

# Mathematics for Machine Learning

— Probability & Distributions (Supplementary):

Gaussian Distribution & Change of Variables/Inverse Transform

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 Gaussian Distribution
  - Marginals and Conditionals of Gaussians
  - Sums and Linear Transformations
- 2 Change of Variables
  - Distribution Function Technique
  - Change of Variables

# Outline

## 1 Gaussian Distribution

- Marginals and Conditionals of Gaussians
- Sums and Linear Transformations

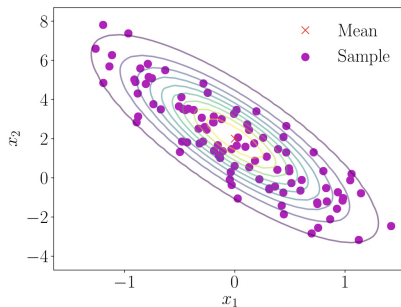
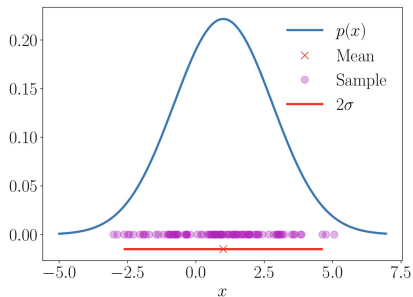
## 2 Change of Variables

- Distribution Function Technique
- Change of Variables

# Introduction

- The Gaussian distribution (a.k.s. normal distribution) is the most well-studied probability distribution for continuous-valued random variables.
- Widely used in statistics and machine learning.

# Gaussian Distributions Overlaid with Samples



# Univariate & Multivariate Gaussian

The probability density functions.

## Univariate

$$p(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right).$$

$$\Sigma = \mathbb{V}_X[\mathbf{x}] = \text{Cov}_X[\mathbf{x}, \mathbf{x}].$$

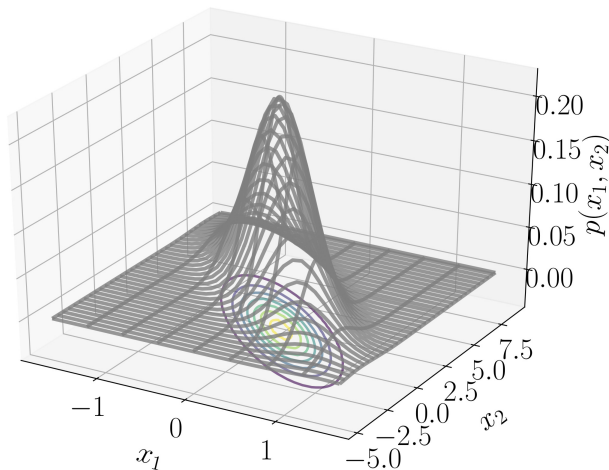
## Multivariate

$$p(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-\frac{D}{2}} \det(\boldsymbol{\Sigma})^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right),$$

for  $\mathbf{x} \in \mathbb{R}^D$ .

We write  $p(\mathbf{x}) = \mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$  or  $X \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ .

Gaussian distribution of two random variables  $x_1, x_2$ .





# Outline

- 1 Gaussian Distribution
  - Marginals and Conditionals of Gaussians
  - Sums and Linear Transformations
- 2 Change of Variables
  - Distribution Function Technique
  - Change of Variables

# Marginals and Conditionals of Gaussians

- Let  $X, Y$  be two multivariate random variables.
- Concatenate their states to be  $[\mathbf{x}^\top, \mathbf{y}^\top]$ .

$$p(\mathbf{x}, \mathbf{y}) = \mathcal{N} \left( \begin{bmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy} \\ \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yy} \end{bmatrix} \right).$$

where  $\boldsymbol{\Sigma}_{xx} = \text{Cov}[\mathbf{x}, \mathbf{x}]$ ,  $\boldsymbol{\Sigma}_{yy} = \text{Cov}[\mathbf{y}, \mathbf{y}]$ ,  $\boldsymbol{\Sigma}_{xy} = \text{Cov}[\mathbf{x}, \mathbf{y}]$ .

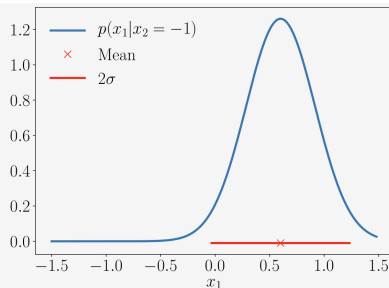
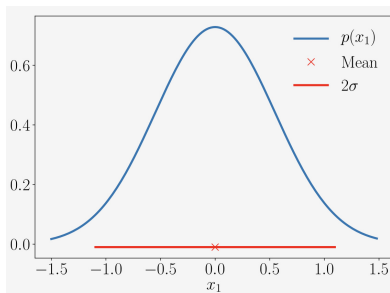
- By [Bishop 2006], the conditional distribution  $p(\mathbf{x} | \mathbf{y})$  is also Gaussian.

$$\begin{aligned} p(\mathbf{x} | \mathbf{y}) &= \mathcal{N}(\boldsymbol{\mu}_{x|y}, \boldsymbol{\Sigma}_{x|y}) \\ \boldsymbol{\mu}_{x|y} &= \boldsymbol{\mu}_x + \boldsymbol{\Sigma}_{xy} \boldsymbol{\Sigma}_{yy}^{-1} (\mathbf{y} - \boldsymbol{\mu}_y) \\ \boldsymbol{\Sigma}_{x|y} &= \boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{xy} \boldsymbol{\Sigma}_{yy}^{-1} \boldsymbol{\Sigma}_{yx}. \end{aligned}$$

$$p(\mathbf{x}) = \int p(\mathbf{x}, \mathbf{y}) d\mathbf{y} = \mathcal{N}(\mathbf{x} | \boldsymbol{\mu}_x, \boldsymbol{\Sigma}_{xx}).$$

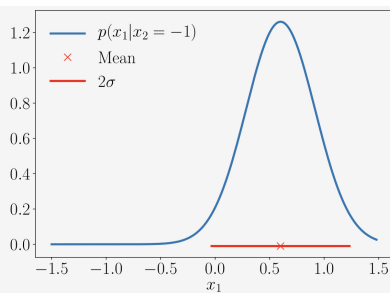
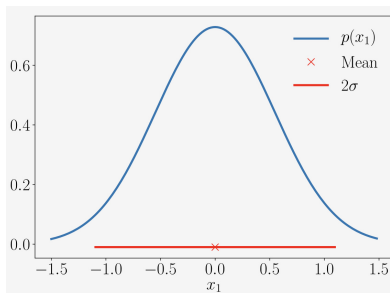
# Example

Consider  $p(x_1, x_2) = \mathcal{N}\left(\begin{bmatrix} 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 0.3 & -1 \\ -1 & 5 \end{bmatrix}\right)$ .



