



# Gatsby Bridging Program

## Probability: Functions of Random Variables

---

Cameron Stewart

Gatsby Computational Neuroscience Unit

# Today's Lecture Topics

1. Prerequisites and Motivation
2. The CDF Method for One-Dimensional Codomains
3. The Method of Transformations for Bijective Functions
4. The Method of Transformations for Non-Injective, Surjective Functions
5. Dynamical Systems as Functions: A Fun Proof to Look at After the Course

## Prerequisites and Motivation

---

# Revision: Injective, Surjective, and Bijective Functions

Today's content will be based heavily on the concepts of injectivity, surjectivity, and bijectivity. So, here is some revision.

$g : \mathcal{X} \rightarrow \mathcal{Y}$  implies that the function  $g$  takes elements from the set  $\mathcal{X}$  and maps them to elements of the set  $\mathcal{Y}$ .  $\mathcal{X}$  is the domain and  $\mathcal{Y}$  is the codomain.

# Revision: Injective, Surjective, and Bijective Functions

Today's content will be based heavily on the concepts of injectivity, surjectivity, and bijectivity. So, here is some revision.

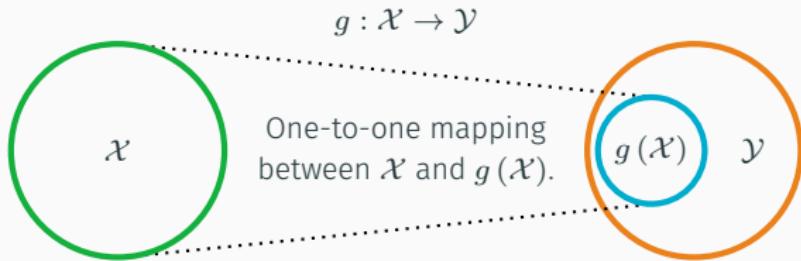
$g : \mathcal{X} \rightarrow \mathcal{Y}$  implies that the function  $g$  takes elements from the set  $\mathcal{X}$  and maps them to elements of the set  $\mathcal{Y}$ .  $\mathcal{X}$  is the domain and  $\mathcal{Y}$  is the codomain.

Not all elements of  $\mathcal{Y}$  may be attainable for a given function. The set of elements which are attainable is called the image, given by

$$g(\mathcal{X}) = \{g(x) \mid x \in \mathcal{X}\} .$$

Clearly,  $g(\mathcal{X}) \subseteq \mathcal{Y}$ .

# Revision: Injective Functions

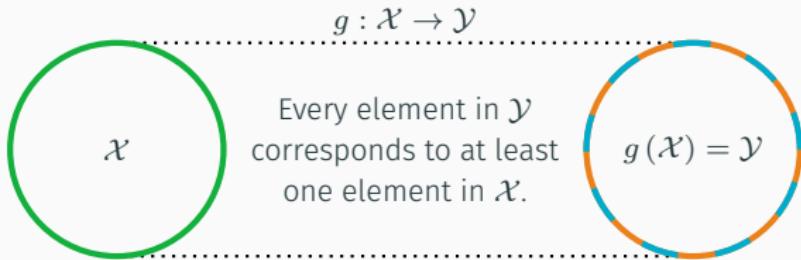


## Injection

If every  $x \in \mathcal{X}$  has a unique mapped value  $g(x)$ , then  $g$  is injective.

$$\forall x_1, x_2 \in \mathcal{X}, g(x_1) = g(x_2) \Rightarrow x_1 = x_2$$

# Revision: Surjective Functions

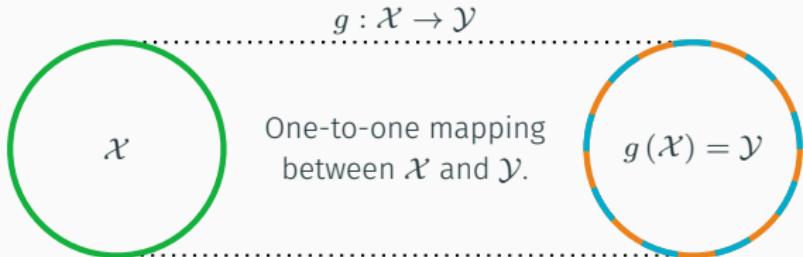


## Surjection

If every  $y \in \mathcal{Y}$  has at least one solution to  $y = g(x)$  for  $x \in \mathcal{X}$ , then  $g$  is surjective.

$$\forall y \in \mathcal{Y}, \exists x \in \mathcal{X} \text{ such that } y = g(x)$$

# Revision: Bijective Functions



## Bijection

If  $g$  is injective and surjective, then it is called bijective.

$$\forall y \in \mathcal{Y}, \exists! x \in \mathcal{X} \text{ such that } y = g(x)$$

Bijective functions have inverses. We denote  $g^{-1}$  as the inverse of  $g$ .

# Revision: Injective, Surjective, and Bijective Functions

## Example 1

Classify the functions  $g$  given by  $g(x) = x^2$  with the following domains and codomains:

- $g : \mathbb{R} \rightarrow \mathbb{R}$ .
- $g : [0, \infty) \rightarrow \mathbb{R}$ .
- $g : \mathbb{R} \rightarrow [0, \infty)$ .
- $g : [0, \infty) \rightarrow [0, \infty)$ .

# Revision: Injective, Surjective, and Bijective Functions

## Example 1

Classify the functions  $g$  given by  $g(x) = x^2$  with the following domains and codomains:

- $g : \mathbb{R} \rightarrow \mathbb{R}$ . Neither injective or surjective.
- $g : [0, \infty) \rightarrow \mathbb{R}$ .
- $g : \mathbb{R} \rightarrow [0, \infty)$ .
- $g : [0, \infty) \rightarrow [0, \infty)$ .

# Revision: Injective, Surjective, and Bijective Functions

## Example 1

Classify the functions  $g$  given by  $g(x) = x^2$  with the following domains and codomains:

- $g : \mathbb{R} \rightarrow \mathbb{R}$ . Neither injective or surjective.
- $g : [0, \infty) \rightarrow \mathbb{R}$ . Injective but not surjective.
- $g : \mathbb{R} \rightarrow [0, \infty)$ .
- $g : [0, \infty) \rightarrow [0, \infty)$ .

# Revision: Injective, Surjective, and Bijective Functions

## Example 1

Classify the functions  $g$  given by  $g(x) = x^2$  with the following domains and codomains:

- $g : \mathbb{R} \rightarrow \mathbb{R}$ . Neither injective or surjective.
- $g : [0, \infty) \rightarrow \mathbb{R}$ . Injective but not surjective.
- $g : \mathbb{R} \rightarrow [0, \infty)$ . Surjective but not injective.
- $g : [0, \infty) \rightarrow [0, \infty)$ .

