

# Mathematics for Machine Learning

— Probability & Distributions (Supplementary):

Sum Rule, Product Rule, Bayes' Theorem & Summary Statistics

Joseph Chuang-Chieh Lin

Department of Computer Science & Information Engineering,  
Tamkang University

Fall 2023

## Credits for the resource

- The slides are based on the textbooks:
  - *Marc Peter Deisenroth, A. Aldo Faisal, and Cheng Soon Ong: Mathematics for Machine Learning. Cambridge University Press. 2020.*
  - *Howard Anton, Chris Rorres, Anton Kaul: Elementary Linear Algebra. Wiley. 2019.*
- We could partially refer to the monograph:  
*Francesco Orabona: A Modern Introduction to Online Learning.*  
<https://arxiv.org/abs/1912.13213>

# Outline

- 1 Sum & Product Rule
- 2 Bayes' Theorem
- 3 Means & Covariances
- 4 Sums & Transformations of Random Variables
- 5 Statistical Independence

# Outline

- 1 Sum & Product Rule
- 2 Bayes' Theorem
- 3 Means & Covariances
- 4 Sums & Transformations of Random Variables
- 5 Statistical Independence

## Sum Rule (1/2)

- $\mathbf{x}, \mathbf{y}$ : random variables (vectors).
- $p(\mathbf{x}, \mathbf{y})$ : joint distribution of  $\mathbf{x}, \mathbf{y}$ .
- $p(\mathbf{y} \mid \mathbf{x})$ : conditional probability of  $\mathbf{y}$  given  $\mathbf{x}$ .

### Sum Rule

$$p(\mathbf{x}) = \begin{cases} \sum_{\mathbf{y} \in \mathcal{Y}} p(\mathbf{x}, \mathbf{y}) & \text{if } \mathbf{y} \text{ is discrete} \\ \int_{\mathcal{Y}} p(\mathbf{x}, \mathbf{y}) d\mathbf{y} & \text{if } \mathbf{y} \text{ is continuous} \end{cases}$$

where  $\mathcal{Y}$  stands for the states of the target space of random variable  $Y$ .

- Marginalization property.

## Sum Rule (2/2)

For  $\mathbf{x} = [x_1, \dots, x_D]^\top$ , the marginal

$$p(x_i) = \int p(x_1, \dots, x_D) d\mathbf{x}_{-i}$$

, where “ $-i$ ” means all except  $i$ .

# Product Rule

## Product Rule

$$p(\mathbf{x}, \mathbf{y}) = p(\mathbf{y} \mid \mathbf{x})p(\mathbf{x})$$

# Outline

- 1 Sum & Product Rule
- 2 Bayes' Theorem
- 3 Means & Covariances
- 4 Sums & Transformations of Random Variables
- 5 Statistical Independence



# Bayes' Theorem

## Bayes' Theorem

$$\underbrace{p(\mathbf{x} | \mathbf{y})}_{\text{posterior}} = \frac{\overbrace{p(\mathbf{y} | \mathbf{x})}^{\text{likelihood}} \overbrace{p(\mathbf{x})}^{\text{prior}}}{\underbrace{p(\mathbf{y})}_{\text{evidence}}}.$$

- Prior: subjective prior knowledge (before observing data).
- Likelihood  $p(\mathbf{y} | \mathbf{x})$ : the probability of  $\mathbf{y}$  if we were to know the latent variable  $\mathbf{x}$ .
  - We call it “the likelihood of  $\mathbf{x}$ ”.
- Posterior  $p(\mathbf{x} | \mathbf{y})$ : the quantity that we know about  $\mathbf{x}$  after having observed  $\mathbf{y}$ .