# Example

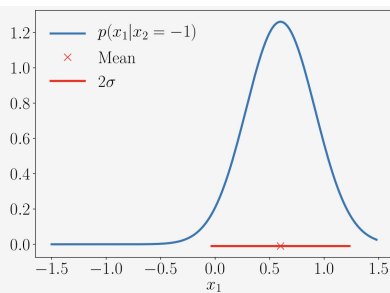
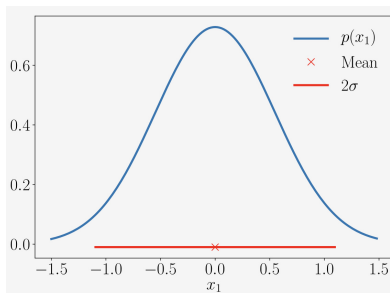
Consider  $p(x_1, x_2) = \mathcal{N}\left(\begin{bmatrix} 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 0.3 & -1 \\ -1 & 5 \end{bmatrix}\right)$ .



Conditioned on  $x_2 = -1$ ,  $\mu_{x_1|x_2=-1} = 0 + (-1) \cdot 0.2 \cdot (-1 - 2) = 0.6$

# Example

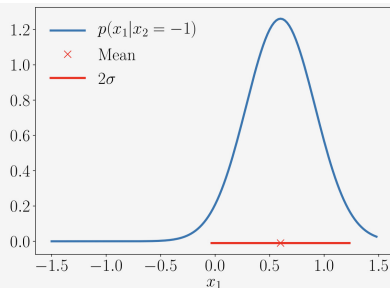
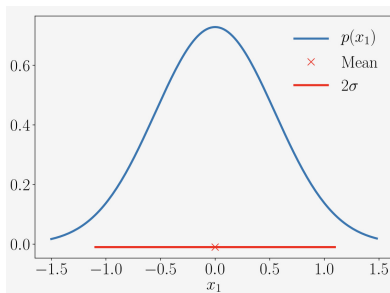
Consider  $p(x_1, x_2) = \mathcal{N}\left(\begin{bmatrix} 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 0.3 & -1 \\ -1 & 5 \end{bmatrix}\right)$ .



Conditioned on  $x_2 = -1$ ,  $\mu_{x_1|x_2=-1} = 0 + (-1) \cdot 0.2 \cdot (-1 - 2) = 0.6$   
 and  $\sigma_{x_1|x_2=-1}^2 =$

# Example

Consider  $p(x_1, x_2) = \mathcal{N}\left(\begin{bmatrix} 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 0.3 & -1 \\ -1 & 5 \end{bmatrix}\right)$ .

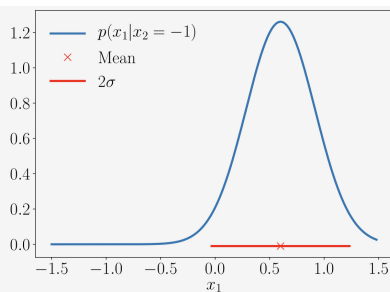
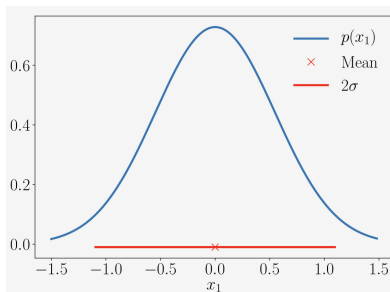


Conditioned on  $x_2 = -1$ ,  $\mu_{x_1|x_2=-1} = 0 + (-1) \cdot 0.2 \cdot (-1 - 2) = 0.6$   
 and  $\sigma^2_{x_1|x_2=-1} = 0.3 - (-1) \cdot 0.2 \cdot (-1) = 0.1$ .

Thus,  $p(x_1 | x_2 = -1) =$

# Example

Consider  $p(x_1, x_2) = \mathcal{N}\left(\begin{bmatrix} 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 0.3 & -1 \\ -1 & 5 \end{bmatrix}\right)$ .

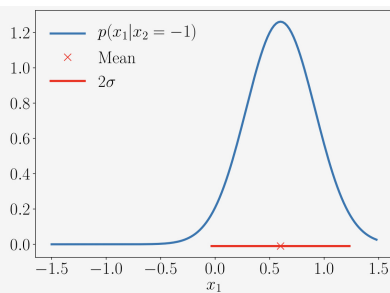
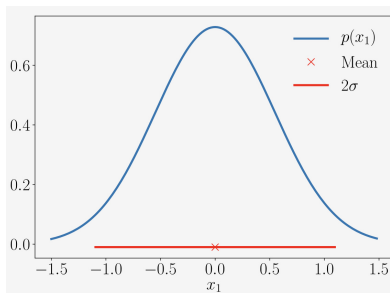


Conditioned on  $x_2 = -1$ ,  $\mu_{x_1|x_2=-1} = 0 + (-1) \cdot 0.2 \cdot (-1 - 2) = 0.6$   
 and  $\sigma^2_{x_1|x_2=-1} = 0.3 - (-1) \cdot 0.2 \cdot (-1) = 0.1$ .

Thus,  $p(x_1 | x_2 = -1) = \mathcal{N}(0.6, 0.1)$ ,

# Example

Consider  $p(x_1, x_2) = \mathcal{N}\left(\begin{bmatrix} 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 0.3 & -1 \\ -1 & 5 \end{bmatrix}\right)$ .



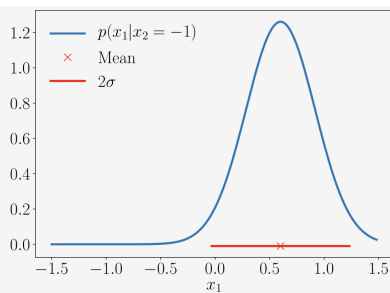
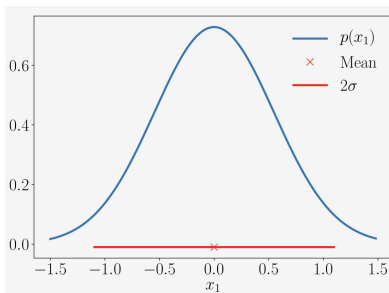
Conditioned on  $x_2 = -1$ ,  $\mu_{x_1|x_2=-1} = 0 + (-1) \cdot 0.2 \cdot (-1 - 2) = 0.6$   
 and  $\sigma^2_{x_1|x_2=-1} = 0.3 - (-1) \cdot 0.2 \cdot (-1) = 0.1$ .