# Revision: Injective, Surjective, and Bijective Functions

## Example 1

Classify the functions  $g$  given by  $g(x) = x^2$  with the following domains and codomains:

- $g : \mathbb{R} \rightarrow \mathbb{R}$ . Neither injective or surjective.
- $g : [0, \infty) \rightarrow \mathbb{R}$ . Injective but not surjective.
- $g : \mathbb{R} \rightarrow [0, \infty)$ . Surjective but not injective.
- $g : [0, \infty) \rightarrow [0, \infty)$ . Injective and surjective, hence bijective. Bijectivity implies an inverse exists:  $g^{-1}(y) = \sqrt{y}$  with  $g^{-1} : [0, \infty) \rightarrow [0, \infty)$ .

# Revision: Injective, Surjective, and Bijective Functions

## Example 1

Classify the functions  $g$  given by  $g(x) = x^2$  with the following domains and codomains:

- $g : \mathbb{R} \rightarrow \mathbb{R}$ . Neither injective or surjective.
- $g : [0, \infty) \rightarrow \mathbb{R}$ . Injective but not surjective.
- $g : \mathbb{R} \rightarrow [0, \infty)$ . Surjective but not injective.
- $g : [0, \infty) \rightarrow [0, \infty)$ . Injective and surjective, hence bijective. Bijectivity implies an inverse exists:  $g^{-1}(y) = \sqrt{y}$  with  $g^{-1} : [0, \infty) \rightarrow [0, \infty)$ .

In today's lecture, you may assume all the functions used in examples are surjective even if the codomains aren't explicitly mentioned.

# Motivation

Functions of random variables arise in various situations. You will learn about some of these situations in this lecture. Some of the most impressive uses of functions of random variables can be found in deep generative models. Two examples of this:

- **Generative adversarial networks (GANs)**: The generator function takes a multivariate Gaussian random vector and outputs a random vector of much higher dimensionality. It is non-surjective and likely non-injective. The generator is trained so that the distribution over the output resembles that of natural images.

# Motivation

Functions of random variables arise in various situations. You will learn about some of these situations in this lecture. Some of the most impressive uses of functions of random variables can be found in deep generative models. Two examples of this:

- **Generative adversarial networks (GANs)**: The generator function takes a multivariate Gaussian random vector and outputs a random vector of much higher dimensionality. It is non-surjective and likely non-injective. The generator is trained so that the distribution over the output resembles that of natural images.
- **Normalising flows**: A function takes a multivariate Gaussian random vector and outputs a random vector of the same dimensionality. This function is constrained to be bijective. Like GANs, it is trained so that the distribution over the output resembles that of natural images.

# Motivation

**Figure 1:** Samples from a normalising flow model created by OpenAI in 2018<sup>1</sup>.

---

<sup>1</sup>Prafulla Dhariwal and Durk Kingma. July 2018. URL: <https://openai.com/research/glow>.

# The CDF Method for One-Dimensional Codomains

---

## The CDF Method

Here we will look at one method for determining the CDF of the random variable  $Y = g(\mathbf{X})$  for some function  $g : \mathcal{X} \rightarrow \mathcal{Y}$ . We will constrain ourselves to dealing with cases where  $Y$  is one-dimensional.

# The CDF Method

Here we will look at one method for determining the CDF of the random variable  $Y = g(\mathbf{X})$  for some function  $g : \mathcal{X} \rightarrow \mathcal{Y}$ . We will constrain ourselves to dealing with cases where  $Y$  is one-dimensional.

## The CDF Method

Given  $Y = g(\mathbf{X})$  and the CDF  $F_{\mathbf{X}}$  for  $\mathbf{X}$ , we can find the CDF  $F_Y$  for  $Y$  with the following steps:

1. Start by writing

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \leq y) \\ &= \mathbb{P}(g(\mathbf{X}) \leq y). \end{aligned}$$

# The CDF Method

Here we will look at one method for determining the CDF of the random variable  $Y = g(\mathbf{X})$  for some function  $g : \mathcal{X} \rightarrow \mathcal{Y}$ . We will constrain ourselves to dealing with cases where  $Y$  is one-dimensional.

## The CDF Method

Given  $Y = g(\mathbf{X})$  and the CDF  $F_{\mathbf{X}}$  for  $\mathbf{X}$ , we can find the CDF  $F_Y$  for  $Y$  with the following steps:

1. Start by writing

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \leq y) \\ &= \mathbb{P}(g(\mathbf{X}) \leq y). \end{aligned}$$

2. Manipulate  $\mathbb{P}(g(\mathbf{X}) \leq y)$  until it is written in terms of known functions (e.g.  $F_{\mathbf{X}}$ ). Substitute in these functions and simplify.

# The CDF Method

Here we will look at one method for determining the CDF of the random variable  $Y = g(\mathbf{X})$  for some function  $g : \mathcal{X} \rightarrow \mathcal{Y}$ . We will constrain ourselves to dealing with cases where  $Y$  is one-dimensional.

## The CDF Method

Given  $Y = g(\mathbf{X})$  and the CDF  $F_{\mathbf{X}}$  for  $\mathbf{X}$ , we can find the CDF  $F_Y$  for  $Y$  with the following steps:

1. Start by writing

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \leq y) \\ &= \mathbb{P}(g(\mathbf{X}) \leq y). \end{aligned}$$

2. Manipulate  $\mathbb{P}(g(\mathbf{X}) \leq y)$  until it is written in terms of known functions (e.g.  $F_{\mathbf{X}}$ ). Substitute in these functions and simplify.
3. (Optional) Given the CDF, find the PMF/PDF.

# The CDF Method

A reminder regarding step 3:

- For discrete  $Y$ , we take differences of the CDF to find the PMF. E.g.  
 $f_Y(y) = F_Y(y) - F_Y(y - 1)$  for  $y \in \mathcal{Y}$  when the support is a subset of the integers. In general,  $\mathbb{P}(Y = y) = \mathbb{P}(Y \leq y) - \mathbb{P}(Y < y)$ .

# The CDF Method

A reminder regarding step 3:

- For discrete  $Y$ , we take differences of the CDF to find the PMF. E.g.  
 $f_Y(y) = F_Y(y) - F_Y(y - 1)$  for  $y \in \mathcal{Y}$  when the support is a subset of the integers. In general,  $\mathbb{P}(Y = y) = \mathbb{P}(Y \leq y) - \mathbb{P}(Y < y)$ .
- For continuous  $Y$ , we take the derivative of the CDF with respect to  $y$  to find the PDF:  $f_Y(y) = \frac{d}{dy}F_Y(y)$ .

# A Simple Discrete Example

## Example 2

Given  $X \sim \mathcal{U}\{1, 3\}$ , what is the PMF of  $Y = 2X$ ?

# A Simple Discrete Example

## Example 2

Given  $X \sim \mathcal{U}\{1, 3\}$ , what is the PMF of  $Y = 2X$ ?

First, consider the supports. The support of  $X$  is  $\mathcal{X} = \{1, 2, 3\}$ , so the support of  $Y$  will be  $\mathcal{Y} = \{2, 4, 6\}$ . Also, we know that

$$F_X(x) = \begin{cases} 0 & \text{if } x < 1 \\ \frac{1}{3} & \text{if } 1 \leq x < 2 \\ \frac{2}{3} & \text{if } 2 \leq x < 3 \\ 1 & \text{if } x \geq 3 \end{cases},$$

so let's try to find an expression for  $F_Y$  in terms of  $F_X$ .

