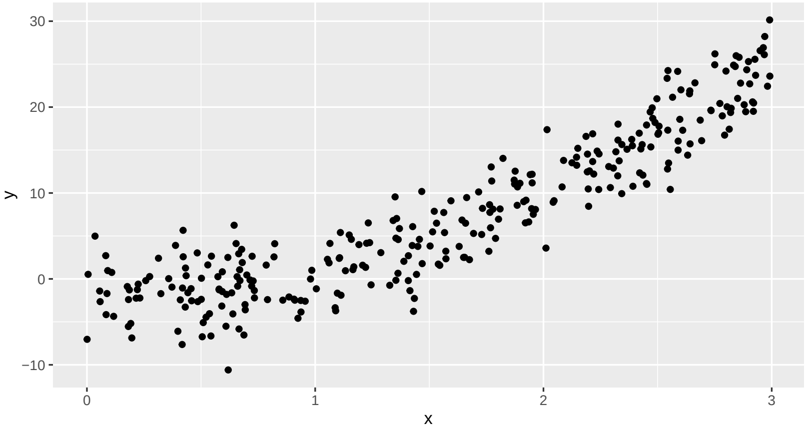


Figure: Data generated from some 2D distribution

Dependency, Example

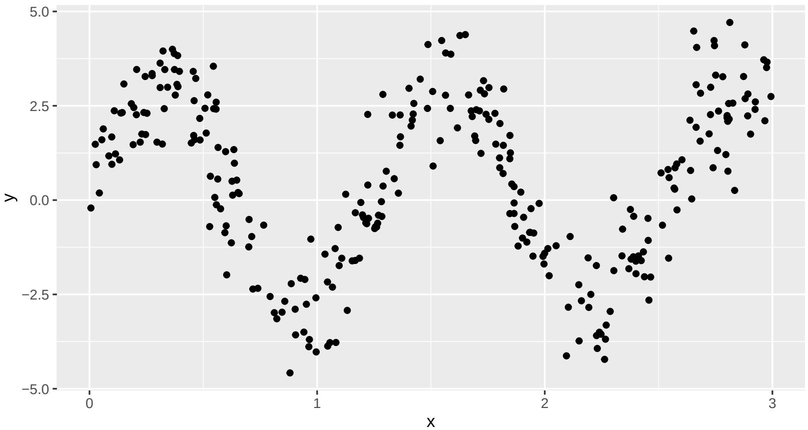


Figure: Data generated from some 2D distribution

Dependency, Example

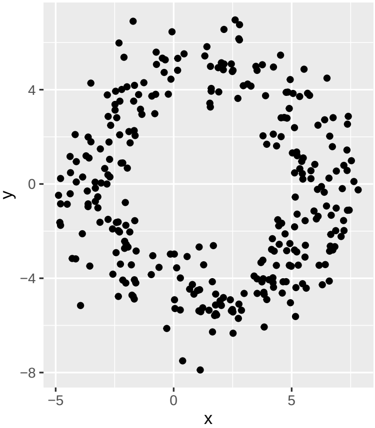


Figure: Data generated from some 2D distribution

Dependency, Example

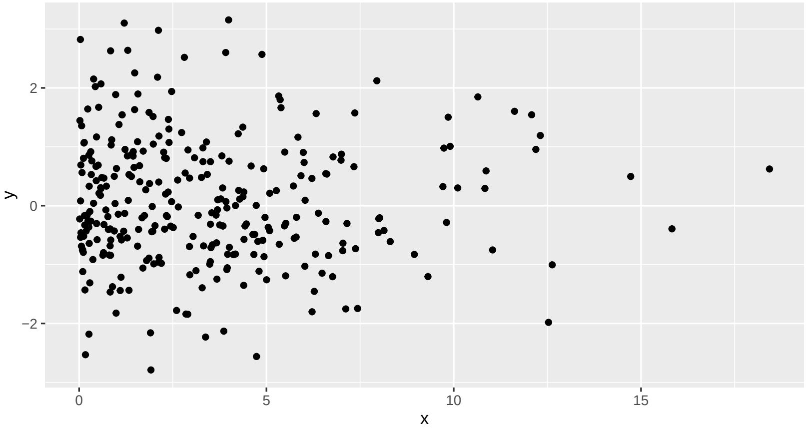


Figure: Data generated from some 2D distribution

Dependency, Example

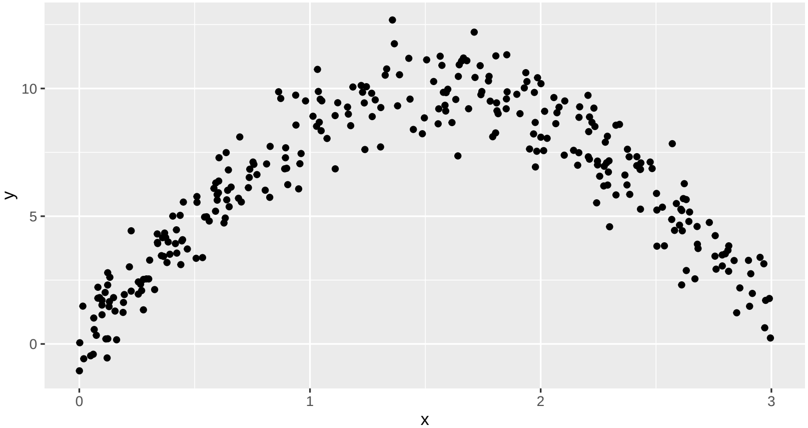


