Some reminders on statistics

Copernicus master
2025–2026

Thomas Corpetti

Outline

Outline

Outline

Random Variable

- A **random variable** : a real-valued function that depends on the outcome of an experiment
  - Example : grades in a class : $\mathcal{X} = \{0, 1, 2, 3, ..., 20\}$
  - Example : price of a liter of milk : $\mathcal{X} \in [0, 10]$
  - ...
  - It is denoted as $\mathcal{X} = \{x_1, ..., x_P\}$
- The **probability distribution** of a random variable is a pair that associates a value x_i of the random variable with the probability p_i that an experiment on this random variable takes that value
  - Example : grades in a class : $\{(0, 0.01), ..., (9, 0.2), (10, 0.4), (11, 0.2), (20, 0.01)\}$
  - Example : price of a liter of milk : Gaussian centered at 0.35 euros
  - ...

Random Variable

- A **random variable** : a real-valued function that depends on the outcome of an experiment
  - Example : grades in a class : $\mathcal{X} = \{0, 1, 2, 3, ..., 20\}$
  - Example : price of a liter of milk : $\mathcal{X} \in [0, 10]$
  - ...
  - It is denoted as $\mathcal{X} = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_P\}$
- The **probability distribution** of a random variable is a pair that associates a value x_i of the random variable with the probability p_i that an experiment on this random variable takes that value
  - Example : grades in a class : $\{(0, 0.01), ..., (9, 0.2), (10, 0.4), (11, 0.2), (20, 0.01)\}$
  - Example : price of a liter of milk : Gaussian centered at 0.35 euros
  - ...

Random Variable

- A **random variable** : a real-valued function that depends on the outcome of an experiment
    - Example : grades in a class : $\mathcal{X} = \{0, 1, 2, 3, ..., 20\}$
    - Example : price of a liter of milk : $\mathcal{X} \in [0, 10]$
    - ...
    - It is denoted as $\mathcal{X} = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_P\}$
- The **probability distribution** of a random variable is a pair that associates a value x_i of the random variable with the probability p_i that an experiment on this random variable takes that value
    - Example : grades in a class : $\{(0, 0.01), ..., (9, 0.2), (10, 0.4), (11, 0.2), (20, 0.01)\}$
    - Example : price of a liter of milk : Gaussian centered at 0.35 euros
    - ...

Outline

Expectation of a Random Variable

- **Expectation** of a random variable defined on a sample space Ω is its *mean* (first moment) :

$$E(X) = \int_{\Omega} \boldsymbol{x} p(\mathcal{X} = \boldsymbol{x}) d\boldsymbol{x}$$

- **Variance** is a measure characterizing the spread of a sample or distribution. It is calculated from the expectation as $E(\mathcal{X}) = \mu$ (second moment) :

$$\text{var}(\mathcal{X}) = \int_{\Omega} (\boldsymbol{x} - \mu)^2 p(\mathcal{X} = \boldsymbol{x}) d\boldsymbol{x}$$

$$\text{var}(\mathcal{X}) = E\left[(\mathcal{X} - E(\mathcal{X}))^2\right] = \underbrace{E\left[\mathcal{X}^2\right] - E\left[\mathcal{X}\right]^2}_{\text{König-Huyghens}}$$

- Definition of **standard deviation** from variance : $\sigma_{\mathcal{X}} = \sqrt{\text{var}(\mathcal{X})}$

Expectation of a Random Variable

- **Expectation** of a random variable defined on a sample space Ω is its *mean* (first moment) :

$$E(X) = \int_\Omega \boldsymbol{x} p(\mathcal{X} = \boldsymbol{x}) d\boldsymbol{x}$$

- **Variance** is a measure characterizing the spread of a sample or distribution. It is calculated from the expectation as $E(\mathcal{X}) = \mu$ (second moment) :

$$\text{var}(\mathcal{X}) = \int_\Omega (\boldsymbol{x} - \mu)^2 p(\mathcal{X} = \boldsymbol{x}) dx$$

$$\text{var}(\mathcal{X}) = E\left[(\mathcal{X} - E(\mathcal{X}))^2\right] = \underbrace{E\left[\mathcal{X}^2\right] - E\left[\mathcal{X}\right]^2}_{\text{König-Huyghens}}$$

- Definition of **standard deviation** from variance : $\sigma_\mathcal{X} = \sqrt{\text{var}(\mathcal{X})}$