# A Simple Discrete Example

## Example 2

The CDF for  $Y$  is given by

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \leq y) \\ &= \mathbb{P}(2X \leq y) \\ &= \mathbb{P}\left(X \leq \frac{y}{2}\right) \\ &= F_X\left(\frac{y}{2}\right) \\ &= \begin{cases} 0 & \text{if } y < 2 \\ \frac{1}{3} & \text{if } 2 \leq y < 4 \\ \frac{2}{3} & \text{if } 4 \leq y < 6 \\ 1 & \text{if } y \geq 6 \end{cases}. \end{aligned}$$

# A Simple Discrete Example

## Example 2

For all  $y \in \mathcal{Y}$ , the PMF is then given by

$$\begin{aligned}f_Y(y) &= F_Y(y) - F_Y(y-2) \\&= F_X\left(\frac{y}{2}\right) - F_X\left(\frac{y}{2}-1\right) \\&= \begin{cases} \frac{1}{3} & \text{if } y \in \mathcal{Y} \\ 0 & \text{otherwise} \end{cases}.\end{aligned}$$

# A Simple Continuous Example

## Example 3

Given  $X \sim \mathcal{U}[0, 1]$ , what is the PDF of  $Y = -2X$ ?

# A Simple Continuous Example

## Example 3

Given  $X \sim \mathcal{U}[0, 1]$ , what is the PDF of  $Y = -2X$ ?

First, consider the supports. The support of  $X$  is  $\mathcal{X} = [0, 1]$ , so the support of  $Y$  will be  $\mathcal{Y} = [-2, 0]$ . Also, we know that  $F_X(x) = x$  for  $x \in \mathcal{X}$ , so let's try to find an expression for  $F_Y$  in terms of  $F_X$ . The CDF for  $Y$  is given by

$$\begin{aligned}F_Y(y) &= \mathbb{P}(Y \leq y) \\&= \mathbb{P}(-2X \leq y) \\&= \mathbb{P}\left(X \geq -\frac{y}{2}\right) \\&= 1 - \mathbb{P}\left(X \leq -\frac{y}{2}\right) \\&= 1 - F_X\left(-\frac{y}{2}\right) \\&= 1 + \frac{y}{2} \quad \text{for } y \in \mathcal{Y}.\end{aligned}$$

# A Simple Continuous Example

## Example 3

For all  $y \in \mathcal{Y}$ , the PDF is then given by

$$\begin{aligned}f_Y(y) &= \frac{d}{dy} F_Y(y) \\&= \frac{d}{dy} \left(1 + \frac{y}{2}\right) \\&= \frac{1}{2}.\end{aligned}$$

# A Simple Continuous Example

## Example 3

For all  $y \in \mathcal{Y}$ , the PDF is then given by

$$\begin{aligned}f_Y(y) &= \frac{d}{dy} F_Y(y) \\&= \frac{d}{dy} \left(1 + \frac{y}{2}\right) \\&= \frac{1}{2}.\end{aligned}$$

Take note that the support interval doubled in size, and the PDF halved in magnitude to compensate. This is in contrast to the discrete example, where the support size and probabilities remained the same.

## Non-Injective Examples

So far we have only considered injective functions in our examples. Let's consider some non-injective functions. These problems may require a bit more thought and effort.

# Sums of Random Variables

## Example 4

Let  $X$  and  $Y$  be arbitrary independent random variables with supports  $\mathcal{X}$  and  $\mathcal{Y}$  respectively. What is the PMF/PDF of  $Z = X + Y$ ?

# Sums of Random Variables

## Example 4

Let  $X$  and  $Y$  be arbitrary independent random variables with supports  $\mathcal{X}$  and  $\mathcal{Y}$  respectively. What is the PMF/PDF of  $Z = X + Y$ ?

Let's start with the CDF:

$$\begin{aligned} F_Z(z) &= \mathbb{P}(Z \leq z) \\ &= \mathbb{P}(X + Y \leq z) \end{aligned}$$

Discrete  $X$ :

$$\begin{aligned} &= \sum_{x \in \mathcal{X}} \mathbb{P}(x + Y \leq z, X = x) \\ &= \sum_{x \in \mathcal{X}} \mathbb{P}(Y \leq z - x) \mathbb{P}(X = x) \\ &= \sum_{x \in \mathcal{X}} F_Y(z - x) f_X(x) \end{aligned}$$

Continuous  $X$ :

$$\begin{aligned} &= \int_{\mathcal{X}} \mathbb{P}(x + Y \leq z) f_X(x) dx \\ &= \int_{\mathcal{X}} \mathbb{P}(Y \leq z - x) f_X(x) dx \\ &= \int_{\mathcal{X}} F_Y(z - x) f_X(x) dx \end{aligned}$$

# Sums of Random Variables

## Example 4

Now we can find the PMF/PDF. In the left column we assume the support of  $Y$  is a subset of the integers. In the right column we assume  $Y$  is continuous.

$$f_Z(z)$$

Discrete  $X$  and  $Y$ :

$$\begin{aligned} &= F_Z(z) - F_Z(z-1) \\ &= \sum_{x \in \mathcal{X}} (F_Y(z-x) \\ &\quad - F_Y(z-1-x)) f_X(x) \\ &= \sum_{x \in \mathcal{X}} f_Y(z-x) f_X(x) \end{aligned}$$

Continuous  $X$  and  $Y$ :

$$\begin{aligned} &= \frac{d}{dz} F_Z(z) \\ &= \int_{\mathcal{X}} \frac{d}{dz} F_Y(z-x) f_X(x) dx \\ &= \int_{\mathcal{X}} f_Y(z-x) f_X(x) dx \end{aligned}$$

These operations are called convolutions.

# Squaring a Gaussian

## Example 5

What is the PDF of  $Y = X^2$  given  $X \sim \mathcal{N}(0, 1)$ ?

# Squaring a Gaussian

## Example 5

What is the PDF of  $Y = X^2$  given  $X \sim \mathcal{N}(0, 1)$ ?

The CDF for  $Y$  is given by

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \leq y) \\ &= \mathbb{P}(X^2 \leq y) \\ &= \mathbb{P}(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= \mathbb{P}(X \leq \sqrt{y}) - \mathbb{P}(X \leq -\sqrt{y}) \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}) \\ &= \frac{1}{2} \left( 1 + \text{erf} \left( \sqrt{\frac{y}{2}} \right) \right) - \frac{1}{2} \left( 1 + \text{erf} \left( -\sqrt{\frac{y}{2}} \right) \right) \\ &= \frac{1}{2} \left( \text{erf} \left( \sqrt{\frac{y}{2}} \right) - \text{erf} \left( -\sqrt{\frac{y}{2}} \right) \right). \end{aligned}$$

# Squaring a Gaussian

## Example 5

The PDF is then given by

$$\begin{aligned}f_Y(y) &= \frac{d}{dy} F_Y(y) \\&= \frac{d}{dy} \frac{1}{2} \left( \operatorname{erf}\left(\sqrt{\frac{y}{2}}\right) - \operatorname{erf}\left(-\sqrt{\frac{y}{2}}\right) \right) \\&= \frac{1}{2} \left( \frac{2}{\sqrt{\pi}} e^{-y/2} \cdot \frac{1}{2} \sqrt{\frac{2}{y}} \cdot \frac{1}{2} + \frac{2}{\sqrt{\pi}} e^{-y/2} \cdot \frac{1}{2} \sqrt{\frac{2}{y}} \cdot \frac{1}{2} \right) \\&= \frac{1}{\sqrt{2\pi y}} e^{-y/2}.\end{aligned}$$

# The Maximum of a Set of Random Variables

## Example 6

If  $Y = \max \{X_1, \dots, X_n\}$  for  $n$  independent continuous random variables  $X_1, \dots, X_n$ , what is the PDF of  $Y$ ?