Thus,  $p(x_1 | x_2 = -1) = \mathcal{N}(0.6, 0.1)$ ,  $p(x_1) =$



# Example

Consider  $p(x_1, x_2) = \mathcal{N}\left(\begin{bmatrix} 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 0.3 & -1 \\ -1 & 5 \end{bmatrix}\right)$ .



Conditioned on  $x_2 = -1$ ,  $\mu_{x_1|x_2=-1} = 0 + (-1) \cdot 0.2 \cdot (-1 - 2) = 0.6$   
 and  $\sigma^2_{x_1|x_2=-1} = 0.3 - (-1) \cdot 0.2 \cdot (-1) = 0.1$ .

Thus,  $p(x_1 | x_2 = -1) = \mathcal{N}(0.6, 0.1)$ ,  $p(x_1) = \mathcal{N}(0, 0.3)$ .

# Outline

- 1 Gaussian Distribution
  - Marginals and Conditionals of Gaussians
  - Sums and Linear Transformations
- 2 Change of Variables
  - Distribution Function Technique
  - Change of Variables

# Sum of Gaussians

Say  $X, Y$  are two independent Gaussian random variables with

$$X \sim \mathcal{N}(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x) \quad \text{and} \quad Y \sim \mathcal{N}(\boldsymbol{\mu}_y, \boldsymbol{\Sigma}_y).$$

# Sum of Gaussians

Say  $X, Y$  are two independent Gaussian random variables with

$$X \sim \mathcal{N}(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x) \text{ and } Y \sim \mathcal{N}(\boldsymbol{\mu}_y, \boldsymbol{\Sigma}_y).$$

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

# Sum of Gaussians

Say  $X, Y$  are two independent Gaussian random variables with

$$X \sim \mathcal{N}(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x) \text{ and } Y \sim \mathcal{N}(\boldsymbol{\mu}_y, \boldsymbol{\Sigma}_y).$$

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

Then  $X + Y$  is also a Gaussian distribution with

$$X + Y \sim \mathcal{N}(\boldsymbol{\mu}_x + \boldsymbol{\mu}_y, \boldsymbol{\Sigma}_x + \boldsymbol{\Sigma}_y)$$

# Sum of Gaussians

Say  $X, Y$  are two independent Gaussian random variables with

$$X \sim \mathcal{N}(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x) \text{ and } Y \sim \mathcal{N}(\boldsymbol{\mu}_y, \boldsymbol{\Sigma}_y).$$

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

Then  $X + Y$  is also a Gaussian distribution with

$$X + Y \sim \mathcal{N}(\boldsymbol{\mu}_x + \boldsymbol{\mu}_y, \boldsymbol{\Sigma}_x + \boldsymbol{\Sigma}_y)$$

Please recall  $\mathbb{E}[\mathbf{x} + \mathbf{y}]$  and  $\mathbb{V}[\mathbf{x} + \mathbf{y}]$ .

# Example

## Linear Combination of Gaussians

$$p(ax + by) =$$

# Example

## Linear Combination of Gaussians

$$p(a\mathbf{x} + b\mathbf{y}) = \mathcal{N}(a\mu_x + b\mu_y,$$



# Example

## Linear Combination of Gaussians

$$p(a\mathbf{x} + b\mathbf{y}) = \mathcal{N}(a\boldsymbol{\mu}_x + b\boldsymbol{\mu}_y, a^2\boldsymbol{\Sigma}_x + b^2\boldsymbol{\Sigma}_y).$$

# Example

## Linear Combination of Gaussians

$$p(a\mathbf{x} + b\mathbf{y}) = \mathcal{N}(a\boldsymbol{\mu}_x + b\boldsymbol{\mu}_y, a^2\boldsymbol{\Sigma}_x + b^2\boldsymbol{\Sigma}_y).$$

## Theorem [Mixture of Two Univariate Gaussian Densities]

Consider a mixture of two univariate Gaussian densities

$$p(x) = \alpha p_1(x) + (1 - \alpha)p_2(x)$$

for the **mixture weight**  $0 < \alpha < 1$  and  $(\mu_1, \sigma_1^2) \neq (\mu_2, \sigma_2^2)$ . Then,

$$\mathbb{E}[x] = \alpha\mu_1 + (1 - \alpha)\mu_2$$

$$\begin{aligned}\mathbb{V}[x] &= [\alpha\sigma_1^2 + (1 - \alpha)\sigma_2^2] \\ &\quad + ([\alpha\mu_1^2 + (1 - \alpha)\mu_2^2] - [\alpha\mu_1 + (1 - \alpha)\mu_2]^2).\end{aligned}$$

# Proof of the Theorem

Sketch:

$$\begin{aligned} \textcircled{1} \quad \mathbb{E}[x] &= \int_{-\infty}^{\infty} xp(x)dx = \int_{-\infty}^{\infty} (\alpha xp_1(x) + (1 - \alpha)xp_2(x))dx \\ &= \alpha\mu_1 + (1 - \alpha)\mu_2. \end{aligned}$$

$$\textcircled{2} \quad \mathbb{E}[x^2] =$$

# Proof of the Theorem

Sketch:

$$\begin{aligned} \textcircled{1} \quad \mathbb{E}[x] &= \int_{-\infty}^{\infty} xp(x)dx = \int_{-\infty}^{\infty} (\alpha xp_1(x) + (1 - \alpha)xp_2(x))dx \\ &= \alpha\mu_1 + (1 - \alpha)\mu_2. \end{aligned}$$

$$\begin{aligned} \textcircled{2} \quad \mathbb{E}[x^2] &= \int_{-\infty}^{\infty} x^2p(x)dx = \int_{-\infty}^{\infty} (\alpha x^2p_1(x) + (1 - \alpha)x^2p_2(x))dx \\ &= \alpha(\mu_1^2 + \sigma_1^2) + (1 - \alpha)(\mu_2^2 + \sigma_2^2). \end{aligned}$$

# Proof of the Theorem

Sketch:

$$\begin{aligned} \textcircled{1} \quad \mathbb{E}[x] &= \int_{-\infty}^{\infty} xp(x)dx = \int_{-\infty}^{\infty} (\alpha xp_1(x) + (1 - \alpha)xp_2(x))dx \\ &= \alpha\mu_1 + (1 - \alpha)\mu_2. \end{aligned}$$

$$\begin{aligned} \textcircled{2} \quad \mathbb{E}[x^2] &= \int_{-\infty}^{\infty} x^2p(x)dx = \int_{-\infty}^{\infty} (\alpha x^2p_1(x) + (1 - \alpha)x^2p_2(x))dx \\ &= \alpha(\mu_1^2 + \sigma_1^2) + (1 - \alpha)(\mu_2^2 + \sigma_2^2). \end{aligned}$$