# Marginal Likelihood/Evidence

$$p(\mathbf{y}) := \sum_{\mathbf{x} \in \mathcal{X}} p(\mathbf{y} | \mathbf{x}) p(\mathbf{x}) = \mathbb{E}_{\mathcal{X}}[p(\mathbf{y} | \mathbf{x})]$$

$$p(\mathbf{y}) := \int_{\mathbf{x} \in \mathcal{X}} p(\mathbf{y} | \mathbf{x}) p(\mathbf{x}) d\mathbf{x} = \mathbb{E}_{\mathcal{X}}[p(\mathbf{y} | \mathbf{x})].$$

# Outline

- 1 Sum & Product Rule
- 2 Bayes' Theorem
- 3 Means & Covariances**
- 4 Sums & Transformations of Random Variables
- 5 Statistical Independence

# Expected Value

## Expected value

The expected value of a function  $g : \mathbb{R} \mapsto \mathbb{R}$  of a random variable  $X \sim p(x)$  is

$$\mathbb{E}_X[g(x)] = \int_{\mathcal{X}} g(x)p(x)dx,$$

or

$$\mathbb{E}_X[g(x)] = \sum_{x \in \mathcal{X}} g(x)p(x).$$

# Expected Value

## Expected value

The expected value of a function  $g : \mathbb{R} \mapsto \mathbb{R}$  of a random variable  $X \sim p(x)$  is

$$\mathbb{E}_X[g(x)] = \int_{\mathcal{X}} g(x)p(x)dx,$$

or

$$\mathbb{E}_X[g(x)] = \sum_{x \in \mathcal{X}} g(x)p(x).$$

## Multivariate $X$

$$\mathbb{E}_X[g(\mathbf{x})] = \begin{bmatrix} \mathbb{E}_{X_1}[g(x_1)] \\ \vdots \\ \mathbb{E}_{X_D}[g(x_D)] \end{bmatrix} \in \mathbb{R}^D,$$

where  $\mathbb{E}_{X_d}$ : taking the expectation w.r.t. the  $x_d$ .

# Linearity of Expectation

Let  $f(\mathbf{x}) = ag(\mathbf{x}) + bh(\mathbf{x})$  for  $a, b \in \mathbb{R}$  and  $\mathbf{x} \in \mathbb{R}^D$ .

$$\begin{aligned}\mathbb{E}_X[f(\mathbf{x})] &= \int f(\mathbf{x})p(\mathbf{x})d\mathbf{x} \\ &= \int [ag(\mathbf{x}) + bh(\mathbf{x})]d\mathbf{x} \\ &= a \int g(\mathbf{x})p(\mathbf{x})d\mathbf{x} + b \int h(\mathbf{x})p(\mathbf{x})d\mathbf{x} \\ &= a\mathbb{E}_X[g(\mathbf{x})] + b\mathbb{E}_X[h(\mathbf{x})]\end{aligned}$$

# Covariance

The (univariate) covariance between two univariate random variables  $X, Y \in \mathbb{R}$  is

$$\text{Cov}_{X,Y}[x, y] := \mathbb{E}_{X,Y}[(x - \mathbb{E}_X[x])(y - \mathbb{E}_Y[y])].$$

Omit the subscript.

$$\text{Cov}[x, y] := \mathbb{E}[xy] - \mathbb{E}[x]\mathbb{E}[y].$$

# Covariance

The (univariate) covariance between two univariate random variables  $X, Y \in \mathbb{R}$  is

$$\text{Cov}_{X,Y}[x, y] := \mathbb{E}_{X,Y}[(x - \mathbb{E}_X[x])(y - \mathbb{E}_Y[y])).$$

Omit the subscript.

$$\text{Cov}[x, y] := \mathbb{E}[xy] - \mathbb{E}[x]\mathbb{E}[y].$$

Note that

$$\text{Cov}[x, x] := \mathbb{E}[x^2] - (\mathbb{E}[x])^2$$

is the **variance** and denoted by  $\mathbb{V}_X[x]$  and  $\sqrt{\text{Cov}[x, x]}$  denoted by  $\sigma(x)$  is called the **standard deviation**.