Figure: Data generated from some 2D distribution

Dependency, Example

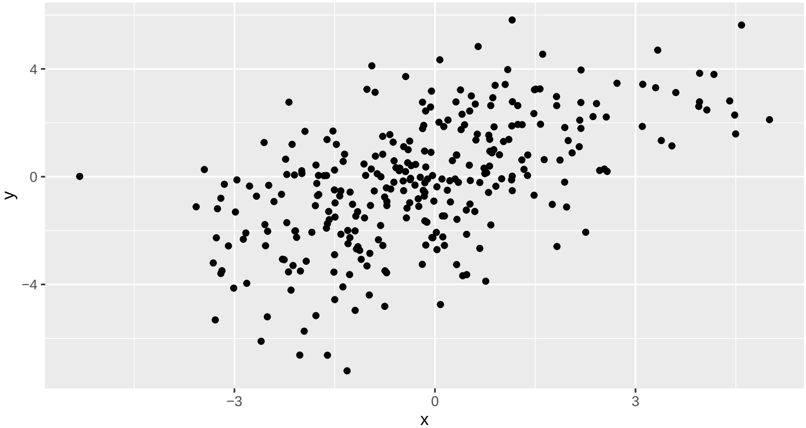


Figure: Data generated from some 2D distribution

Dependency, Example

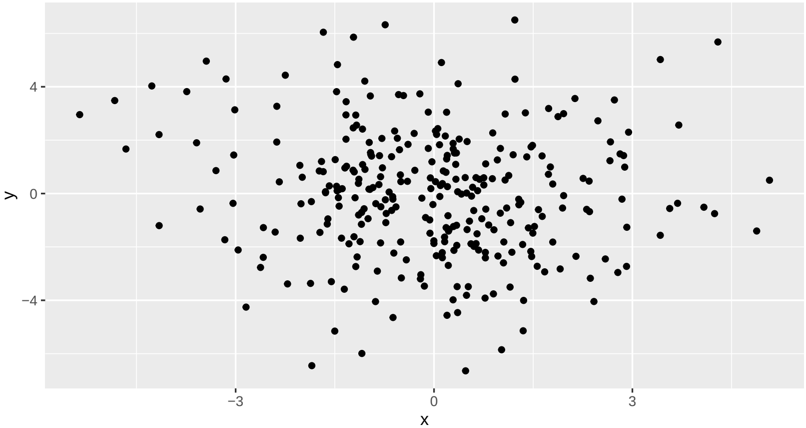


Figure: Data generated from some 2D distribution

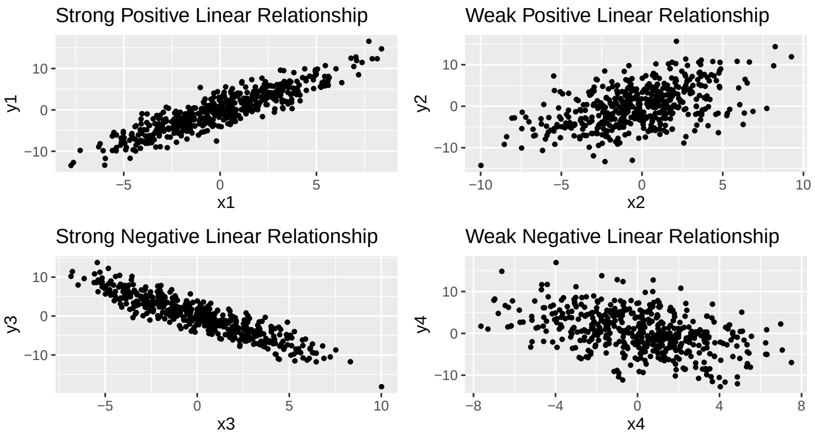
Linear Relationship

Note: In the above graphs, some realizations of, data from (X, Y) are depicted. In Probability, we want to talk about the dependency of r.v.s X and Y, of *Processes generating that data*.