- **Recall:**  $\mathbb{V}_X[x] = \mathbb{E}_X[x^2] - (\mathbb{E}_X[x])^2.$

# Proof of the Theorem

Sketch:

$$\begin{aligned} \textcircled{1} \quad \mathbb{E}[x] &= \int_{-\infty}^{\infty} xp(x)dx = \int_{-\infty}^{\infty} (\alpha xp_1(x) + (1 - \alpha)xp_2(x))dx \\ &= \alpha\mu_1 + (1 - \alpha)\mu_2. \end{aligned}$$

$$\begin{aligned} \textcircled{2} \quad \mathbb{E}[x^2] &= \int_{-\infty}^{\infty} x^2p(x)dx = \int_{-\infty}^{\infty} (\alpha x^2p_1(x) + (1 - \alpha)x^2p_2(x))dx \\ &= \alpha(\mu_1^2 + \sigma_1^2) + (1 - \alpha)(\mu_2^2 + \sigma_2^2). \end{aligned}$$

- **Recall:**  $\mathbb{V}_X[x] = \mathbb{E}_X[x^2] - (\mathbb{E}_X[x])^2$ .

Using  $\textcircled{1}$  &  $\textcircled{2}$  we can prove the theorem.

# Linear Transformation by a Matrix (1/2)

$X \sim \mathcal{N}(\mu, \Sigma)$  and  $\mathbf{y} = \mathbf{A}\mathbf{x}$

- The expectation:  $\mathbb{E}[\mathbf{y}] = \mathbb{E}[\mathbf{A}\mathbf{x}] =$

# Linear Transformation by a Matrix (1/2)

$X \sim \mathcal{N}(\mu, \Sigma)$  and  $\mathbf{y} = \mathbf{A}\mathbf{x}$

- The expectation:  $\mathbb{E}[\mathbf{y}] = \mathbb{E}[\mathbf{A}\mathbf{x}] = \mathbf{A}\mathbb{E}[\mathbf{x}] = \mathbf{A}\mu$ .



# Linear Transformation by a Matrix (1/2)

$X \sim \mathcal{N}(\mu, \Sigma)$  and  $\mathbf{y} = \mathbf{A}\mathbf{x}$

- The expectation:  $\mathbb{E}[\mathbf{y}] = \mathbb{E}[\mathbf{A}\mathbf{x}] = \mathbf{A}\mathbb{E}[\mathbf{x}] = \mathbf{A}\mu$ .
- The variance:  $\mathbb{V}[\mathbf{y}] = \mathbb{V}[\mathbf{A}\mathbf{x}] =$

# Linear Transformation by a Matrix (1/2)

$X \sim \mathcal{N}(\mu, \Sigma)$  and  $\mathbf{y} = \mathbf{A}\mathbf{x}$

- The expectation:  $\mathbb{E}[\mathbf{y}] = \mathbb{E}[\mathbf{A}\mathbf{x}] = \mathbf{A}\mathbb{E}[\mathbf{x}] = \mathbf{A}\mu$ .
- The variance:  $\mathbb{V}[\mathbf{y}] = \mathbb{V}[\mathbf{A}\mathbf{x}] = \mathbf{A}\mathbb{V}[\mathbf{x}]\mathbf{A}^\top = \mathbf{A}\Sigma\mathbf{A}^\top$ .

# Linear Transformation by a Matrix (1/2)

$\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  and  $\mathbf{y} = \mathbf{A}\mathbf{x}$

- The expectation:  $\mathbb{E}[\mathbf{y}] = \mathbb{E}[\mathbf{A}\mathbf{x}] = \mathbf{A}\mathbb{E}[\mathbf{x}] = \mathbf{A}\boldsymbol{\mu}$ .
- The variance:  $\mathbb{V}[\mathbf{y}] = \mathbb{V}[\mathbf{A}\mathbf{x}] = \mathbf{A}\mathbb{V}[\mathbf{x}]\mathbf{A}^\top = \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top$ .
- Thus, we have

# Linear Transformation by a Matrix (1/2)

$X \sim \mathcal{N}(\mu, \Sigma)$  and  $y = Ax$

- The expectation:  $\mathbb{E}[y] = \mathbb{E}[Ax] = A\mathbb{E}[x] = A\mu$ .
- The variance:  $\mathbb{V}[y] = \mathbb{V}[Ax] = A\mathbb{V}[x]A^\top = A\Sigma A^\top$ .
- Thus, we have

$$Y \sim \mathcal{N}(A\mu, A\Sigma A^\top).$$

## Linear Transformation by a Matrix (2/2)

Let's consider the *reverse transformation*.

$Y \sim \mathcal{N}(\mu_y, \Sigma)$ ,  $\mathbf{y} = \mathbf{A}\mathbf{x}$  for  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^M$ , a full rank  $\mathbf{A} \in \mathbb{R}^{M \times N}$ ,  $M \geq N$

- $p(\mathbf{y}) = \mathcal{N}(\mathbf{y} \mid \mathbf{A}\mathbf{x}, \Sigma)$ .
- **Note:**  $\mathbf{A}$  might not be invertible.

## Linear Transformation by a Matrix (2/2)

Let's consider the *reverse transformation*.

$Y \sim \mathcal{N}(\mu_y, \Sigma)$ ,  $\mathbf{y} = \mathbf{A}\mathbf{x}$  for  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^M$ , a full rank  $\mathbf{A} \in \mathbb{R}^{M \times N}$ ,  $M \geq N$

- $p(\mathbf{y}) = \mathcal{N}(\mathbf{y} \mid \mathbf{A}\mathbf{x}, \Sigma)$ .
  - **Note:**  $\mathbf{A}$  might not be invertible.
- $\mathbf{y} = \mathbf{A}\mathbf{x}$

## Linear Transformation by a Matrix (2/2)

Let's consider the *reverse transformation*.

$Y \sim \mathcal{N}(\mu_y, \Sigma)$ ,  $\mathbf{y} = \mathbf{A}\mathbf{x}$  for  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^M$ , a full rank  $\mathbf{A} \in \mathbb{R}^{M \times N}$ ,  $M \geq N$

- $p(\mathbf{y}) = \mathcal{N}(\mathbf{y} \mid \mathbf{A}\mathbf{x}, \Sigma)$ .
  - **Note:**  $\mathbf{A}$  might not be invertible.
- $\mathbf{y} = \mathbf{A}\mathbf{x} \iff \mathbf{A}^\top \mathbf{y} = \mathbf{A}^\top \mathbf{A}\mathbf{x}$

## Linear Transformation by a Matrix (2/2)

Let's consider the *reverse transformation*.