# Covariance of Multivariate R.V.'s

## Covariance (Multivariate)

Consider random variables  $X$  and  $Y$  with states  $\mathbf{x} \in \mathbb{R}^D$  and  $\mathbf{y} \in \mathbb{R}^E$ . The covariance between  $X$  and  $Y$ :

$$\text{Cov}[\mathbf{x}, \mathbf{y}] =$$

# Covariance of Multivariate R.V.'s

## Covariance (Multivariate)

Consider random variables  $X$  and  $Y$  with states  $\mathbf{x} \in \mathbb{R}^D$  and  $\mathbf{y} \in \mathbb{R}^E$ . The covariance between  $X$  and  $Y$ :

$$\text{Cov}[\mathbf{x}, \mathbf{y}] = \mathbb{E}[\mathbf{x}\mathbf{y}^\top] - \mathbb{E}[\mathbf{x}]\mathbb{E}[\mathbf{y}]^\top$$

# Covariance of Multivariate R.V.'s

## Covariance (Multivariate)

Consider random variables  $X$  and  $Y$  with states  $\mathbf{x} \in \mathbb{R}^D$  and  $\mathbf{y} \in \mathbb{R}^E$ . The covariance between  $X$  and  $Y$ :

$$\text{Cov}[\mathbf{x}, \mathbf{y}] = \mathbb{E}[\mathbf{x}\mathbf{y}^\top] - \mathbb{E}[\mathbf{x}]\mathbb{E}[\mathbf{y}]^\top = \text{Cov}[\mathbf{y}, \mathbf{x}]^\top \in \mathbb{R}^{D \times E}.$$

# Variance (Multivariate)

## Variance (Multivariate)

The variance of a random variables  $X$  with states  $\mathbf{x} \in \mathbb{R}^D$  and mean  $\boldsymbol{\mu} \in \mathbb{R}^D$  is

$$\mathbb{V}_X[\mathbf{x}] = \text{Cov}_X[\mathbf{x}, \mathbf{x}] = \mathbb{E}_X[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^\top]$$

# Variance (Multivariate)

## Variance (Multivariate)

The variance of a random variables  $X$  with states  $\mathbf{x} \in \mathbb{R}^D$  and mean  $\boldsymbol{\mu} \in \mathbb{R}^D$  is

$$\mathbb{V}_X[\mathbf{x}] = \text{Cov}_X[\mathbf{x}, \mathbf{x}] = \mathbb{E}_X[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^\top] = \mathbb{E}_X[\mathbf{x}\mathbf{x}^\top] - \mathbb{E}_X[\mathbf{x}]\mathbb{E}_X[\mathbf{x}]^\top$$

# Variance (Multivariate)

## Variance (Multivariate)

The variance of a random variables  $X$  with states  $\mathbf{x} \in \mathbb{R}^D$  and mean  $\boldsymbol{\mu} \in \mathbb{R}^D$  is

$$\begin{aligned} \mathbb{V}_X[\mathbf{x}] &= \text{Cov}_X[\mathbf{x}, \mathbf{x}] = \mathbb{E}_X[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^\top] = \mathbb{E}_X[\mathbf{x}\mathbf{x}^\top] - \mathbb{E}_X[\mathbf{x}]\mathbb{E}_X[\mathbf{x}]^\top \\ &= \begin{bmatrix} \text{Cov}[x_1, x_1] & \text{Cov}[x_1, x_2] & \cdots & \text{Cov}[x_1, x_D] \\ \text{Cov}[x_2, x_1] & \text{Cov}[x_2, x_2] & \cdots & \text{Cov}[x_2, x_D] \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}[x_D, x_1] & \text{Cov}[x_D, x_2] & \cdots & \text{Cov}[x_D, x_D] \end{bmatrix}. \end{aligned}$$

# Variance (Multivariate)

## Variance (Multivariate)

The variance of a random variables  $X$  with states  $\mathbf{x} \in \mathbb{R}^D$  and mean  $\boldsymbol{\mu} \in \mathbb{R}^D$  is

$$\begin{aligned} \mathbb{V}_X[\mathbf{x}] &= \text{Cov}_X[\mathbf{x}, \mathbf{x}] = \mathbb{E}_X[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^\top] = \mathbb{E}_X[\mathbf{x}\mathbf{x}^\top] - \mathbb{E}_X[\mathbf{x}]\mathbb{E}_X[\mathbf{x}]^\top \\ &= \begin{bmatrix} \text{Cov}[x_1, x_1] & \text{Cov}[x_1, x_2] & \cdots & \text{Cov}[x_1, x_D] \\ \text{Cov}[x_2, x_1] & \text{Cov}[x_2, x_2] & \cdots & \text{Cov}[x_2, x_D] \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}[x_D, x_1] & \text{Cov}[x_D, x_2] & \cdots & \text{Cov}[x_D, x_D] \end{bmatrix}. \end{aligned}$$

- The **covariance matrix** of the multivariate  $X$ .