In this lecture we want to introduce a deterministic quantity to measure the linear relationship, association between 2 r.v.s. It will give us a tool to measure if we have a

- strong positive linear relationship;
- week positive linear relationship;
- strong negative linear relationship;
- week negative linear relationship;
- no linear relationship at all.

Linear Relationship Types



Linear Relationship Types

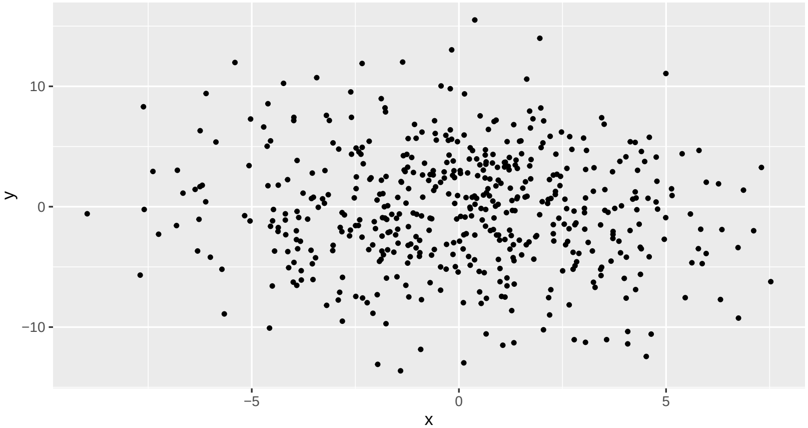


Figure: No Linear Relationship

Linear Relationship Types

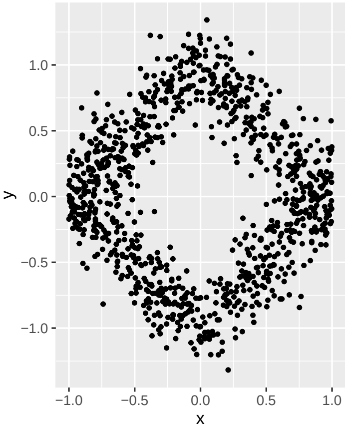


Figure: No Linear Relationship

Linear Relationship Types

We will consider the Correlation Coefficient as the measure to quantify the linear relationship between r.v.s. But first we will introduce the Covariance of r.v.s, since the Correlation Coefficient is obtained from the Covariance.

Covariance of Random Variables

Covariance of r.v.s

Assume X and Y are Jointly Distributed r.v.s.

Covariance of r.v.s

The **Covariance** of r.v.s X and Y is the following number:

$$Cov(X, Y) = \sigma_{XY} = \mathbb{E}\Big((X - \mathbb{E}(X)) \cdot (Y - \mathbb{E}(Y))\Big).$$

Note: Covariance = Co-Variance, measures how X and Y Co-Vary, i.e. vary together.

Uncorrelated r.v.s

We say that X and Y are **uncorrelated**, if $Cov(X, Y) = 0$.

And, if X and Y are not Uncorrelated, we say that they are Correlated $\ddot\smile$

Covariance of r.v.s, Interpretation

Interpretation: It is not so easy to interpret the magnitude of Covariance (we will later talk about this), but the sign of Covariance is giving us the direction of linear relationship.
Assume $Cov(X, Y) > 0$. This means that

$$Cov(X, Y) = \mathbb{E}\Big((X - \mathbb{E}(X)) \cdot (Y - \mathbb{E}(Y))\Big) > 0.$$

So

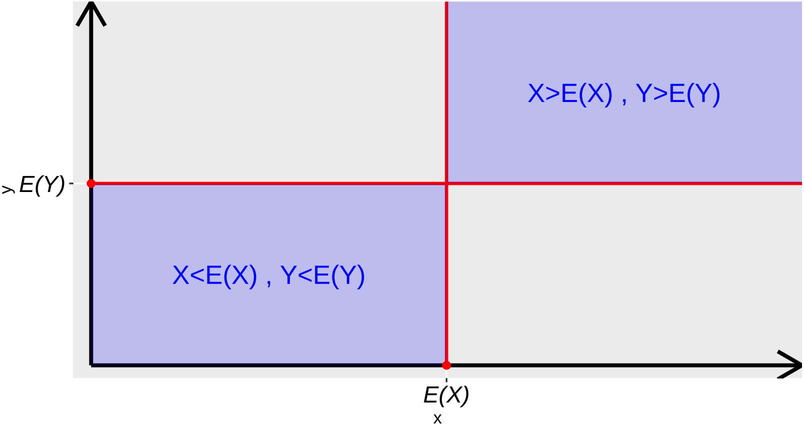
the Mean of $\quad (X - \mathbb{E}(X)) \cdot (Y - \mathbb{E}(Y)) \quad$ is positive.