$Y \sim \mathcal{N}(\mu_y, \Sigma)$ ,  $\mathbf{y} = \mathbf{A}\mathbf{x}$  for  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^M$ , a full rank  $\mathbf{A} \in \mathbb{R}^{M \times N}$ ,  $M \geq N$

- $p(\mathbf{y}) = \mathcal{N}(\mathbf{y} \mid \mathbf{A}\mathbf{x}, \Sigma)$ .
  - **Note:**  $\mathbf{A}$  might not be invertible.
- $\mathbf{y} = \mathbf{A}\mathbf{x} \iff \mathbf{A}^\top \mathbf{y} = \mathbf{A}^\top \mathbf{A}\mathbf{x} \iff (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{y} = \mathbf{x}.$



## Linear Transformation by a Matrix (2/2)

Let's consider the *reverse transformation*.

$Y \sim \mathcal{N}(\mu_y, \Sigma)$ ,  $\mathbf{y} = \mathbf{A}\mathbf{x}$  for  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^M$ , a full rank  $\mathbf{A} \in \mathbb{R}^{M \times N}$ ,  $M \geq N$

- $p(\mathbf{y}) = \mathcal{N}(\mathbf{y} \mid \mathbf{A}\mathbf{x}, \Sigma)$ .
  - **Note:**  $\mathbf{A}$  might not be invertible.
- $\mathbf{y} = \mathbf{A}\mathbf{x} \iff \mathbf{A}^\top \mathbf{y} = \mathbf{A}^\top \mathbf{A}\mathbf{x} \iff (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{y} = \mathbf{x}$ .
  - This works even for non-invertible  $\mathbf{A}$ !

## Linear Transformation by a Matrix (2/2)

Let's consider the *reverse transformation*.

$Y \sim \mathcal{N}(\mu_y, \Sigma)$ ,  $\mathbf{y} = \mathbf{A}\mathbf{x}$  for  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^M$ , a full rank  $\mathbf{A} \in \mathbb{R}^{M \times N}$ ,  $M \geq N$

- $p(\mathbf{y}) = \mathcal{N}(\mathbf{y} \mid \mathbf{A}\mathbf{x}, \Sigma)$ .
  - **Note:**  $\mathbf{A}$  might not be invertible.
- $\mathbf{y} = \mathbf{A}\mathbf{x} \iff \mathbf{A}^\top \mathbf{y} = \mathbf{A}^\top \mathbf{A}\mathbf{x} \iff (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{y} = \mathbf{x}$ .
  - This works even for non-invertible  $\mathbf{A}$ !
- The variance:  $\mathbb{V}[\mathbf{x}] = \mathbb{V}[(\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{y}] =$

## Linear Transformation by a Matrix (2/2)

Let's consider the *reverse transformation*.

$Y \sim \mathcal{N}(\mu_y, \Sigma)$ ,  $\mathbf{y} = \mathbf{A}\mathbf{x}$  for  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^M$ , a full rank  $\mathbf{A} \in \mathbb{R}^{M \times N}$ ,  $M \geq N$

- $p(\mathbf{y}) = \mathcal{N}(\mathbf{y} \mid \mathbf{A}\mathbf{x}, \Sigma)$ .
  - **Note:**  $\mathbf{A}$  might not be invertible.
- $\mathbf{y} = \mathbf{A}\mathbf{x} \iff \mathbf{A}^\top \mathbf{y} = \mathbf{A}^\top \mathbf{A}\mathbf{x} \iff (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{y} = \mathbf{x}$ .
  - This works even for non-invertible  $\mathbf{A}$ !
- The variance:  $\mathbb{V}[\mathbf{x}] = \mathbb{V}[(\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{y}] = (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \Sigma \mathbf{A} (\mathbf{A}^\top \mathbf{A})^{-1}$ .

## Linear Transformation by a Matrix (2/2)

Let's consider the *reverse transformation*.

$Y \sim \mathcal{N}(\mu_y, \Sigma)$ ,  $\mathbf{y} = \mathbf{A}\mathbf{x}$  for  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^M$ , a full rank  $\mathbf{A} \in \mathbb{R}^{M \times N}$ ,  $M \geq N$

- $p(\mathbf{y}) = \mathcal{N}(\mathbf{y} \mid \mathbf{A}\mathbf{x}, \Sigma)$ .
  - **Note:**  $\mathbf{A}$  might not be invertible.
- $\mathbf{y} = \mathbf{A}\mathbf{x} \iff \mathbf{A}^\top \mathbf{y} = \mathbf{A}^\top \mathbf{A}\mathbf{x} \iff (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{y} = \mathbf{x}$ .
  - This works even for non-invertible  $\mathbf{A}$ !
- The variance:  $\mathbb{V}[\mathbf{x}] = \mathbb{V}[(\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{y}] = (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \Sigma \mathbf{A} (\mathbf{A}^\top \mathbf{A})^{-1}$ .
- Thus, we have

## Linear Transformation by a Matrix (2/2)

Let's consider the *reverse transformation*.

$Y \sim \mathcal{N}(\mu_y, \Sigma)$ ,  $y = \mathbf{A}x$  for  $x, y \in \mathbb{R}^M$ , a full rank  $\mathbf{A} \in \mathbb{R}^{M \times N}$ ,  $M \geq N$

- $p(y) = \mathcal{N}(y \mid \mathbf{A}x, \Sigma)$ .
  - **Note:**  $\mathbf{A}$  might not be invertible.
- $y = \mathbf{A}x \iff \mathbf{A}^\top y = \mathbf{A}^\top \mathbf{A}x \iff (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top y = x$ .
  - This works even for non-invertible  $\mathbf{A}$ !
- The variance:  $\mathbb{V}[x] = \mathbb{V}[(\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top y] = (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \Sigma \mathbf{A} (\mathbf{A}^\top \mathbf{A})^{-1}$ .
- Thus, we have

$$X \sim \mathcal{N}((\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mu_y, (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \Sigma \mathbf{A} (\mathbf{A}^\top \mathbf{A})^{-1}).$$

# Exercise

Another example of *reverse transformation*.