Expectation of a Random Variable

- **Expectation** of a random variable defined on a sample space Ω is its *mean* (first moment) :

$$E(X) = \int_\Omega \boldsymbol{x} p(\mathcal{X} = \boldsymbol{x}) d\boldsymbol{x}$$

- **Variance** is a measure characterizing the spread of a sample or distribution. It is calculated from the expectation as $E(\mathcal{X}) = \mu$ (second moment) :

$$\text{var}(\mathcal{X}) = \int_\Omega (\boldsymbol{x} - \mu)^2 p(\mathcal{X} = \boldsymbol{x}) dx$$

$$\text{var}(\mathcal{X}) = E\left[(\mathcal{X} - E(\mathcal{X}))^2\right] = \underbrace{E\left[\mathcal{X}^2\right] - E\left[\mathcal{X}\right]^2}_{\text{König-Huyghens}}$$

- Definition of **standard deviation** from variance : $\sigma_\mathcal{X} = \sqrt{\text{var}(\mathcal{X})}$

Expectation of a Random Variable

- **Expectation** of a random variable defined on a sample space Ω is its *mean* (first moment) :

$$E(X) = \int_{\Omega} \boldsymbol{x} p(\mathcal{X} = \boldsymbol{x}) d\boldsymbol{x}$$

- **Variance** is a measure characterizing the spread of a sample or distribution. It is calculated from the expectation as $E(\mathcal{X}) = \mu$ (second moment) :

$$\text{var}(\mathcal{X}) = \int_{\Omega} (\boldsymbol{x} - \mu)^2 p(\mathcal{X} = \boldsymbol{x}) dx$$

$$\text{var}(\mathcal{X}) = E\left[(\mathcal{X} - E(\mathcal{X}))^2\right] = \underbrace{E\left[\mathcal{X}^2\right] - E\left[\mathcal{X}\right]^2}_{\text{König-Huyghens}}$$

- Definition of **standard deviation** from variance : $\sigma_{\mathcal{X}} = \sqrt{\text{var}(\mathcal{X})}$

Variable centrée/réduite

- A variable X is **centered and reduced** if its mean is zero, and its variance and standard deviation are both 1.

$\Longrightarrow$ To center and reduce a variable Y, you apply

$$\frac{Y - E(Y)}{\sqrt{\text{var}(Y)}} = \frac{Y - E(Y)}{\sigma_Y}$$

Variable centrée/réduite

- A variable X is **centered and reduced** if its mean is zero, and its variance and standard deviation are both 1.

  $\Longrightarrow$ To center and reduce a variable Y, you apply

$$\frac{Y - E(Y)}{\sqrt{\text{var}(Y)}} = \frac{Y - E(Y)}{\sigma_Y}$$

Plan

Card Game

- We have a deck of 32 cards. A player places a bet and draws a card at random.
  - If they draw an ace, they **win 3 times** their bet.
  - If they draw a king, they **win 2 times** their bet.
  - If they draw any other card, they **lose** their bet.
- Let X be the variable associated with the player's winnings.
  - What values can X take ?
  - What is its probability distribution ?
  - What are its expectations, variances, and standard deviations ?

Card Game

- We have a deck of 32 cards. A player places a bet and draws a card at random.
  - If they draw an ace, they **win 3 times** their bet.
  - If they draw a king, they **win 2 times** their bet.
  - If they draw any other card, they **lose** their bet.
- Let X be the variable associated with the player's winnings.
  - What values can X take ?
  - What is its probability distribution ?
  - What are its expectations, variances, and standard deviations ?

Plan

Expectation of Expectations

- Notations
  - X is a random variable
  - $(x_1, ..., x_P)$ is a P-tuple corresponding to values taken by this random variable for P individuals

$\Longrightarrow$ We consider this P-tuple as the realization of a random vector $(X_1, ..., X_P)$ where all X_i follow the same distribution as X

- Estimation Case
  - We extract a sample of size P, which is considered as realizations of P independent random variables $\mathcal{X}_P = (X_1, ..., X_P)$ following the distribution of X
  - We have

$$E(\mathcal{X}_P) = E\left(\frac{1}{P}\sum_{i=1}^{P} X_i\right) = E(X) \;\; ; \;\; \text{var}(\mathcal{X}_P) = \frac{1}{P}V(X)$$

$\Longrightarrow$ The precision of estimating the mean increases with the number of realizations