We can interpret this as:

mostly, $\quad (X - \mathbb{E}(X)) \cdot (Y - \mathbb{E}(Y)) \quad$ is positive.

Covariance of r.v.s, Interpretation

This means that mostly, $X - \mathbb{E}(X)$ and $Y - \mathbb{E}(Y)$ are of the same sign, i.e., the values of (X, Y) are mostly in the following region:



Covariance of r.v.s, Interpretation

Example:

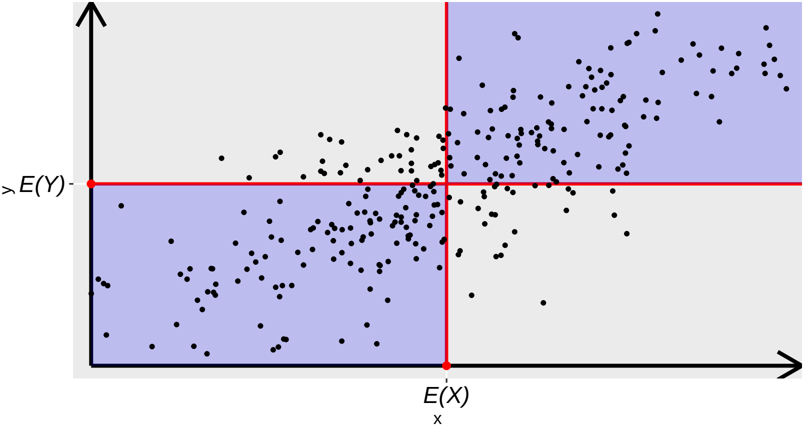
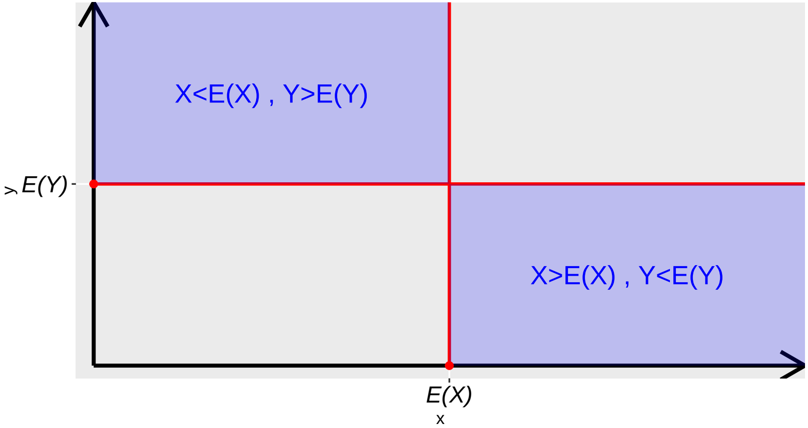


Figure: Data from (X, Y), with $Cov(X, Y) > 0$

Covariance of r.v.s, Interpretation

In contrast, if we have $Cov(X, Y) < 0$, we can think as mostly $X - \mathbb{E}(X)$ and $Y - \mathbb{E}(Y)$ are of the different signs, i.e., the values of (X, Y) are mostly in the following region:



Covariance of r.v.s, Interpretation

Example:

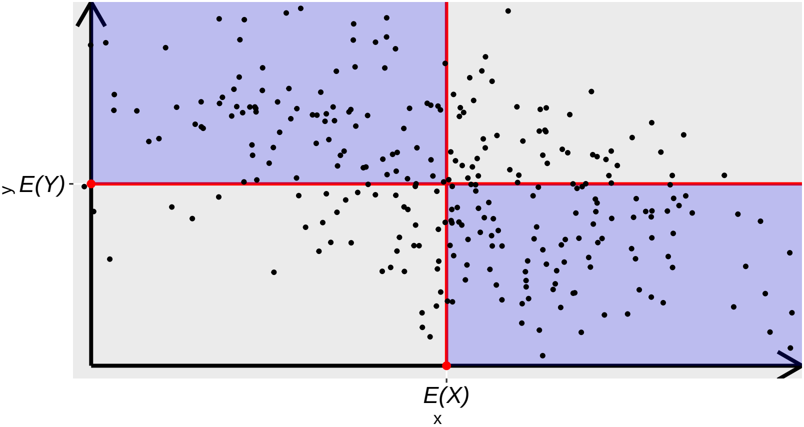


Figure: Data from (X, Y), with $Cov(X, Y) < 0$

Covariance of r.v.s, Interpretation

Note: Please note that I am a little bit cheating when saying that if the Mean of r.v. is positive, then mostly, that r.v. is positive. That is, if $\mathbb{E}(Y) > 0$, that doesn't mean that $Y > 0$ most of the time, with high Probability. In fact, we can have have the following case:

$$Y \sim \begin{pmatrix} -1 & 10000000 \\ 0.99 & 0.01 \end{pmatrix}$$

Clearly, $\mathbb{E}(Y) > 0$, but, with high Probability, $Y < 0$:

$$\mathbb{P}(Y < 0) = 0.99.$$

Covariance of r.v.s, Calculation formulas

Recall the definition of the Covariance:

$$Cov(X, Y) = \mathbb{E}\Big((X - \mathbb{E}(X)) \cdot (Y - \mathbb{E}(Y))\Big).$$

The following formula gives another, simpler method to calculate the Covariance:

Calculation of Covariance

$$Cov(X, Y) = \mathbb{E}(X \cdot Y) - \mathbb{E}(X) \cdot \mathbb{E}(Y).$$

Covariance of r.v.s, Calculation formulas

Now, we can calculate the Covariance for 2 r.v.s:

- **Discrete Case:** If X and Y are Discrete, then

$$Cov(X, Y) = \sum_{i,j} x_i \cdot y_j \cdot \mathbb{P}(X = x_i, Y = y_j) - \mathbb{E}(X) \cdot \mathbb{E}(Y);$$

We can calculate the product $\mathbb{E}(X) \cdot \mathbb{E}(Y)$ in two (equivalent) ways: either by first calculating the Marginal PMFs, and then calculating

$$\mathbb{E}(X) = \sum_{i} x_i \cdot \mathbb{P}(X = x_i), \qquad \mathbb{E}(Y) = \sum_{j} y_j \cdot \mathbb{P}(Y = y_j),$$

or we can avoid the calculation of the Marginals by using

$$\mathbb{E}(X) = \sum_{i,j} x_i \cdot \mathbb{P}(X = x_i, Y = y_j), \quad \mathbb{E}(Y) = \sum_{i,j} y_j \cdot \mathbb{P}(X = x_i, Y = y_j).$$

Covariance of r.v.s, Calculation formulas

- **Continuous Case:** If X and Y are Continuous, with the Joint PDF $f(x,y)$, then

$$Cov(X,Y) = \iint_{\mathbb{R}^2} x \cdot y \cdot f(x,y)\, dxdy - \mathbb{E}(X) \cdot \mathbb{E}(Y).$$

  Here again we can calculate $\mathbb{E}(X) \cdot \mathbb{E}(Y)$ in two ways:
  - By first calculating the Marginal PDFs $f_X(x)$ and $f_Y(y)$, and then

$$\mathbb{E}(X) = \int_{\mathbb{R}} x \cdot f_X(x)\, dx, \qquad \mathbb{E}(Y) = \int_{\mathbb{R}} y \cdot f_Y(y)\, dy;$$

  - By avoiding the calculation of Marginals and just using

$$\mathbb{E}(X) = \iint_{\mathbb{R}^2} x \cdot f(x,y)\, dxdy, \qquad \mathbb{E}(Y) = \iint_{\mathbb{R}^2} y \cdot f(x,y)\, dxdy.$$

Covariance, Example

Example 36.1: Calculate the Covariance of X and Y, if X and Y are Jointly Discrete r.v.s with the following Joint PMF:

$Y \setminus X$	-1	5
0	0.1	0.2
4	0.3	0.4

Covariance, Example

Example 36.2: Calculate the Covariance of X and Y, if X and Y are Jointly Continuous with the following Joint PDF:

$$f(x, y) = \begin{cases} K \cdot (2x + 3y), & x, y \in [0, 1] \\ 0, & \text{otherwise.} \end{cases}$$

Covariance, Example

Example 36.3: Calculate the Covariance of X and Y, if $(X, Y) \sim Unif(D)$, where

a. $D = [0, 2] \times [3, 9]$;

b. D is the triangle with vertices at $(0, 0), (1, 0), (0, 1)$.