$Y \sim \mathcal{N}(\mu_y, \Sigma)$  and  $\mathbf{y} = \mathbf{A}\mathbf{x}$  for  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^M$ , and  $\mathbf{A}$  is invertible

- $p(\mathbf{y}) = \mathcal{N}(\mathbf{y} \mid \mathbf{A}\mathbf{x}, \Sigma)$ .
- Compute  $\mathbb{E}[\mathbf{x}]$ .
- Compute  $\mathbb{V}[\mathbf{x}]$ .
- Derive  $X \sim \mathcal{N}(?, ?)$ .

# A Sampling Approach

We want to obtain samples from a multivariate  $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ .

- However, we only have a sampler of  $\mathcal{N}(\mathbf{0}, \mathbf{I})$  at hand.

# A Sampling Approach

We want to obtain samples from a multivariate  $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ .

- However, we only have a sampler of  $\mathcal{N}(\mathbf{0}, \mathbf{I})$  at hand.

- Assume that we have  $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ .
- Then, define  $\mathbf{y} = \mathbf{A}\mathbf{x} + \boldsymbol{\mu}$ , where  $\mathbf{A}\mathbf{A}^\top = \mathbf{A}\mathbf{A}^\top = \boldsymbol{\Sigma}$ .



# A Sampling Approach

We want to obtain samples from a multivariate  $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ .

- However, we only have a sampler of  $\mathcal{N}(\mathbf{0}, \mathbf{I})$  at hand.

- Assume that we have  $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ .
- Then, define  $\mathbf{y} = \mathbf{A}\mathbf{x} + \boldsymbol{\mu}$ , where  $\mathbf{A}\mathbf{A}^\top = \mathbf{A}\mathbf{A}^\top = \boldsymbol{\Sigma}$ .
- To derive  $\mathbf{A}$ :

# A Sampling Approach

We want to obtain samples from a multivariate  $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ .

- However, we only have a sampler of  $\mathcal{N}(\mathbf{0}, \mathbf{I})$  at hand.

- Assume that we have  $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ .
- Then, define  $\mathbf{y} = \mathbf{A}\mathbf{x} + \boldsymbol{\mu}$ , where  $\mathbf{A}\mathbf{A}^\top = \mathbf{A}\mathbf{A}^\top = \boldsymbol{\Sigma}$ .
- To derive  $\mathbf{A}$ : Use **Cholesky decomposition** of the covariance matrix  $\boldsymbol{\Sigma}$ .
  - $\mathbf{A}$  will be triangular and efficient for computation.

# Outline

- 1 Gaussian Distribution
  - Marginals and Conditionals of Gaussians
  - Sums and Linear Transformations
- 2 Change of Variables
  - Distribution Function Technique
  - Change of Variables

# Motivation

Consider the following examples.

- Assuming that  $X$  is a random variable distributed according to some well-known distribution, then **what is the distribution of  $X^2$** ?
- Assuming that  $X_1, X_2$  are two univariate standard normal distributions, then **what is the distribution of  $\frac{1}{2}(X_1 + X_2)$** ?

# Motivation

Consider the following examples.

- Assuming that  $X$  is a random variable distributed according to some well-known distribution, then **what is the distribution of  $X^2$** ?
- Assuming that  $X_1, X_2$  are two univariate standard normal distributions, then **what is the distribution of  $\frac{1}{2}(X_1 + X_2)$** ?
- What if the transformation is **nonlinear**?

# Motivation

Consider the following examples.

- Assuming that  $X$  is a random variable distributed according to some well-known distribution, then **what is the distribution of  $X^2$** ?
- Assuming that  $X_1, X_2$  are two univariate standard normal distributions, then **what is the distribution of  $\frac{1}{2}(X_1 + X_2)$** ?
- What if the transformation is **nonlinear**?
  - Closed-form expressions are not readily available.

# Straightforward for Discrete Random Variables

## Example: Univariate Random Variables

Given

- A discrete random variable  $X$  with pmf  $\Pr[X = x]$ .
- An invertible function  $U(x)$ .

Consider the transformed random variable  $Y := U(X)$  with pmf  $\Pr[Y = y]$ . Then

$$\begin{aligned}\Pr[Y = y] &= \Pr[U(X) = y] && \text{(transformation of interest)} \\ &= \Pr[X = U^{-1}(y)] && \text{(inverse)}\end{aligned}$$

where we can observe  $x = U^{-1}(y)$ .

# Two Approaches

- We consider the discrete case (e.g.,  $\Pr[X = x]$ ).
- Two approaches:
  - ① Cumulative distribution (Distribution Function Technique).
  - ② Change-of-variable.



# Outline

- 1 Gaussian Distribution
  - Marginals and Conditionals of Gaussians
  - Sums and Linear Transformations
- 2 Change of Variables
  - Distribution Function Technique
  - Change of Variables

# Distribution Function Technique

**Note:** a cdf of  $X$ :  $F_X(x) = \Pr[X \leq x]$ .

Goal: Find the cdf of the random variable  $Y := U(X)$

- 1 Find the cdf

$$F_Y(y) = \Pr[Y \leq y].$$

- 2 Differentiating  $F_Y(y)$  to get the pdf  $f_Y(y)$ :

$$f_Y(y) = \frac{d}{dy} F_Y(y).$$

# Distribution Function Technique

**Note:** a cdf of  $X$ :  $F_X(x) = \Pr[X \leq x]$ .

Goal: Find the cdf of the random variable  $Y := U(X)$

- 1 Find the cdf

$$F_Y(y) = \Pr[Y \leq y].$$

- 2 Differentiating  $F_Y(y)$  to get the pdf  $f_Y(y)$ :

$$f_Y(y) = \frac{d}{dy} F_Y(y).$$

**Note:** The domain of the random variable may have changed!

# Example

## Example

Let  $X$  be a continuous random variable with pdf  $f_X : [0, 1] \mapsto [0, 1]$ :

$$f_X(x) = 3x^2.$$

**Goal:** Find the pdf of  $Y = X^2$ .

$$F_Y(y) = \Pr[Y \leq y]$$

# Example

## Example

Let  $X$  be a continuous random variable with pdf  $f_X : [0, 1] \mapsto [0, 1]$ :

$$f_X(x) = 3x^2.$$

**Goal:** Find the pdf of  $Y = X^2$ .

$$\begin{aligned} F_Y(y) &= \Pr[Y \leq y] = \Pr[X^2 \leq y] \\ &= \Pr[X \leq y^{\frac{1}{2}}] \\ &= F_X(y^{\frac{1}{2}}) \end{aligned}$$