# Correlation

## Correlation

The correlation between two random variables  $X, Y$  is

$$\text{corr}[x, y] = \frac{\text{Cov}[x, y]}{\sqrt{\mathbb{V}[x]\mathbb{V}[y]}} \in [-1, 1].$$



# Empirical Means & Covariances

In machine learning, we need to learn from empirical observations of data.

## Empirical Mean & Covariance

The **empirical mean** vector: arithmetic average of the observations for each variable:

$$\bar{\mathbf{x}} := \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i,$$

for  $\mathbf{x}_i \in \mathbb{R}^D$ . The **empirical covariance** matrix is a  $D \times D$  matrix

$$\Sigma := \frac{1}{N} \sum_{i=1}^N (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^\top.$$

# Empirical Means & Covariances

In machine learning, we need to learn from empirical observations of data.

## Empirical Mean & Covariance

The **empirical mean** vector: arithmetic average of the observations for each variable:

$$\bar{\mathbf{x}} := \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i,$$

for  $\mathbf{x}_i \in \mathbb{R}^D$ . The **empirical covariance** matrix is a  $D \times D$  matrix

$$\Sigma := \frac{1}{N} \sum_{i=1}^N (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^\top.$$

- $\Sigma$  is symmetric, positive semidefinite.

# Computing the Empirical Variance

Approaches:

- ①  $\mathbb{V}_X[x] := \mathbb{E}_X[(x - \mu)^2]$ .
- ②  $\mathbb{V}_X[x] = \mathbb{E}_X[x^2] - (\mathbb{E}_X[x])^2$ .
  - One-pass; more efficient
- ③ Averaging pairwise differences between all pairs of observations.

$$\frac{1}{N^2} \sum_{i,j=1}^N (x_i - x_j)^2 = \textcolor{red}{2} \left[ \frac{1}{N} \sum_{i=1}^N x_i^2 - \left( \frac{1}{N} \sum_{i=1}^N x_i \right)^2 \right].$$

# Computing the Empirical Variance

Approaches:

- 1  $\mathbb{V}_X[x] := \mathbb{E}_X[(x - \mu)^2]$ .
- 2  $\mathbb{V}_X[x] = \mathbb{E}_X[x^2] - (\mathbb{E}_X[x])^2$ .
  - One-pass; more efficient
- 3 Averaging pairwise differences between all pairs of observations.

$$\frac{1}{N^2} \sum_{i,j=1}^N (x_i - x_j)^2 = 2 \left[ \frac{1}{N} \sum_{i=1}^N x_i^2 - \left( \frac{1}{N} \sum_{i=1}^N x_i \right)^2 \right].$$

- Twice of the 2nd approach.
- Interesting perspective to compute the left-hand side target.

# Outline

- 1 Sum & Product Rule
- 2 Bayes' Theorem
- 3 Means & Covariances
- 4 Sums & Transformations of Random Variables**
- 5 Statistical Independence

# Basic Rules

## Simple Rules & Exercise

Consider two random variables  $X, Y$  with states  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^D$ . Then,

$$\mathbb{E}[\mathbf{x} \pm \mathbf{y}] = \mathbb{E}[\mathbf{x}] \pm \mathbb{E}[\mathbf{y}]$$

$$\mathbb{V}[\mathbf{x} \pm \mathbf{y}] = \mathbb{V}[\mathbf{x}] \pm \mathbb{V}[\mathbf{y}] \pm \text{Cov}[\mathbf{x}, \mathbf{y}] \pm \text{Cov}[\mathbf{y}, \mathbf{x}] \quad (\text{Exercise}).$$