# The Maximum of a Set of Random Variables

## Example 6

If  $Y = \max \{X_1, \dots, X_n\}$  for  $n$  independent continuous random variables  $X_1, \dots, X_n$ , what is the PDF of  $Y$ ?

The CDF for  $Y$  is given by

$$\begin{aligned}F_Y(y) &= \mathbb{P}(Y \leq y) \\&= \mathbb{P}(\max \{X_1, \dots, X_n\} \leq y) \\&= \mathbb{P}(X_1 \leq y, \dots, X_n \leq y) \\&= \prod_{i=1}^n \mathbb{P}(X_i \leq y) \\&= \prod_{i=1}^n F_{X_i}(y).\end{aligned}$$

# The Maximum of a Set of Random Variables

## Example 6

The PDF for  $Y$  is then given by

$$\begin{aligned}f_Y(y) &= \frac{d}{dy} F_Y(y) \\&= \frac{d}{dy} \prod_{i=1}^n F_{X_i}(y) \\&= \sum_{i=1}^n f_{X_i}(y) \prod_{j \neq i} F_{X_j}(y).\end{aligned}$$

Intermission

# The Method of Transformations for Bijective Functions

---

## Bijective Functions of Discrete Random Variables

If  $\mathbf{g}$  is a bijective function and  $\mathbf{X}$  is a discrete random variable with PMF  $f_{\mathbf{X}}$ , then  $\mathbf{Y} = \mathbf{g}(\mathbf{X})$  has PMF  $f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(\mathbf{g}^{-1}(\mathbf{y}))$ . I.e. the mapping may change the support, but doesn't change probabilities. Because the probabilities don't change,  $f_{\mathbf{Y}}(\mathbf{y})$  still sums to one over the transformed support.

# Bijective Functions of Discrete Random Variables

If  $\mathbf{g}$  is a bijective function and  $\mathbf{X}$  is a discrete random variable with PMF  $f_{\mathbf{X}}$ , then  $\mathbf{Y} = \mathbf{g}(\mathbf{X})$  has PMF  $f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(\mathbf{g}^{-1}(\mathbf{y}))$ . I.e. the mapping may change the support, but doesn't change probabilities. Because the probabilities don't change,  $f_{\mathbf{Y}}(\mathbf{y})$  still sums to one over the transformed support.

*Proof:*

$$\begin{aligned}\mathbb{P}(\mathbf{Y} = \mathbf{y}) &= \mathbb{P}(\mathbf{g}(\mathbf{X}) = \mathbf{y}) \\ f_{\mathbf{Y}}(\mathbf{y}) &= \mathbb{P}(\mathbf{X} = \mathbf{g}^{-1}(\mathbf{y})) \\ &= f_{\mathbf{X}}(\mathbf{g}^{-1}(\mathbf{y}))\end{aligned}$$

□

The first example in this lecture demonstrates this well.

# Bijective Functions of Continuous Random Variables

The discrete case is very simple. The continuous case is more complicated.

## The Method of Transformations

For continuous random variables  $\mathbf{X}$  and  $\mathbf{Y} = \mathbf{g}(\mathbf{X})$ , where  $\mathbf{g}$  is a bijective function, we have the following relationship between the PDFs  $f_{\mathbf{X}}$  and  $f_{\mathbf{Y}}$ :

$$\begin{aligned}f_{\mathbf{Y}}(\mathbf{y}) &= f_{\mathbf{X}}(\mathbf{g}^{-1}(\mathbf{y})) |\det(D\mathbf{g}^{-1}(\mathbf{y}))| \\&= f_{\mathbf{X}}(\mathbf{g}^{-1}(\mathbf{y})) |\det(D\mathbf{g}(\mathbf{g}^{-1}(\mathbf{y})))|^{-1}.\end{aligned}$$

The notation  $D\mathbf{g}(\mathbf{x})$  means "the Jacobian of  $\mathbf{g}$ , evaluated at  $\mathbf{x}$ ".

# Bijective Functions of Continuous Random Variables

The discrete case is very simple. The continuous case is more complicated.

## The Method of Transformations

For continuous random variables  $\mathbf{X}$  and  $\mathbf{Y} = \mathbf{g}(\mathbf{X})$ , where  $\mathbf{g}$  is a bijective function, we have the following relationship between the PDFs  $f_{\mathbf{X}}$  and  $f_{\mathbf{Y}}$ :

$$\begin{aligned}f_{\mathbf{Y}}(\mathbf{y}) &= f_{\mathbf{X}}(\mathbf{g}^{-1}(\mathbf{y})) |\det(D\mathbf{g}^{-1}(\mathbf{y}))| \\&= f_{\mathbf{X}}(\mathbf{g}^{-1}(\mathbf{y})) |\det(D\mathbf{g}(\mathbf{g}^{-1}(\mathbf{y})))|^{-1}.\end{aligned}$$

The notation  $D\mathbf{g}(\mathbf{x})$  means "the Jacobian of  $\mathbf{g}$ , evaluated at  $\mathbf{x}$ ".

In the one-dimensional case, this simplifies to:

$$\begin{aligned}f_Y(y) &= f_X(g^{-1}(y)) \left| \frac{dg^{-1}}{dy}(y) \right| \\&= f_X(g^{-1}(y)) \left| \frac{dg}{dx}(g^{-1}(y)) \right|^{-1}.\end{aligned}$$

# Bijective Functions of Continuous Random Variables

*Proof:*

Let  $A$  be some subset of the support of  $\mathbf{Y}$ . Then

$$\begin{aligned}\mathbb{P}(\mathbf{Y} \in A) &= \mathbb{P}(\mathbf{g}(\mathbf{X}) \in A) \\ \int_A f_{\mathbf{Y}}(\mathbf{y}) d\mathbf{y} &= \mathbb{P}(\mathbf{X} \in \mathbf{g}^{-1}(A)) \\ &= \int_{\mathbf{g}^{-1}(A)} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}.\end{aligned}$$

Performing integration by substitution on the right side of the equation with  $\mathbf{y} = \mathbf{g}(\mathbf{x})$ , giving us  $d\mathbf{x} = |\det(D\mathbf{g}^{-1}(\mathbf{y}))| d\mathbf{y}$ , we see that

$$\begin{aligned}\int_A f_{\mathbf{Y}}(\mathbf{y}) d\mathbf{y} &= \int_A f_{\mathbf{X}}(\mathbf{g}^{-1}(\mathbf{y})) |\det(D\mathbf{g}^{-1}(\mathbf{y}))| d\mathbf{y} \\ \therefore f_{\mathbf{Y}}(\mathbf{y}) &= f_{\mathbf{X}}(\mathbf{g}^{-1}(\mathbf{y})) |\det(D\mathbf{g}^{-1}(\mathbf{y}))|.\end{aligned}$$