  - θ is an unknown parameter of a random variable (mean, variance,

Expectation of Expectations

- Notations
  - X is a random variable
  - $(x_1, ..., x_P)$ is a P-tuple corresponding to values taken by this random variable for P individuals

$\Longrightarrow$ We consider this P-tuple as the realization of a random vector $(X_1, ..., X_P)$ where all X_i follow the same distribution as X

- Estimation Case
  - We extract a sample of size P, which is considered as realizations of P independent random variables $\mathcal{X}_P = (X_1, ..., X_P)$ following the distribution of X
  - We have

$$E(\mathcal{X}_P) = E\left(\frac{1}{P}\sum_{i=1}^{P} X_i\right) = E(X) \;\; ; \;\; \text{var}(\mathcal{X}_P) = \frac{1}{P}V(X)$$

$\Longrightarrow$ The precision of estimating the mean increases with the number of realizations

  - θ is an unknown parameter of a random variable (mean, variance,

Expectation of Expectations

- Notations
  - X is a random variable
  - $(x_1, ..., x_P)$ is a P-tuple corresponding to values taken by this random variable for P individuals

$\Longrightarrow$ We consider this P-tuple as the realization of a random vector $(X_1, ..., X_P)$ where all X_i follow the same distribution as X

- Estimation Case
  - We extract a sample of size P, which is considered as realizations of P independent random variables $\mathcal{X}_P = (X_1, ..., X_P)$ following the distribution of X
  - We have

$$E(\mathcal{X}_P) = E\left(\frac{1}{P}\sum_{i=1}^{P} X_i\right) = E(X) \;\; ; \;\; \text{var}(\mathcal{X}_P) = \frac{1}{P}V(X)$$

$\Longrightarrow$ The precision of estimating the mean increases with the number of realizations

  - θ is an unknown parameter of a random variable (mean, variance,

Expectation of Expectations

- Notations
  - X is a random variable
  - $(x_1, ..., x_P)$ is a P-tuple corresponding to values taken by this random variable for P individuals

$\Longrightarrow$ We consider this P-tuple as the realization of a random vector $(X_1, ..., X_P)$ where all X_i follow the same distribution as X

- Estimation Case
  - We extract a sample of size P, which is considered as realizations of P independent random variables $\mathcal{X}_P = (X_1, ..., X_P)$ following the distribution of X
  - We have

$$E(\mathcal{X}_P) = E\left(\frac{1}{P}\sum_{i=1}^{P} X_i\right) = E(X) \quad ; \quad \text{var}(\mathcal{X}_P) = \frac{1}{P}V(X)$$

$\Longrightarrow$ The precision of estimating the mean increases with the number of realizations

  - θ is an unknown parameter of a random variable (mean, variance,

Expectation of Expectations

- Notations
  - X is a random variable
  - $(x_1, ..., x_P)$ is a P-tuple corresponding to values taken by this random variable for P individuals

$\Longrightarrow$ We consider this P-tuple as the realization of a random vector $(X_1, ..., X_P)$ where all X_i follow the same distribution as X

- Estimation Case
  - We extract a sample of size P, which is considered as realizations of P independent random variables $\mathcal{X}_P = (X_1, ..., X_P)$ following the distribution of X
  - We have

$$E(\mathcal{X}_P) = E\left(\frac{1}{P}\sum_{i=1}^{P} X_i\right) = E(X) \quad ; \quad \text{var}(\mathcal{X}_P) = \frac{1}{P}V(X)$$

$\Longrightarrow$ The precision of estimating the mean increases with the number of realizations

  - θ is an unknown parameter of a random variable (mean, variance,

Expectation of Expectations

- Notations
  - X is a random variable
  - $(x_1, ..., x_P)$ is a P-tuple corresponding to values taken by this random variable for P individuals

$\Longrightarrow$ We consider this P-tuple as the realization of a random vector $(X_1, ..., X_P)$ where all X_i follow the same distribution as X

- Estimation Case
  - We extract a sample of size P, which is considered as realizations of P independent random variables $\mathcal{X}_P = (X_1, ..., X_P)$ following the distribution of X
  - We have

$$E(\mathcal{X}_P) = E\left(\frac{1}{P}\sum_{i=1}^{P} X_i\right) = E(X) \;\; ; \;\; \text{var}(\mathcal{X}_P) = \frac{1}{P}V(X)$$

$\Longrightarrow$ The precision of estimating the mean increases with the number of realizations