# Example

## Example

Let  $X$  be a continuous random variable with pdf  $f_X : [0, 1] \mapsto [0, 1]$ :

$$f_X(x) = 3x^2.$$

**Goal:** Find the pdf of  $Y = X^2$ .

$$\begin{aligned}F_Y(y) &= \Pr[Y \leq y] = \Pr[X^2 \leq y] \\&= \Pr[X \leq y^{\frac{1}{2}}] \\&= F_X(y^{\frac{1}{2}}) = \int_0^{y^{\frac{1}{2}}} 3t^2 dt \\&= [t^3]_0^{y^{\frac{1}{2}}} = y^{\frac{3}{2}}, \quad 0 \leq y \leq 1.\end{aligned}$$

# Example

## Example

Let  $X$  be a continuous random variable with pdf  $f_X : [0, 1] \mapsto [0, 1]$ :

$$f_X(x) = 3x^2.$$

**Goal:** Find the pdf of  $Y = X^2$ .

$$\begin{aligned} F_Y(y) &= \Pr[Y \leq y] = \Pr[X^2 \leq y] && \text{Thus,} \\ &= \Pr[X \leq y^{\frac{1}{2}}] && f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{3}{2} y^{\frac{1}{2}} \\ &= F_X(y^{\frac{1}{2}}) = \int_0^{y^{\frac{1}{2}}} 3t^2 dt && \text{for } 0 \leq y \leq 1. \\ &= [t^3]_0^{y^{\frac{1}{2}}} = y^{\frac{3}{2}}, \quad 0 \leq y \leq 1. \end{aligned}$$

# Exercise

## Theorem [Casella & Berger (2002)]

Let  $X$  be a continuous random variable with a *strictly monotone* cumulative distribution function  $F_X(x)$ . Then, the random variable  $Y$  defined as

$$Y := F_X(X)$$

has a **uniform distribution**.

## Exercise

Consider  $f_X(x) = 3x^2$  in the previous example. Show that  $Y := F_X(X)$  attains a uniform distribution.



# Remark

The first approach relies on the following facts:

- We can transform the cdf of  $Y$  into an expression that is a cdf of  $X$ .
- We can differentiate the cdf to obtain the pdf.

# Outline

- 1 Gaussian Distribution
  - Marginals and Conditionals of Gaussians
  - Sums and Linear Transformations
- 2 Change of Variables
  - Distribution Function Technique
  - Change of Variables

# What We have Learnt From the Calculus Course

$$\int f(g(x))g'(x)dx = \int f(u)du, \text{ where } u = g(x).$$

# What We have Learnt From the Calculus Course

$$\int f(g(x))g'(x)dx = \int f(u)du, \text{ where } u = g(x).$$

- Intuitively, considering  $du \approx \Delta u = g'(x)\Delta x$  as the “small changes”.

# The Roadmap (1/2)

- Consider a univariate random variable  $X$  and an invertible function  $U$  such that  $Y := U(X)$ .
- Assume that  $X$  has states  $x \in [a, b]$ .
- By the definition of a cdf, we have

# The Roadmap (1/2)

- Consider a univariate random variable  $X$  and an invertible function  $U$  such that  $Y := U(X)$ .
- Assume that  $X$  has states  $x \in [a, b]$ .
- By the definition of a cdf, we have

$$F_Y(y) = \Pr[Y \leq y]$$

# The Roadmap (1/2)

- Consider a univariate random variable  $X$  and an invertible function  $U$  such that  $Y := U(X)$ .
- Assume that  $X$  has states  $x \in [a, b]$ .
- By the definition of a cdf, we have

$$F_Y(y) = \Pr[Y \leq y] = \Pr[U(X) \leq y]$$

# The Roadmap (1/2)

- Consider a univariate random variable  $X$  and an invertible function  $U$  such that  $Y := U(X)$ .
- Assume that  $X$  has states  $x \in [a, b]$ .
- By the definition of a cdf, we have

$$F_Y(y) = \Pr[Y \leq y] = \Pr[U(X) \leq y]$$

If  $U$  is *strictly increasing*, then so is its inverse  $U^{-1}$ .

$$\Pr[U(X) \leq y] = \Pr[U^{-1}(U(X)) \leq U^{-1}(y)]$$



# The Roadmap (1/2)

- Consider a univariate random variable  $X$  and an invertible function  $U$  such that  $Y := U(X)$ .
- Assume that  $X$  has states  $x \in [a, b]$ .
- By the definition of a cdf, we have

$$F_Y(y) = \Pr[Y \leq y] = \Pr[U(X) \leq y]$$

If  $U$  is *strictly increasing*, then so is its inverse  $U^{-1}$ .

$$\Pr[U(X) \leq y] = \Pr[U^{-1}(U(X)) \leq U^{-1}(y)] = \Pr[X \leq U^{-1}(y)].$$

$$\text{Then, } F_Y(y) = \Pr[X \leq U^{-1}(y)] = \int_a^{U^{-1}(y)} f_X(x) dx$$

## The Roadmap (2/2)

- To obtain the pdf, we differentiate  $F_Y(y)$  w.r.t.  $y$ :

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} \int_a^{U^{-1}(y)} f_X(x) dx.$$

## The Roadmap (2/2)

- To obtain the pdf, we differentiate  $F_Y(y)$  w.r.t.  $y$ :

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} \int_a^{U^{-1}(y)} f_X(x) dx.$$

- The integral on the right-hand side is w.r.t.  $x$ , but we need an integral w.r.t.  $y$  ( $\because$  we are differentiating w.r.t.  $y$ ...)
- Change-of-variable comes to the rescue!

## The Roadmap (2/2)

- To obtain the pdf, we differentiate  $F_Y(y)$  w.r.t.  $y$ :

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} \int_a^{U^{-1}(y)} f_X(x) dx.$$

- The integral on the right-hand side is w.r.t.  $x$ , but we need an integral **w.r.t.  $y$**  ( $\because$  we are differentiating w.r.t.  $y$ ...)
  - Change-of-variable comes to the rescue!
- $\int f_X(U^{-1}(y)) U^{-1'}(y) dy = \int f_X(x) dx$ , where  $x = U^{-1}(y)$ .

## The Roadmap (2/2)

- To obtain the pdf, we differentiate  $F_Y(y)$  w.r.t.  $y$ :

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} \int_a^{U^{-1}(y)} f_X(x) dx.$$

- The integral on the right-hand side is w.r.t.  $x$ , but we need an integral **w.r.t.  $y$**  ( $\because$  we are differentiating w.r.t.  $y$ ...)
- Change-of-variable comes to the rescue!

- $\int f_X(U^{-1}(y)) U^{-1}'(y) dy = \int f_X(x) dx$ , where  $x = U^{-1}(y)$ .