- **Note:** For a constant vector  $\mathbf{b} \in \mathbb{R}^D$ ,  $\mathbb{V}(\mathbf{x} \pm \mathbf{b}) = \mathbb{V}[\mathbf{x}]$  because  $\mathbb{V}[\mathbf{b}] = \mathbb{E}[\mathbf{b}\mathbf{b}^\top] - \mathbb{E}[\mathbf{b}]\mathbb{E}[\mathbf{b}]^\top = \mathbf{b}\mathbf{b}^\top - \mathbf{b}\mathbf{b}^\top = \mathbf{0}$  and

$$\text{Cov}(\mathbf{x}, \mathbf{b})$$

# Basic Rules

## Simple Rules & Exercise

Consider two random variables  $X, Y$  with states  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^D$ . Then,

$$\mathbb{E}[\mathbf{x} \pm \mathbf{y}] = \mathbb{E}[\mathbf{x}] \pm \mathbb{E}[\mathbf{y}]$$

$$\mathbb{V}[\mathbf{x} \pm \mathbf{y}] = \mathbb{V}[\mathbf{x}] \pm \mathbb{V}[\mathbf{y}] \pm \text{Cov}[\mathbf{x}, \mathbf{y}] \pm \text{Cov}[\mathbf{y}, \mathbf{x}] \quad (\text{Exercise}).$$

- **Note:** For a constant vector  $\mathbf{b} \in \mathbb{R}^D$ ,  $\mathbb{V}(\mathbf{x} \pm \mathbf{b}) = \mathbb{V}[\mathbf{x}]$  because  $\mathbb{V}[\mathbf{b}] = \mathbb{E}[\mathbf{b}\mathbf{b}^\top] - \mathbb{E}[\mathbf{b}]\mathbb{E}[\mathbf{b}]^\top = \mathbf{b}\mathbf{b}^\top - \mathbf{b}\mathbf{b}^\top = \mathbf{0}$  and

$$\text{Cov}(\mathbf{x}, \mathbf{b}) = \mathbb{E}[\mathbf{x}\mathbf{b}^\top] - \mathbb{E}[\mathbf{x}]\mathbb{E}[\mathbf{b}]^\top$$

# Basic Rules

## Simple Rules & Exercise

Consider two random variables  $X, Y$  with states  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^D$ . Then,

$$\mathbb{E}[\mathbf{x} \pm \mathbf{y}] = \mathbb{E}[\mathbf{x}] \pm \mathbb{E}[\mathbf{y}]$$

$$\mathbb{V}[\mathbf{x} \pm \mathbf{y}] = \mathbb{V}[\mathbf{x}] \pm \mathbb{V}[\mathbf{y}] \pm \text{Cov}[\mathbf{x}, \mathbf{y}] \pm \text{Cov}[\mathbf{y}, \mathbf{x}] \quad (\text{Exercise}).$$

- **Note:** For a constant vector  $\mathbf{b} \in \mathbb{R}^D$ ,  $\mathbb{V}(\mathbf{x} \pm \mathbf{b}) = \mathbb{V}[\mathbf{x}]$  because  $\mathbb{V}[\mathbf{b}] = \mathbb{E}[\mathbf{b}\mathbf{b}^\top] - \mathbb{E}[\mathbf{b}]\mathbb{E}[\mathbf{b}]^\top = \mathbf{b}\mathbf{b}^\top - \mathbf{b}\mathbf{b}^\top = \mathbf{0}$  and

$$\text{Cov}(\mathbf{x}, \mathbf{b}) = \mathbb{E}[\mathbf{x}\mathbf{b}^\top] - \mathbb{E}[\mathbf{x}]\mathbb{E}[\mathbf{b}]^\top = \mathbb{E}[\mathbf{x}]\mathbf{b}^\top - \mathbb{E}[\mathbf{x}]\mathbf{b}^\top = \mathbf{0}.$$

- **Question:** Why does the second equality hold?