  - θ is an unknown parameter of a random variable (mean, variance,

Expectation of Expectations

- Notations
  - X is a random variable
  - $(x_1, ..., x_P)$ is a P-tuple corresponding to values taken by this random variable for P individuals

$\Longrightarrow$ We consider this P-tuple as the realization of a random vector $(X_1, ..., X_P)$ where all X_i follow the same distribution as X

- Estimation Case
  - We extract a sample of size P, which is considered as realizations of P independent random variables $\mathcal{X}_P = (X_1, ..., X_P)$ following the distribution of X
  - We have

$$E(\mathcal{X}_P) = E\left(\frac{1}{P}\sum_{i=1}^{P} X_i\right) = E(X) \quad ; \quad \text{var}(\mathcal{X}_P) = \frac{1}{P}V(X)$$

$\Longrightarrow$ The precision of estimating the mean increases with the number of realizations

- θ is an unknown parameter of a random variable (mean, variance,

Expectation of Expectations

- Notations
  - X is a random variable
  - $(x_1, ..., x_P)$ is a P-tuple corresponding to values taken by this random variable for P individuals

$\Longrightarrow$ We consider this P-tuple as the realization of a random vector $(X_1, ..., X_P)$ where all X_i follow the same distribution as X

- Estimation Case
  - We extract a sample of size P, which is considered as realizations of P independent random variables $\mathcal{X}_P = (X_1, ..., X_P)$ following the distribution of X
  - We have

$$E(\mathcal{X}_P) = E\left(\frac{1}{P}\sum_{i=1}^{P} X_i\right) = E(X) \quad ; \quad \text{var}(\mathcal{X}_P) = \frac{1}{P}V(X)$$

$\Longrightarrow$ The precision of estimating the mean increases with the number of realizations

- θ is an unknown parameter of a random variable (mean, variance,

Expectation of Expectations

- Notations
  - X is a random variable
  - $(x_1, ..., x_P)$ is a P-tuple corresponding to values taken by this random variable for P individuals

$\Longrightarrow$ We consider this P-tuple as the realization of a random vector $(X_1, ..., X_P)$ where all X_i follow the same distribution as X

- Estimation Case
  - We extract a sample of size P, which is considered as realizations of P independent random variables $\mathcal{X}_P = (X_1, ..., X_P)$ following the distribution of X
  - We have

$$E(\mathcal{X}_P) = E\left(\frac{1}{P}\sum_{i=1}^{P} X_i\right) = E(X) \quad ; \quad \text{var}(\mathcal{X}_P) = \frac{1}{P}V(X)$$

$\Longrightarrow$ The precision of estimating the mean increases with the number of realizations

  - θ is an unknown parameter of a random variable (mean, variance,

Table of Contents

Case of Two Random Variables

- Let $X = \{x_1, \ldots, x_N\}$ and $Y = \{y_1, \ldots, y_N\}$ be two realizations of random variables (for simplicity, we'll refer to X and Y as random variables) :
- We can question how X and Y are related.

  $\Longrightarrow$ **Covariance** : a measure of the joint variation of 2 variables

$$\text{cov}(X, Y) = E\big[[X - E[X]][Y - E[Y]]\big]$$

- We have $\text{var}(X) = \text{cov}(X, X)$
- We have $\text{cov}(X, Y) = E(XY) - E(X)E(Y) \Longrightarrow$ Covariance is zero for two independent random variables
- Covariance becomes positive if we have many pairs of values that deviate from their means in the same direction, and vice versa.

Case of Two Random Variables

- Let $X = \{x_1, \ldots, x_N\}$ and $Y = \{y_1, \ldots, y_N\}$ be two realizations of random variables (for simplicity, we'll refer to X and Y as random variables) :
- We can question how X and Y are related.

$\Longrightarrow$ **Covariance** : a measure of the joint variation of 2 variables

$$\text{cov}(X, Y) = E\big[[X - E[X]][Y - E[Y]]\big]$$

- We have $\text{var}(X) = \text{cov}(X, X)$
- We have $\text{cov}(X, Y) = E(XY) - E(X)E(Y) \Longrightarrow$ Covariance is zero for two independent random variables
- Covariance becomes positive if we have many pairs of values that deviate from their means in the same direction, and vice versa.

Case of Two Random Variables

- Let $X = \{x_1, \ldots, x_N\}$ and $Y = \{y_1, \ldots, y_N\}$ be two realizations of random variables (for simplicity, we'll refer to X and Y as random variables) :
- We can question how X and Y are related.