- Thus,

$$\begin{aligned} f_Y(y) &= \frac{d}{dy} \int_a^{U^{-1}(y)} f_X(U^{-1}(y)) U^{-1}'(y) dy \\ &= f_X(U^{-1}(y)) \cdot \left( \frac{d}{dy} U^{-1}(y) \right). \end{aligned}$$

## Remark

For decreasing functions,

$$f_Y(y) = -f_X(U^{-1}(y)) \cdot \left( \frac{d}{dy} U^{-1}(y) \right).$$

# Remark

For decreasing functions,

$$f_Y(y) = -f_X(U^{-1}(y)) \cdot \left( \frac{d}{dy} U^{-1}(y) \right).$$

So for both increasing and decreasing  $U$ ,

$$f_Y(y) = f_X(U^{-1}(y)) \cdot \left| \frac{d}{dy} U^{-1}(y) \right|.$$

## Remark

For decreasing functions,

$$f_Y(y) = -f_X(U^{-1}(y)) \cdot \left( \frac{d}{dy} U^{-1}(y) \right).$$

So for both increasing and decreasing  $U$ ,

$$f_Y(y) = f_X(U^{-1}(y)) \cdot \left| \frac{d}{dy} U^{-1}(y) \right|.$$

- The term  $\left| \frac{d}{dy} U^{-1}(y) \right|$  measures how much a unit volume changes when applying  $U$ .



# The Main Theorem

## Theorem [Billingsley (1995)]

Let  $f_X(\mathbf{x})$  be the pdf of the multivariate continuous random variable  $X$ . If the **vector-valued** function  $\mathbf{y} = U(\mathbf{x})$  is **differentiable** and **invertible** for all values within the domain of  $\mathbf{x}$ , then for corresponding values of  $\mathbf{y}$ , the pdf of  $Y = U(X)$  is given by

$$f(\mathbf{y}) = f_{\mathbf{x}}(U^{-1}(\mathbf{y})) \cdot \left| \det \left( \frac{\partial}{\partial \mathbf{y}} U^{-1}(\mathbf{y}) \right) \right|.$$

# Example

## Example

Consider a bivariate random variable  $X$  with states  $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$  and pdf

$$f\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \frac{1}{2\pi} \exp\left(-\frac{1}{2} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}^\top \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right).$$

Then, consider a matrix  $\mathbf{A} \in \mathbb{R}^{2 \times 2}$  defined as

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

**Goal:** Find the pdf of the random variable  $Y$  with states  $\mathbf{y} = \mathbf{A}\mathbf{x}$ .

- $\mathbf{y} = \mathbf{A}\mathbf{x}$

$$\bullet \mathbf{y} = \mathbf{A}\mathbf{x} \implies \mathbf{x} = \mathbf{A}^{-1}\mathbf{y}.$$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mathbf{A}^{-1} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

$$\bullet \mathbf{y} = \mathbf{A}\mathbf{x} \implies \mathbf{x} = \mathbf{A}^{-1}\mathbf{y}.$$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mathbf{A}^{-1} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}.$$

- $\mathbf{y} = \mathbf{A}\mathbf{x} \implies \mathbf{x} = \mathbf{A}^{-1}\mathbf{y}.$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mathbf{A}^{-1} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}.$$

- The corresponding pdf is given by

$$f(\mathbf{x}) = f(\mathbf{A}^{-1}\mathbf{y}) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}\mathbf{y}^\top \mathbf{A}^{-1}\mathbf{A}^{-1}\mathbf{y}\right)$$

- $\mathbf{y} = \mathbf{A}\mathbf{x} \implies \mathbf{x} = \mathbf{A}^{-1}\mathbf{y}.$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mathbf{A}^{-1} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}.$$

- The corresponding pdf is given by

$$f(\mathbf{x}) = f(\mathbf{A}^{-1}\mathbf{y}) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}\mathbf{y}^\top \mathbf{A}^{-1} \mathbf{A}^{-1} \mathbf{y}\right)$$

- $\frac{\partial}{\partial \mathbf{y}} \mathbf{A}^{-1} \mathbf{y} =$

- $\mathbf{y} = \mathbf{A}\mathbf{x} \implies \mathbf{x} = \mathbf{A}^{-1}\mathbf{y}.$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mathbf{A}^{-1} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}.$$

- The corresponding pdf is given by

$$f(\mathbf{x}) = f(\mathbf{A}^{-1}\mathbf{y}) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}\mathbf{y}^\top \mathbf{A}^{-1}\mathbf{A}^{-1}\mathbf{y}\right)$$

- $\frac{\partial}{\partial \mathbf{y}} \mathbf{A}^{-1}\mathbf{y} = \mathbf{A}^{-1}$ . So,  $\det\left(\frac{\partial}{\partial \mathbf{y}} \mathbf{A}^{-1}\mathbf{y}\right) = \det(\mathbf{A}^{-1}) =$



- $\mathbf{y} = \mathbf{A}\mathbf{x} \implies \mathbf{x} = \mathbf{A}^{-1}\mathbf{y}.$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mathbf{A}^{-1} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}.$$

- The corresponding pdf is given by

$$f(\mathbf{x}) = f(\mathbf{A}^{-1}\mathbf{y}) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}\mathbf{y}^\top \mathbf{A}^{-1}\mathbf{A}^{-1}\mathbf{y}\right)$$

- $\frac{\partial}{\partial \mathbf{y}} \mathbf{A}^{-1}\mathbf{y} = \mathbf{A}^{-1}$ . So,  $\det\left(\frac{\partial}{\partial \mathbf{y}} \mathbf{A}^{-1}\mathbf{y}\right) = \det(\mathbf{A}^{-1}) = \frac{1}{ad - bc}.$

- $\mathbf{y} = \mathbf{A}\mathbf{x} \implies \mathbf{x} = \mathbf{A}^{-1}\mathbf{y}.$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mathbf{A}^{-1} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}.$$

- The corresponding pdf is given by

$$f(\mathbf{x}) = f(\mathbf{A}^{-1}\mathbf{y}) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}\mathbf{y}^\top \mathbf{A}^{-1} \mathbf{A}^{-1} \mathbf{y}\right)$$

- $\frac{\partial}{\partial \mathbf{y}} \mathbf{A}^{-1} \mathbf{y} = \mathbf{A}^{-1}$ . So,  $\det\left(\frac{\partial}{\partial \mathbf{y}} \mathbf{A}^{-1} \mathbf{y}\right) = \det(\mathbf{A}^{-1}) = \frac{1}{ad - bc}.$

- Thus,  $f(\mathbf{y}) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}\mathbf{y}^\top \mathbf{A}^{-1} \mathbf{A}^{-1} \mathbf{y}\right) \cdot \frac{1}{ad - bc}.$

# Discussions