Properties of the Covariance

Covariance Properties

Assume X, Y, Z,... are r.v.s, $\alpha, \beta, \dots$ are any real numbers. Then

- $Cov(X, c) = 0$, if $c = const$;
- $Cov(X, Y) = Cov(Y, X)$;
- $Cov(X, X) = Var(X)$;
- $Cov(\alpha X + \beta Y, Z) = \alpha Cov(X, Z) + \beta Cov(Y, Z)$;
- $$Cov\left(\sum_{i=1}^{n} \alpha_i X_i, \sum_{j=1}^{m} \beta_j Y_j\right) = \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_i \beta_j \cdot Cov(X_i, Y_j);$$

Covariance Properties, Cont'd

- If $X \perp\!\!\!\perp Y$, then $Cov(X, Y) = 0$. The inverse is **NOT TRUE**, in general;
- $$Var(X + Y) = Var(X) + 2Cov(X, Y) + Var(Y)$$

  and

  $$Var(X - Y) = Var(X) - 2Cov(X, Y) + Var(Y);$$

  In particular,

  $$Var(X \pm Y) = Var(X) + Var(Y) \qquad \text{iff} \qquad Cov(X, Y) = 0.$$

Covariance Properties, Cont'd

- $$Var(X_1 + X_2 + ... + X_n) = \sum_{i,j=1}^{n} Cov(X_i, X_j) =$$
  $$= \sum_{i=1}^{n} Var(X_i) + 2 \cdot \sum_{i<j} Cov(X_i, X_j).$$

- $$|Cov(X, Y)| \leq \sqrt{Var(X)} \cdot \sqrt{Var(Y)}.$$

Covariance Properties, Example

Example 36.4: Assume $X \sim Pois(2.5)$. Calculate

$$Cov(X, 2X - 1).$$

Covariance Properties, Example

Example 36.5: Calculate $Cov(2X, 3X - Y)$, if $SD(X) = 2$, $Cov(Y, X) = 0.5$.

Covariance Properties, Example

Example 36.6: Assume $Var(X) = 2$, $Var(Y) = 1$. Is it possible that $Cov(X, Y) = 2$?

Covariance Properties, Example

Example 36.7: Calculate

$$Var\Big(X - 3Y + \mathbb{E}(X)\Big),$$

if $Var(X) = Var(Y) = 2$, and $Cov(X, Y) = 1.5$.

Covariance Properties, Example

Example 36.8: Calculate

$$Var(2X - 5Y + Z + 2\pi),$$

if $X \perp\!\!\!\perp Z$, $Y \perp\!\!\!\perp Z$, $Var(X) = 1$, $Var(Y) = 2$, $Var(Z) = 3$ and $Cov(X, Y) = 1$.

Covariance Properties, Example

Example 36.9: Assume $X \sim Unif[-1, 1]$ and $Y = X^2$. Clearly, Y depends on X, so X and Y are not Independent. Show that X and Y are Uncorrelated:

$$Cov(X, Y) = 0.$$

Note: Above, instead of $Unif[-1, 1]$, one can take any Continuous r.v. with symmetric PDF around 0.

Correlation Coefficient and its Properties

Correlation of R.V.s, Intro

We learned that the Covariance measures the relationship (linear relationship!) between the r.v.s. But this measure is not so convenient from several points of view:

- Its value depends on the units we choose:
  Example: Say, we are calculating the height H of a person in meters, and the weight W in Kgs, and we are interested in the Covariance of H and W, $Cov(H, W)$. Now, if we will write the same person's height in centimeters, $h = 100H$ and the weight in grams, $w = 1000\,W$, then the Covariance between h and w will be:

  $$Cov(h, w) = Cov(100H, 1000\,W) = 10^5 \cdot Cov(H, W)$$

  So Covariance will not describe the relationship between the Height and Weight, rather, it will describe the relationship between the Height and Weight in Kg and M, or relationship between the Height and Weight in g and cm

Correlation of R.V.s, Intro

- Exactly on the same basis, we cannot compare two different r.v. pairs relationships using Covariances, say, we cannot say that the relationship between X and Y is stronger than the relationship between Z and T, if, say, $|Cov(X, Y)| > |Cov(Z, T)|$.

For the previous example, we would like to report the relationship between the Height and Weight, not different values for different units of measure.
So we need to introduce another measure for the relationship between r.v.s, which is, as we will see, scale/unit-invariant.