□

# Bijective Functions of Continuous Random Variables

*Proof:*

Let  $A$  be some subset of the support of  $\mathbf{Y}$ . Then

$$\begin{aligned}\mathbb{P}(\mathbf{Y} \in A) &= \mathbb{P}(\mathbf{g}(\mathbf{X}) \in A) \\ \int_A f_{\mathbf{Y}}(\mathbf{y}) d\mathbf{y} &= \mathbb{P}(\mathbf{X} \in \mathbf{g}^{-1}(A)) \\ &= \int_{\mathbf{g}^{-1}(A)} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}.\end{aligned}$$

Performing integration by substitution on the right side of the equation with  $\mathbf{y} = \mathbf{g}(\mathbf{x})$ , giving us  $d\mathbf{x} = |\det(D\mathbf{g}^{-1}(\mathbf{y}))| d\mathbf{y}$ , we see that

$$\begin{aligned}\int_A f_{\mathbf{Y}}(\mathbf{y}) d\mathbf{y} &= \int_A f_{\mathbf{X}}(\mathbf{g}^{-1}(\mathbf{y})) |\det(D\mathbf{g}^{-1}(\mathbf{y}))| d\mathbf{y} \\ \therefore f_{\mathbf{Y}}(\mathbf{y}) &= f_{\mathbf{X}}(\mathbf{g}^{-1}(\mathbf{y})) |\det(D\mathbf{g}^{-1}(\mathbf{y}))|.\end{aligned}$$

Alternatively, we could have used  $d\mathbf{y} = |\det(D\mathbf{g}(\mathbf{x}))| d\mathbf{x}$  in the substitution. Remember, the Jacobian of a function is the same as the reciprocal of the Jacobian of the inverse function. □

# The Log-Normal Distribution

## Example 7

If  $Y = e^X$  and  $X \sim \mathcal{N}(\mu, \sigma^2)$ , then  $Y$  is log-normally distributed. What is the PDF of the log-normal distribution?

# The Log-Normal Distribution

## Example 7

If  $Y = e^X$  and  $X \sim \mathcal{N}(\mu, \sigma^2)$ , then  $Y$  is log-normally distributed. What is the PDF of the log-normal distribution?

The support and PDF of  $X$  are  $(-\infty, \infty)$  and

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

respectively. Clearly the support of  $Y$  will be  $[0, \infty)$ .

# The Log-Normal Distribution

## Example 7

Writing  $x = \log(y)$ , we see that  $\frac{dx}{dy} = \frac{1}{y}$ . Given this information we proceed as follows:

$$\begin{aligned}f_Y(y) &= f_X(x) \left| \frac{dx}{dy} \right| \\&= f_X(\log(y)) \left| \frac{1}{y} \right| \\&= \frac{1}{y\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\log(y) - \mu)^2}{2\sigma^2}\right).\end{aligned}$$

# Inverse Transform Sampling

## Example 8

The PDF of a univariate distribution with support  $\mathcal{Y}$  is known, but we need a way to sample from it. Given a sample  $x$  from  $\mathcal{U}[0, 1]$ , what must the bijective and strictly increasing function  $g : [0, 1] \rightarrow \mathcal{Y}$  be such that  $g(x)$  is a sample from the desired distribution?

# Inverse Transform Sampling

## Example 8

The PDF of a univariate distribution with support  $\mathcal{Y}$  is known, but we need a way to sample from it. Given a sample  $x$  from  $\mathcal{U}[0, 1]$ , what must the bijective and strictly increasing function  $g : [0, 1] \rightarrow \mathcal{Y}$  be such that  $g(x)$  is a sample from the desired distribution?

Let  $Y = g(X)$  for  $X \sim \mathcal{U}[0, 1]$ .  $f_X(x) = 1$  on the support. As  $g$  is strictly increasing,  $\frac{dg}{dx} > 0$ , which also means  $\frac{dg^{-1}}{dy} > 0$ . Then

$$\begin{aligned}f_Y(y) &= f_X(g^{-1}(y)) \left| \frac{dg^{-1}}{dy}(y) \right| \\&= \frac{dg^{-1}}{dy}(y) \\g^{-1}(y) &= \int_{-\infty}^y f_Y(z) dz + c \\&= F_Y(y) + c.\end{aligned}$$

# Inverse Transform Sampling

## Example 8

When  $y \rightarrow -\infty$  we have:

$$\begin{aligned}\lim_{y \rightarrow -\infty} g^{-1}(y) &= \lim_{y \rightarrow -\infty} F_Y(y) + c \\ 0 &= 0 + c \\ c &= 0.\end{aligned}$$

Therefore,  $g(x) = F_Y^{-1}(x)$ .

# Inverse Transform Sampling

## Example 8

When  $y \rightarrow -\infty$  we have:

$$\begin{aligned}\lim_{y \rightarrow -\infty} g^{-1}(y) &= \lim_{y \rightarrow -\infty} F_Y(y) + c \\ 0 &= 0 + c \\ c &= 0.\end{aligned}$$

Therefore,  $g(x) = F_Y^{-1}(x)$ .

We have seen a proof for the inverse transform sampling method in a previous lecture. This is an alternative proof for why this method works.

# Linear Transformations of Multivariate Gaussians

## Example 9

Given  $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, I)$  with support  $\mathbb{R}^n$ , what is the PDF of  $\mathbf{Y} = A\mathbf{X}$  for some real and invertible  $n \times n$  matrix  $A$ ?

# Linear Transformations of Multivariate Gaussians

## Example 9

Given  $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, I)$  with support  $\mathbb{R}^n$ , what is the PDF of  $\mathbf{Y} = A\mathbf{X}$  for some real and invertible  $n \times n$  matrix  $A$ ?

We have  $f_{\mathbf{X}}(\mathbf{x}) = (2\pi)^{-n/2} \exp\left(-\frac{1}{2}\mathbf{x}^T \mathbf{x}\right)$ . Let's start by recognising that  $\mathbf{x} = A^{-1}\mathbf{y}$  and  $D\mathbf{y}(\mathbf{x}) = A$ . Given this information we proceed as follows:

$$\begin{aligned}f_{\mathbf{Y}}(\mathbf{y}) &= f_{\mathbf{X}}(\mathbf{x}) |\det(D\mathbf{y}(\mathbf{x}))|^{-1} \\&= f_{\mathbf{X}}(A^{-1}\mathbf{y}) |\det(A)|^{-1} \\&= (2\pi)^{-n/2} \exp\left(-\frac{1}{2}(A^{-1}\mathbf{y})^T (A^{-1}\mathbf{y})\right) (\det(A) \det(A))^{-1/2} \\&= (2\pi)^{-n/2} \exp\left(-\frac{1}{2}\mathbf{y}^T (A^T)^{-1} A^{-1}\mathbf{y}\right) (\det(A) \det(A^T))^{-1/2} \\&= (2\pi)^{-n/2} (\det(AA^T))^{-1/2} \exp\left(-\frac{1}{2}\mathbf{y}^T (AA^T)^{-1} \mathbf{y}\right).\end{aligned}$$

Therefore  $\mathbf{Y}$  is also multivariate Gaussian.  $\mathbf{Y} \sim \mathcal{N}(\mathbf{0}, AA^T)$ .