Consider  $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}$  and let  $\Sigma := \mathbb{V}_X[\mathbf{x}]$ .

$$\mathbb{E}_Y[\mathbf{y}] = \mathbb{E}_X[\mathbf{Ax} + \mathbf{b}] = \mathbf{A}\mathbb{E}_X[\mathbf{x}] + \mathbf{b}$$

$$\mathbb{V}_Y[\mathbf{y}] = \mathbb{V}_X[\mathbf{A}\mathbf{x} + \mathbf{b}] = \mathbb{V}_X[\mathbf{A}\mathbf{x}] = \mathbf{A}\mathbb{V}_X[\mathbf{x}]\mathbf{A}^\top = \mathbf{A}\Sigma\mathbf{A}^\top.$$

◀ ◻ ▶ ◀ ◻ ▶ ◀ ≡ ▶ ◀ ≡ ▶ ≡

## Affine Transformation of r.v.'s (2/2)

Furthermore, let  $\boldsymbol{\mu} := \mathbb{E}_X[\mathbf{x}]$  and  $\boldsymbol{\Sigma} := \mathbb{V}_X[\mathbf{x}]$ .

$$\begin{aligned}\text{Cov}[\mathbf{x}, \mathbf{y}] &= \mathbb{E}[\mathbf{x}(\mathbf{Ax} + \mathbf{b})^\top] - \mathbb{E}[\mathbf{x}]\mathbb{E}[\mathbf{Ax} + \mathbf{b}]^\top \\ &= \boldsymbol{\mu}\mathbf{b}^\top + \mathbb{E}[\mathbf{xx}^\top]\mathbf{A}^\top - \boldsymbol{\mu}\mathbf{b}^\top - \boldsymbol{\mu}\boldsymbol{\mu}^\top\mathbf{A}^\top \\ &= (\mathbb{E}[\mathbf{xx}^\top] - \boldsymbol{\mu}\boldsymbol{\mu}^\top)\mathbf{A}^\top \\ &= \boldsymbol{\Sigma}\mathbf{A}^\top.\end{aligned}$$

# Outline

- 1 Sum & Product Rule
- 2 Bayes' Theorem
- 3 Means & Covariances
- 4 Sums & Transformations of Random Variables
- 5 Statistical Independence**

## (Statistically) Independent

Two random variables  $X, Y$  are statistically independent if and only if

$$p(\mathbf{x}, \mathbf{y}) = p(\mathbf{x})p(\mathbf{y}).$$

If  $X, Y$  are independent, then

- $p(\mathbf{y} \mid \mathbf{x}) = p(\mathbf{y})$ .
- $p(\mathbf{x} \mid \mathbf{y}) = p(\mathbf{x})$ .
- $\mathbb{V}_{X,Y}[\mathbf{x} + \mathbf{y}] = \mathbb{V}_X[\mathbf{x}] + \mathbb{V}_Y[\mathbf{y}]$ .
- $\text{Cov}_{X,Y}(\mathbf{x}, \mathbf{y}) = \mathbf{0}$ .

Note that  $\text{Cov}_{X,Y}(\mathbf{x}, \mathbf{y}) = \mathbf{0}$  does NOT necessarily imply that  $X$  and  $Y$  are independent.

See Example (6.5).

## Conditional Independence

Two random variables  $X, Y$  are conditionally independent given  $Z$  if and only if

$$p(\mathbf{x}, \mathbf{y} \mid \mathbf{z}) = p(\mathbf{x} \mid \mathbf{z})p(\mathbf{y} \mid \mathbf{z}).$$

for all  $\mathbf{z} \in \mathcal{Z}$ .

By the product rule, we can have

$$p(\mathbf{x}, \mathbf{y} \mid \mathbf{z}) = p(\mathbf{x} \mid \mathbf{y}, \mathbf{z})p(\mathbf{y} \mid \mathbf{z}).$$

Thus,

$$p(\mathbf{x} \mid \mathbf{y}, \mathbf{z}) = p(\mathbf{x} \mid \mathbf{z}).$$

## Heads Up

If  $X, Y$  are independent, then  $\mathbb{V}_{X,Y}[\mathbf{x} + \mathbf{y}] = \mathbb{V}_X[\mathbf{x}] + \mathbb{V}_Y[\mathbf{y}]$ .

$$\therefore \text{Cov}_{X,Y}(\mathbf{x}, \mathbf{y}) = \mathbf{0}$$

# Discussions