Correlation of R.V.s

Correlation of R.V.s

If X and Y are non-constant r.v.s, then the **Pearson's Correlation Coefficient** of X and Y is defined as

$$Cor(X, Y) = \rho_{XY} = \frac{Cov(X, Y)}{\sqrt{Var(X)} \cdot \sqrt{Var(Y)}}.$$

Note: Like in the case of Variance-Standard Deviation pair, here the situation is the same: Covariance is easy to handle with, and the Correlation Coefficient is used for reporting.

Correlation of R.V.s

Note: It is easy to see that we can obtain the Correlation Coefficient $Cor(X, Y)$ in the following way: given X, Y, first we Standardize them, calculate

$$\frac{X - \mathbb{E}(X)}{\sqrt{Var(X)}} \qquad \text{and} \qquad \frac{Y - \mathbb{E}(Y)}{\sqrt{Var(Y)}},$$

then we calculate the Covariance of this Standardized r.v.s:

$$Cov\left(\frac{X - \mathbb{E}(X)}{\sqrt{Var(X)}}, \frac{Y - \mathbb{E}(Y)}{\sqrt{Var(Y)}}\right) = Cor(X, Y).$$

Properties of a Correlation

- $Cor(X, Y) = Cor(Y, X)$;
- $Cor(X, X) = 1$;
- For any non-constant r.v.s X, Y,
$$-1 \leq Cor(X, Y) \leq 1;$$
- Correlation is dimensionless and Scale-invariant:
$$Cor(\alpha X + a, \beta Y + b) = sgn(\alpha \cdot \beta) \cdot Cor(X, Y);$$
- Non-constant r.v.s X and Y are uncorrelated iff $Cor(X, Y) = 0$.

Correlation: Example

Example 36.10: Assume X and Y are Jointly Continuous with the Joint PDF

$$f(x, y) = \begin{cases} K \cdot xy, & x, y \in [0, 2], x + y \leq 2 \\ 0, & \text{otherwise.} \end{cases}$$

Find $Cor(X, Y)$.

Properties of a Correlation

Above we stated that

$$-1 \leq Cor(X, Y) \leq 1.$$

Now, in addition to this, we have:

- $Cor(X, Y) = 1$ if and only if there exist constants $\alpha > 0$ and β such that
$$Y = \alpha \cdot X + \beta.$$

- $Cor(X, Y) = -1$ if and only if there exist constants $\alpha < 0$ and β such that
$$Y = \alpha \cdot X + \beta.$$

Comparison of Covariance and Correlation

- Covariance is linear, correlation is not;
- Correlation is scale-invariant, Covariance is not;
- If $|Cov(X, Y)| > |Cov(Z, T)|$, we cannot state that the relationship between X and Y is stronger than the relationship between Z and T. But if $|Cor(X, Y)| > |Cor(Z, T)|$, we can.

  Note: So it is not so easy to interpret the magnitude of the covariance, but the magnitude of the correlation coefficient is the strength of the linear relationship between r.v.s.

Properties of a Correlation

Note: Very common mistake is that people are confusing the linear relationship slope α with the Correlation coefficient ρ. Say, think about the following question: for which case the Correlation $Cor(X, Y)$ is larger:

$$Y = 0.2 \cdot X + 1 \qquad \text{or} \qquad Y = 10 \cdot X + 1 \ ?$$

Well, of course, the answer is that the Correlation is the same for both cases, $Cor(X, Y) = 1$.

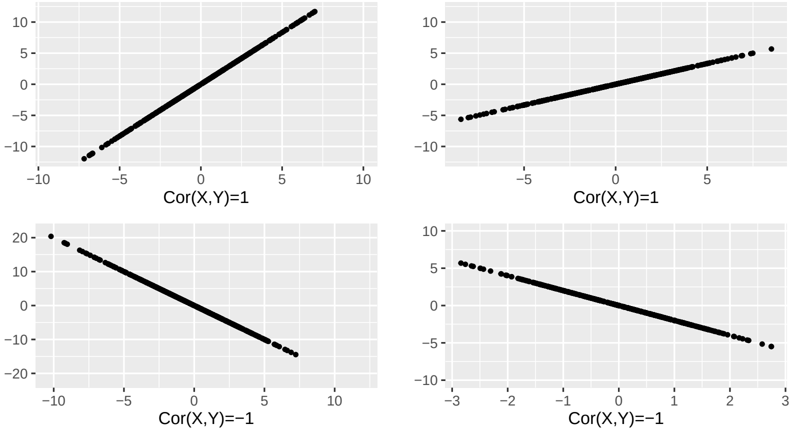
Properties of a Correlation

So $|Cor(X, Y)|$ shows how close is the relationship between X and Y to the linear relationship:

- If $|Cor(X, Y)| = 1$, then we have exactly (perfect) linear relationship between X and Y;
- If $|Cor(X, Y)|$ is close to 1, then we have almost linear relationship between X and Y;
- If $|Cor(X, Y)|$ is substantially smaller than 1, then we have weak linear relationship between X and Y;
- If $Cor(X, Y) = 0$, then we do not have a linear relationship between X and Y.