Intermission

# The Method of Transformations for Non-Injective, Surjective Functions

---

## From Non-Injectivity to Bijectivity

The method of transformations is only applicable in the case of bijective functions. There is no standard way of dealing with all non-bijective functions, but here we will discuss a helpful method for dealing with some non-injective, surjective functions.

# From Non-Injectivity to Bijectivity

The method of transformations is only applicable in the case of bijective functions. There is no standard way of dealing with all non-bijective functions, but here we will discuss a helpful method for dealing with some non-injective, surjective functions.

## The Method of Transformations for Non-Injective, Surjective Functions

For continuous random variables  $\mathbf{X}_1$  and  $\mathbf{Y}_1 = \mathbf{g}(\mathbf{X}_1)$ , where  $\mathbf{g}$  is a non-injective, surjective function, we can attempt to find the PDF  $f_{\mathbf{Y}_1}$  for  $\mathbf{Y}_1$  with the following steps:

1. Introduce new random variables  $\mathbf{X}_2$  and  $\mathbf{Y}_2$  as needed, in addition to  $\mathbf{X}_1$  and  $\mathbf{Y}_1$ . Create a new bijective function  $(\mathbf{Y}_1, \mathbf{Y}_2) = \mathbf{h}(\mathbf{X}_1, \mathbf{X}_2)$  which keeps the marginal distribution of  $\mathbf{Y}_1$  unchanged.

# From Non-Injectivity to Bijectivity

The method of transformations is only applicable in the case of bijective functions. There is no standard way of dealing with all non-bijective functions, but here we will discuss a helpful method for dealing with some non-injective, surjective functions.

## The Method of Transformations for Non-Injective, Surjective Functions

For continuous random variables  $\mathbf{X}_1$  and  $\mathbf{Y}_1 = \mathbf{g}(\mathbf{X}_1)$ , where  $\mathbf{g}$  is a non-injective, surjective function, we can attempt to find the PDF  $f_{\mathbf{Y}_1}$  for  $\mathbf{Y}_1$  with the following steps:

1. Introduce new random variables  $\mathbf{X}_2$  and  $\mathbf{Y}_2$  as needed, in addition to  $\mathbf{X}_1$  and  $\mathbf{Y}_1$ . Create a new bijective function  $(\mathbf{Y}_1, \mathbf{Y}_2) = \mathbf{h}(\mathbf{X}_1, \mathbf{X}_2)$  which keeps the marginal distribution of  $\mathbf{Y}_1$  unchanged.
2. Apply the method of transformations on the bijective problem.

# From Non-Injectivity to Bijectivity

The method of transformations is only applicable in the case of bijective functions. There is no standard way of dealing with all non-bijective functions, but here we will discuss a helpful method for dealing with some non-injective, surjective functions.

## The Method of Transformations for Non-Injective, Surjective Functions

For continuous random variables  $\mathbf{X}_1$  and  $\mathbf{Y}_1 = \mathbf{g}(\mathbf{X}_1)$ , where  $\mathbf{g}$  is a non-injective, surjective function, we can attempt to find the PDF  $f_{\mathbf{Y}_1}$  for  $\mathbf{Y}_1$  with the following steps:

1. Introduce new random variables  $\mathbf{X}_2$  and  $\mathbf{Y}_2$  as needed, in addition to  $\mathbf{X}_1$  and  $\mathbf{Y}_1$ . Create a new bijective function  $(\mathbf{Y}_1, \mathbf{Y}_2) = \mathbf{h}(\mathbf{X}_1, \mathbf{X}_2)$  which keeps the marginal distribution of  $\mathbf{Y}_1$  unchanged.
2. Apply the method of transformations on the bijective problem.
3. Find the marginal distribution of  $\mathbf{Y}_1$  by marginalising out  $\mathbf{Y}_2$ .

# Sums of Independent Continuous Random Variables

## Example 10

Let  $X_1$  and  $X_2$  be independent continuous random variables with supports  $\mathcal{X}_1$  and  $\mathcal{X}_2$  respectively. What is the distribution of  $Y_1 = X_1 + X_2$ ?

# Sums of Independent Continuous Random Variables

## Example 10

Let  $X_1$  and  $X_2$  be independent continuous random variables with supports  $\mathcal{X}_1$  and  $\mathcal{X}_2$  respectively. What is the distribution of  $Y_1 = X_1 + X_2$ ?

This function is non-injective. Let's make a bijective one to help us solve the problem. In addition to  $Y_1 = X_1 + X_2$ , let's introduce  $Y_2 = X_2$ . Given  $Y_1$  and  $Y_2$ , we can recover  $X_1$  and  $X_2$ , so this mapping is bijective.

Writing this function in the form  $\mathbf{h}(\mathbf{x}) = (x_1 + x_2, x_2)$  and letting  $\mathbf{y} = \mathbf{h}(\mathbf{x})$ , we see the Jacobian is

$$D\mathbf{h}(\mathbf{x}) = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

# Sums of Independent Continuous Random Variables

## Example 10

Now, applying the method of transformations:

$$\begin{aligned}f_{\mathbf{Y}}(\mathbf{y}) &= f_{\mathbf{X}}(\mathbf{x}) |\det(D\mathbf{h}(\mathbf{x}))|^{-1} \\&= f_{\mathbf{X}}(y_1 - y_2, y_2) \left| \det \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \right|^{-1} \\&= f_{\mathbf{X}}(y_1 - y_2, y_2).\end{aligned}$$

# Sums of Independent Continuous Random Variables

## Example 10

Now we just need to find the marginal distribution of  $Y_1$ . To do this, recognise that the support of  $Y_2$  is identical to  $X_2$ .

$$\begin{aligned}f_{Y_1}(y_1) &= \int_{y_2} f_{\mathbf{Y}}(\mathbf{y}) dy_2 \\&= \int_{y_2} f_{\mathbf{X}}(y_1 - y_2, y_2) dy_2 \\&= \int_{y_2} f_{X_1}(y_1 - y_2) f_{X_2}(y_2) dy_2 \\&= \int_{x_2} f_{X_1}(y_1 - x_2) f_{X_2}(x_2) dx_2.\end{aligned}$$

This convolution agrees with the result we got earlier in the lecture using the CDF method.

## The Sum of Gaussian Random Variables is Gaussian

We can prove that the sum of independent univariate Gaussian random variables is also Gaussian using convolutions, but this proof involves a particularly horrible mess of algebra. [Here's the proof](#) for anyone wanting to suffer.