$\Longrightarrow$ **Covariance** : a measure of the joint variation of 2 variables

$$\text{cov}(X, Y) = E\big[[X - E[X]][Y - E[Y]]\big]$$

- We have $\text{var}(X) = \text{cov}(X, X)$
- We have $\text{cov}(X, Y) = E(XY) - E(X)E(Y) \Longrightarrow$ Covariance is zero for two independent random variables
- Covariance becomes positive if we have many pairs of values that deviate from their means in the same direction, and vice versa.

Case of Two Random Variables

- Let $X = \{x_1, \ldots, x_N\}$ and $Y = \{y_1, \ldots, y_N\}$ be two realizations of random variables (for simplicity, we'll refer to X and Y as random variables) :
- We can question how X and Y are related.

  $\Longrightarrow$ **Covariance** : a measure of the joint variation of 2 variables

$$\mathrm{cov}(X, Y) = E\big[[X - E[X]][Y - E[Y]]\big]$$

- We have $\mathrm{var}(X) = \mathrm{cov}(X, X)$
- We have $\mathrm{cov}(X, Y) = E(XY) - E(X)E(Y) \Longrightarrow$ Covariance is zero for two independent random variables
- Covariance becomes positive if we have many pairs of values that deviate from their means in the same direction, and vice versa.

Case of Two Random Variables

- Let $X = \{x_1, \ldots, x_N\}$ and $Y = \{y_1, \ldots, y_N\}$ be two realizations of random variables (for simplicity, we'll refer to X and Y as random variables) :
- We can question how X and Y are related.

$\Longrightarrow$ **Covariance** : a measure of the joint variation of 2 variables

$$\text{cov}(X, Y) = E\Big[[X - E[X]][Y - E[Y]]\Big]$$

- We have $\text{var}(X) = \text{cov}(X, X)$
- We have $\text{cov}(X, Y) = E(XY) - E(X)E(Y) \Longrightarrow$ Covariance is zero for two independent random variables
- Covariance becomes positive if we have many pairs of values that deviate from their means in the same direction, and vice versa.

Case of Two Random Variables

- Let $X = \{x_1, \ldots, x_N\}$ and $Y = \{y_1, \ldots, y_N\}$ be two realizations of random variables (for simplicity, we'll refer to X and Y as random variables) :
- We can question how X and Y are related.

$\Longrightarrow$ **Covariance** : a measure of the joint variation of 2 variables

$$\text{cov}(X, Y) = E\big[[X - E[X]][Y - E[Y]]\big]$$

- We have $\text{var}(X) = \text{cov}(X, X)$
- We have $\text{cov}(X, Y) = E(XY) - E(X)E(Y) \Longrightarrow$ Covariance is zero for two independent random variables
- Covariance becomes positive if we have many pairs of values that deviate from their means in the same direction, and vice versa.

Outline

Coefficient of Correlation

- If the units of X and Y are entirely different (e.g., age of consumers, price of milk per liter), it can be challenging to assess the relationship between the two sets of data.

  $\Longrightarrow$ Necessity to standardize the data

- We can center and reduce the random variables (normalize them).
- We can use the **correlation coefficient** : a measure of the linear relationship normalized by the standard deviations of the variables.

$$\text{corr}(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$$

Coefficient of Correlation

- If the units of X and Y are entirely different (e.g., age of consumers, price of milk per liter), it can be challenging to assess the relationship between the two sets of data.

  $\Longrightarrow$ Necessity to standardize the data
- We can center and reduce the random variables (normalize them).
- We can use the **correlation coefficient** : a measure of the linear relationship normalized by the standard deviations of the variables.

$$\mathrm{corr}(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y}$$

Coefficient of Correlation

- If the units of X and Y are entirely different (e.g., age of consumers, price of milk per liter), it can be challenging to assess the relationship between the two sets of data.