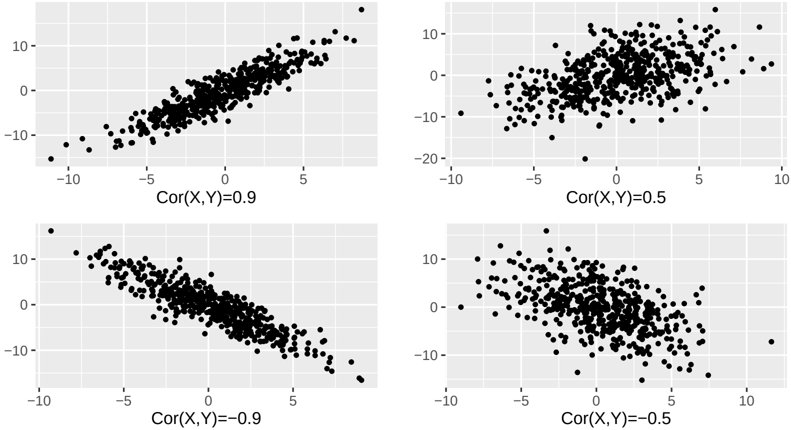
Correlation Coefficient, Interpretation

Example:



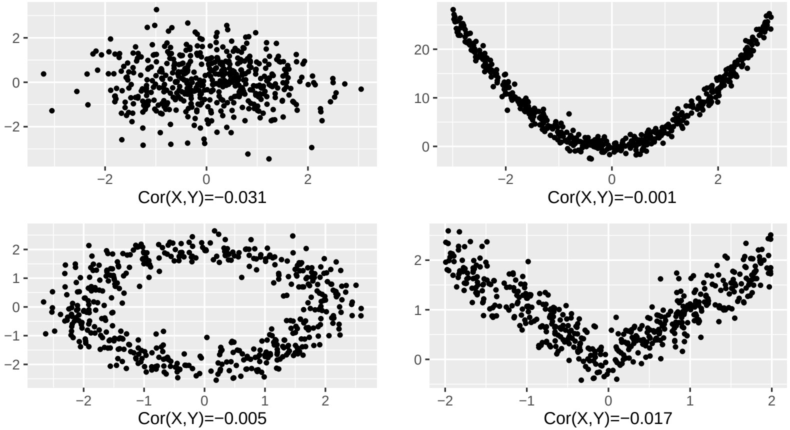
Correlation Coefficient, Interpretation

Example:



Correlation Coefficient, Interpretation

Example:



Covariance Matrix

Covariance Matrix

If we have r.v.s $X_1, X_2, ..., X_n$, then we can make the Covariance Matrix:

Covariance Matrix

The following Matrix is called the Covariance Matrix of $X_1, X_2, ..., X_n$:

$$\Sigma = \begin{bmatrix} Cov(X_1, X_1) & \dots & Cov(X_1, X_n) \\ \vdots & \ddots & \vdots \\ Cov(X_n, X_1) & \dots & Cov(X_n, X_n) \end{bmatrix}$$

Note: It can be proved that the Covariance Matrix is always Symmetric and Positive Semi-Definite (or Non-Negative Definite).

Covariance Matrix

Recall that when talking about the Mutivariate Normal Distribution

$$(X_1, X_2, ..., X_n) \sim \mathcal{N}(\mu, \Sigma),$$

we have called Σ a Covariance Matrix: it turns out that Σ is exactly the Covariance Matrix of $X_1, ..., X_n$. In particular,

- If
$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 4 & 2 \\ 2 & 5 \end{bmatrix} \right),$$
then $Var(X) = 4$, $Var(Y) = 5$ and $Cov(X, Y) = 2$
- If (X, Y) are Jointly Normally Distributed, $\mathbb{E}(X) = 3, \mathbb{E}(Y) = -1$, and $Var(X) = 8$, $Var(Y) = 5$ and $Cov(X, Y) = 1$, then
$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 3 \\ -1 \end{bmatrix}, \begin{bmatrix} 8 & 1 \\ 1 & 5 \end{bmatrix} \right).$$