## The Sum of Gaussian Random Variables is Gaussian

We can prove that the sum of independent univariate Gaussian random variables is also Gaussian using convolutions, but this proof involves a particularly horrible mess of algebra. [Here's the proof](#) for anyone wanting to suffer.

To summarise, if  $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$  and  $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ , then

$$X + Y \sim \mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2) .$$

# The Sum of Gaussian Random Variables is Gaussian

We can prove that the sum of independent univariate Gaussian random variables is also Gaussian using convolutions, but this proof involves a particularly horrible mess of algebra. [Here's the proof](#) for anyone wanting to suffer.

To summarise, if  $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$  and  $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ , then

$$X + Y \sim \mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2) .$$

This is also true for independent multivariate Gaussian random variables. If  $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}_X, \Sigma_X)$  and  $\mathbf{Y} \sim \mathcal{N}(\boldsymbol{\mu}_Y, \Sigma_Y)$ , then

$$\mathbf{X} + \mathbf{Y} \sim \mathcal{N}(\boldsymbol{\mu}_X + \boldsymbol{\mu}_Y, \Sigma_X + \Sigma_Y) .$$

# The Dirac Delta

Before continuing, we need to introduce the Dirac delta "function". It isn't really a regular function, but you can just consider it one for now. It is very loosely defined as

$$\delta(x) \simeq \begin{cases} \infty & \text{if } x = 0 \\ 0 & \text{otherwise} \end{cases},$$

and has the following important properties:

- Can be used as a PDF:

$$\int_{-\infty}^{\infty} \delta(x) dx = 1.$$

# The Dirac Delta

Before continuing, we need to introduce the Dirac delta "function". It isn't really a regular function, but you can just consider it one for now. It is very loosely defined as

$$\delta(x) \simeq \begin{cases} \infty & \text{if } x = 0 \\ 0 & \text{otherwise} \end{cases},$$

and has the following important properties:

- Can be used as a PDF:

$$\int_{-\infty}^{\infty} \delta(x) dx = 1.$$

- Is easy to integrate:

$$\int_{-\infty}^{\infty} g(x) \delta(x - a) dx = g(a).$$

# The Dirac Delta

---

The Dirac delta defines a deterministic distribution, where a single outcome occurs with probability 1. You can consider it as a normal distribution with mean 0 in the limit of the standard deviation approaching 0:

$$\delta(x) = \lim_{\sigma \rightarrow 0^+} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}.$$

# The Dirac Delta

The Dirac delta defines a deterministic distribution, where a single outcome occurs with probability 1. You can consider it as a normal distribution with mean 0 in the limit of the standard deviation approaching 0:

$$\delta(x) = \lim_{\sigma \rightarrow 0^+} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}.$$

One more useful relationship is as follows. If a function  $g$  has  $n$  real roots/zeros  $x_1, \dots, x_n$ , then

$$\delta(g(x)) = \sum_{i=1}^n \frac{\delta(x - x_i)}{|g'(x_i)|}.$$

We will now see an example of how to use a Dirac delta.

# Squaring a Gaussian

## Example 11

What is the PDF of  $Y_1 = g(X_1)$  given  $X_1 \sim \mathcal{N}(0, 1)$ ,  $g : \mathbb{R} \setminus \{0\} \rightarrow (0, \infty)$ , and  $g(x) = x^2$ ? Let's use the method of transformations.

# Squaring a Gaussian

## Example 11

What is the PDF of  $Y_1 = g(X_1)$  given  $X_1 \sim \mathcal{N}(0, 1)$ ,  $g : \mathbb{R} \setminus \{0\} \rightarrow (0, \infty)$ , and  $g(x) = x^2$ ? Let's use the method of transformations.

$g$  is surjective, but not injective. Let's address this by introducing a new function  $\mathbf{h} : \mathbb{R}^2 \setminus \{(0, 0)\} \rightarrow \mathbb{R}^2 \setminus \{(0, 0)\}$  defined by

$$\mathbf{h}(\mathbf{x}) = (x_1 x_2, x_2).$$

This is a bijective function, with  $\mathbf{y} = \mathbf{h}(\mathbf{x})$  having the inverse

$$\mathbf{h}^{-1}(\mathbf{y}) = \left( \frac{y_1}{y_2}, y_2 \right).$$

Our plan here is to force  $x_1$  and  $x_2$  to take the same value, thus resulting in  $y_1 = x_1^2$ .

# Squaring a Gaussian

## Example 11

Now that we've defined a bijective function, let's consider the random vector  $\mathbf{Y} = \mathbf{h}(\mathbf{X})$  with  $X_1 \sim \mathcal{N}(0, 1)$  and the distribution of  $X_2 | X_1$  being deterministic. Specifically, the conditional PDF for  $X_2 | X_1$  should be  $\delta(x_2 - x_1)$ . Now, applying the method of transformations:

$$\begin{aligned} f_{\mathbf{Y}}(\mathbf{y}) &= f_{\mathbf{X}}(\mathbf{h}^{-1}(\mathbf{y})) |\det(D\mathbf{h}^{-1}(\mathbf{y}))| \\ &= f_{\mathbf{X}}\left(\frac{y_1}{y_2}, y_2\right) \left| \det \begin{pmatrix} \frac{1}{y_2} & -\frac{y_1}{y_2^2} \\ 0 & 1 \end{pmatrix} \right| \\ &= f_{\mathbf{X}}\left(\frac{y_1}{y_2}, y_2\right) \left| \frac{1}{y_2} \right| \\ &= f_{X_1}\left(\frac{y_1}{y_2}\right) f_{X_2|X_1}\left(y_2 \mid \frac{y_1}{y_2}\right) \left| \frac{1}{y_2} \right| \\ &= f_{X_1}\left(\frac{y_1}{y_2}\right) \delta\left(y_2 - \frac{y_1}{y_2}\right) \left| \frac{1}{y_2} \right|. \end{aligned}$$

# Squaring a Gaussian

## Example 11

We have the joint distribution of  $Y_1$  and  $Y_2$ . Let's find the marginal of  $Y_1$ .

First we need to decompose the Dirac delta. The roots of  $y_2 - \frac{y_1}{y_2}$  are  $y_2 = \pm\sqrt{y_1}$ . The derivative of  $y_2 - \frac{y_1}{y_2}$  is  $1 + \frac{y_1}{y_2^2}$  which evaluates to 2 at the roots. Now we can continue:

$$\begin{aligned}f_{Y_1}(y_1) &= \int_{\mathbb{R}} f_{\mathbf{Y}}(\mathbf{y}) dy_2 \\&= \int_{\mathbb{R}} f_{X_1}\left(\frac{y_1}{y_2}\right) \delta\left(y_2 - \frac{y_1}{y_2}\right) \left|\frac{1}{y_2}\right| dy_2 \\&= \int_{\mathbb{R}} f_{X_1}\left(\frac{y_1}{y_2}\right) \cdot \frac{1}{2} (\delta(y_2 + \sqrt{y_1}) + \delta(y_2 - \sqrt{y_1})) \left|\frac{1}{y_2}\right| dy_2 \\&= \frac{1}{2} \int_{\mathbb{R}} f_{X_1}\left(\frac{y_1}{y_2}\right) \delta(y_2 + \sqrt{y_1}) \left|\frac{1}{y_2}\right| dy_2 \\&\quad + \frac{1}{2} \int_{\mathbb{R}} f_{X_1}\left(\frac{y_1}{y_2}\right) \delta(y_2 - \sqrt{y_1}) \left|\frac{1}{y_2}\right| dy_2\end{aligned}$$