  $\Longrightarrow$ Necessity to standardize the data

- We can center and reduce the random variables (normalize them).
- We can use the **correlation coefficient** : a measure of the linear relationship normalized by the standard deviations of the variables.

$$\text{corr}(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$$

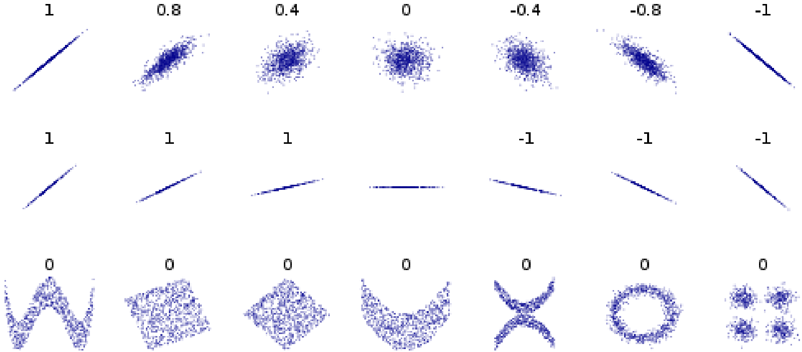
Coefficient of Correlation

- If the units of X and Y are entirely different (e.g., age of consumers, price of milk per liter), it can be challenging to assess the relationship between the two sets of data.

  $\Longrightarrow$ Necessity to standardize the data
- We can center and reduce the random variables (normalize them).
- We can use the **correlation coefficient** : a measure of the linear relationship normalized by the standard deviations of the variables.

$$\mathrm{corr}(X, Y) = \frac{\mathrm{cov}(X, Y)}{\sigma_X \sigma_Y}$$

Illustration



Source : Wikipedia

Multiple Random Variables

- Let $\{X_1, ..., X_N\}$ be N random variables. The covariance matrix is

$$\Sigma_X = M = \begin{bmatrix} \text{var}(X_1) & \text{cov}(X_1, X_2) & ... & \text{cov}(X_1, X_N) \\ \text{cov}(X_2, X_1) & \text{var}(X_2) & ... & \text{cov}(X_2, X_N) \\ ... & ... & ... & ... \\ \text{cov}(X_N, X_1) & \text{cov}(X_N, X_2) & ... & \text{var}(X_N) \end{bmatrix}$$

- **Practical Case** : If X is a $P \times N$ matrix representing P realizations of the N centered random variables, the estimation of the **empirical variance-covariance matrix** is

$$\Sigma_X \approx X^T X$$

- **Correlation Matrix** : A matrix of variance-covariance on standardized variables (centered and reduced).

Multiple Random Variables

- Let $\{X_1, ..., X_N\}$ be N random variables. The covariance matrix is

$$\Sigma_X = M = \begin{bmatrix} \mathrm{var}(X_1) & \mathrm{cov}(X_1, X_2) & \ldots & \mathrm{cov}(X_1, X_N) \\ \mathrm{cov}(X_2, X_1) & \mathrm{var}(X_2) & \ldots & \mathrm{cov}(X_2, X_N) \\ \ldots & \ldots & \ldots & \ldots \\ \mathrm{cov}(X_N, X_1) & \mathrm{cov}(X_N, X_2) & \ldots & \mathrm{var}(X_N) \end{bmatrix}$$

- **Practical Case** : If X is a $P \times N$ matrix representing P realizations of the N centered random variables, the estimation of the **empirical variance-covariance matrix** is

$$\Sigma_X \approx X^T X$$

- **Correlation Matrix** : A matrix of variance-covariance on standardized variables (centered and reduced).

Multiple Random Variables

- Let $\{X_1, ..., X_N\}$ be N random variables. The covariance matrix is

$$\Sigma_X = M = \begin{bmatrix} \text{var}(X_1) & \text{cov}(X_1, X_2) & ... & \text{cov}(X_1, X_N) \\ \text{cov}(X_2, X_1) & \text{var}(X_2) & ... & \text{cov}(X_2, X_N) \\ ... & ... & ... & ... \\ \text{cov}(X_N, X_1) & \text{cov}(X_N, X_2) & ... & \text{var}(X_N) \end{bmatrix}$$

- **Practical Case** : If X is a $P \times N$ matrix representing P realizations of the N centered random variables, the estimation of the **empirical variance-covariance matrix** is

$$\Sigma_X \approx X^T X$$

- **Correlation Matrix** : A matrix of variance-covariance on standardized variables (centered and reduced).

Example / Exercise

- Consider a 3-dimensional dataset with

$$X = \begin{bmatrix} -2 & 3 & -1 \\ -1 & 1 & 0 \\ 2 & -1 & -1 \\ 1 & -3 & 2 \end{bmatrix}$$

1. Calculate the mean of the random variables.
2. Calculate the variance-covariance matrix.
3. Center and standardize the data.
4. Calculate the correlation matrix.