# Squaring a Gaussian

## Example 11

$$\begin{aligned} &= \frac{1}{2} f_{X_1}(-\sqrt{y_1}) \left| \frac{-1}{\sqrt{y_1}} \right| + \frac{1}{2} f_{X_1}(\sqrt{y_1}) \left| \frac{1}{\sqrt{y_1}} \right| \\ &= \frac{1}{2\sqrt{y_1}} \left( \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(-\sqrt{y_1})^2}{2}\right) + \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(\sqrt{y_1})^2}{2}\right) \right) \\ &= \frac{1}{\sqrt{2\pi y_1}} \exp\left(-\frac{y_1}{2}\right). \end{aligned}$$

This agrees with the result we got using the CDF method earlier. By the way, this is a chi-squared distribution.

## Another Resource

---

For a more comprehensive overview of the tools available to deal with functions of continuous random variables, take a look at this [Wikibook](#).

# Dynamical Systems as Functions: A Fun Proof to Look at After the Course

---

# Tying Together the Modules in the Program

This is a bonus section. It covers concepts from linear algebra, calculus, probability, and differential equations, but I don't expect anyone to grasp it immediately even after covering these modules. Treat it as a challenge 😊.

Here we will use  $\mathbf{x}_t$  as shorthand for  $\mathbf{x}(t)$ .

Say we have some stochastic process  $\{\mathbf{X}_t \mid t \in [0, \infty)\}$ . We know the PDF of  $\mathbf{X}_0$  as well as the ODE that determines  $\mathbf{X}_t$  at all times  $t > 0$ . If we have  $\mathbf{X}_0 \sim f_0$  and ODE

$$\frac{d\mathbf{X}_t}{dt} = \mathbf{g}(\mathbf{X}_t),$$

then what is the PDF  $f_t$  of  $\mathbf{X}_t$ ?

## Prerequisite: Flow Functions

If we have an ODE, an initial state  $\mathbf{x}_s$ , and a length of time  $t$  to run the dynamics for, then  $\mathbf{x}_{s+t}$  is determined by a function of  $\mathbf{x}_s$  and  $t$ . If

$$\frac{d\mathbf{x}_t}{dt} = \mathbf{g}(\mathbf{x}_t),$$

then we define this function  $\varphi : \mathcal{X} \times \mathbb{R} \rightarrow \mathcal{X}$  as

$$\begin{aligned}\varphi(\mathbf{x}_s, t) &= \mathbf{x}_s + \int_s^{s+t} \mathbf{g}(\mathbf{x}_\tau) d\tau \\ &= \mathbf{x}_{s+t}\end{aligned}$$

This function is called a flow, and is often written as  $\varphi^t(\mathbf{x}_s)$ . Take note that  $\varphi^s \circ \varphi^t = \varphi^{s+t}$ .

# How the PDF Warps Over Time

Back to our original question: what is the PDF  $f_t$  of  $\mathbf{X}_t$ ? It is given by

$$f_t(\mathbf{x}_t) = f_0(\varphi^{-t}(\mathbf{x}_t)) \exp\left(\int_t^0 \nabla \cdot \mathbf{g}(\mathbf{x}_\tau) d\tau\right).$$

*Proof:*

We need to find the change in volume over the trajectory taken by the dynamics. Here  $D\varphi^t$  is the Jacobian of the flow  $\varphi$  of the dynamics at time  $t$ . Let's start with **Jacobi's formula**:

$$\frac{d}{dt} \det(D\varphi^t(\mathbf{x}_0)) = \det(D\varphi^t(\mathbf{x}_0)) \operatorname{tr}\left(D\varphi^t(\mathbf{x}_0)^{-1} \frac{d}{dt} D\varphi^t(\mathbf{x}_0)\right).$$

# How the PDF Warps Over Time

Therefore, considering the derivative of logarithms:

$$\begin{aligned}\frac{d}{dt} \log (\det (D\varphi^t(\mathbf{x}_0))) &= \text{tr} \left( D\varphi^t(\mathbf{x}_0)^{-1} \frac{d}{dt} D\varphi^t(\mathbf{x}_0) \right) \\ &= \text{tr} \left( D\varphi^{-t}(\varphi^t(\mathbf{x}_0)) D \frac{d\varphi^t}{dt}(\mathbf{x}_0) \right) \\ &= \text{tr} (D\varphi^{-t}(\varphi^t(\mathbf{x}_0)) D(g \circ \varphi^t)(\mathbf{x}_0)) \\ &= \text{tr} (D(g \circ \varphi^t \circ \varphi^{-t})(\varphi^t(\mathbf{x}_0))) \\ &= \text{tr} (Dg(\mathbf{x}_t)) \\ &= \nabla \cdot g(\mathbf{x}_t) .\end{aligned}$$

# How the PDF Warps Over Time

Hence, the log Jacobian determinant of the flow over time  $t$  is given by

$$\begin{aligned}\log (\det (D\varphi^t(\mathbf{x}_0))) &= \log (\det (D\varphi^0(\mathbf{x}_0))) + \int_0^t \nabla \cdot \mathbf{g}(\mathbf{x}_\tau) d\tau \\ &= \log (\det (I)) + \int_0^t \nabla \cdot \mathbf{g}(\mathbf{x}_\tau) d\tau \\ &= \int_0^t \nabla \cdot \mathbf{g}(\mathbf{x}_\tau) d\tau.\end{aligned}$$

Starting with  $\mathbf{x}_t$  and integrating backwards in time, we get

$$\begin{aligned}\log (\det (D\varphi^t(\varphi^{-t}(\mathbf{x}_t)))) &= - \int_t^0 \nabla \cdot \mathbf{g}(\mathbf{x}_\tau) d\tau \\ \det (D\varphi^t(\varphi^{-t}(\mathbf{x}_t))) &= \exp \left( - \int_t^0 \nabla \cdot \mathbf{g}(\mathbf{x}_\tau) d\tau \right).\end{aligned}$$

# How the PDF Warps Over Time

Finally, using the method of transformations, we have

$$\begin{aligned} f_t(\mathbf{x}_t) &= f_0(\varphi^{-t}(\mathbf{x}_t)) |\det(D\varphi^t(\varphi^{-t}(\mathbf{x}_t)))|^{-1} \\ &= f_0(\varphi^{-t}(\mathbf{x}_t)) \exp\left(\int_t^0 \nabla \cdot \mathbf{g}(\mathbf{x}_\tau) d\tau\right). \end{aligned}$$

□

If we can sample from  $f_0$  and approximately simulate the given dynamics, we can get approximate samples from  $f_t$  for any length of time  $t$ . Of equal importance, we can also approximately calculate  $f_t(\mathbf{x}_t)$  using the formula above. Notice how  $f_t$  starts at  $f_0$  and smoothly warps over time.

The End