Probability Theory

SUZUKI, Atsushi

The whole contents

1. Continuous Random Variables

Outline

1 Continuous Random Variables

- Introduction: why are continuous random variables less trivial?
- Probability density function
- Area, integration, and properties of PDF.
- Calculating integral
- Summary statistics of continuous RV and integral
- Jointly continuous random variables and multiple integral
- Relation among jointly continuous RVs
- Exercises

Outline

1 **Continuous Random Variables**

- Introduction: why are continuous random variables less trivial?

Continuous random variables in real AI applications

A discrete RV can take only limited values. However, many real-world phenomena are represented as random variables which can take any real value in a continuous section.

- Inflation rate (economics),
- Position of a vehicle,
- The brightness of scenery,
- The intensity of an acoustic signal,
- Density of air pollution.

Hence, when we want to analyze those phenomena using probability theory, we cannot always use mathematical tools to handle discrete RVs.

For example, those random variables typically have **no probability mass function (PMF)**.

A random variable may not have a PMF.

Consider a simple random variable uniformly distributed in $[0,1]$. Here $\Pr(0 \leq X \leq 1) = 1$.

This random variable have nowhere probability mass, i.e., $\Pr(X = x) = 0$. for any $X \in \mathbb{R}$.

Proof.

Since its support is $[0,1]$, it is trivial that $\Pr(X = x) = 0$ for $x \neq [0,1]$. For $x \in [0,1]$, assume, for the sake of contradiction, that $\Pr(X = x) = \epsilon$, where $\epsilon > 0$. From its uniformity, if $\Pr(X = x) = \epsilon$ holds for one value $x \in [0,1]$, then it holds for all $x \in [0,1]$. Hence, if $A \subset [0,1]$ and A has at least N elements, $\Pr(X \in A) \leq N\epsilon$. However, there are an infinite number of real numbers in $[0,1]$, so $\Pr(X \in [0,1])$ is infinity. It contradicts $\Pr(X \in [0,1]) = 1$. □

A random variable may not have a PMF.

Consider a simple random variable uniformly distributed in $[0,1]$. Here $\Pr(0 \le X \le 1) = 1$.

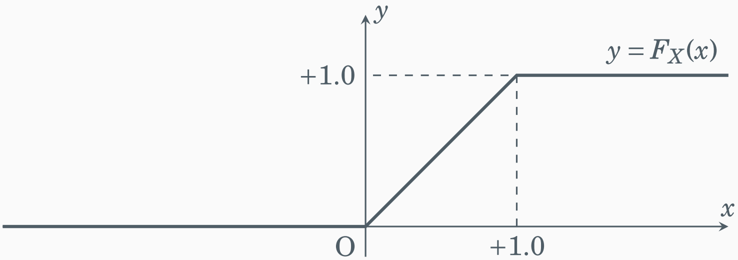
This random variable have nowhere probability mass, i.e., $\Pr(X = x) = 0$. for any $X \in \mathbb{R}$.

Other random variables whose support is a section in the real line have the same problem. Hence, we need another way to represent a random variable.

Fortunately, any univariate random variable has a cumulative distribution function (CDF)

Example of CDF for a non-discrete RV

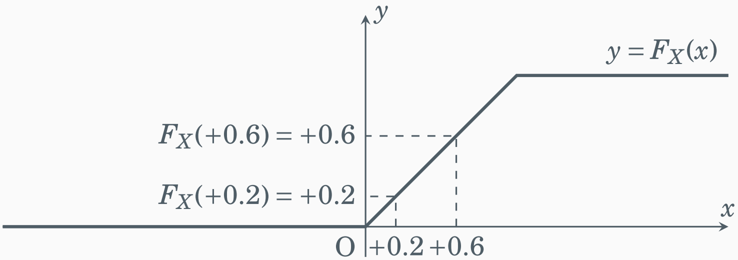
The CDF of a random variable X uniformly distributed in $[0, 1]$ is:



$$F_X(x) = \begin{cases} 0 & \text{if } x \leq 0, \\ x & \text{if } 0 \leq x \leq 1, \\ 1 & \text{if } x \geq 1. \end{cases} \tag{1}$$

Example of CDF for a non-discrete RV

The CDF of a random variable X uniformly distributed in $[0,1]$ is:

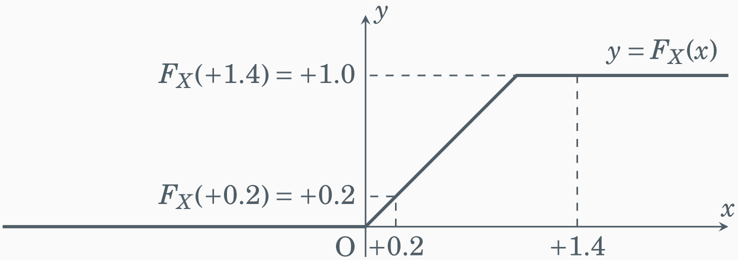


Using the CDF, we can calculate the probability of various events. For example,

$$
\begin{aligned}
\Pr(0.2 < X \leq 0.6) &= \Pr(X \leq 0.6) - \Pr(X \leq 0.2) \\
&= F_X(0.6) - F_X(0.2) \\
&= 0.6 - 0.2 = 0.4.
\end{aligned}
\tag{1}
$$

Example of CDF for a non-discrete RV

The CDF of a random variable X uniformly distributed in $[0,1]$ is:

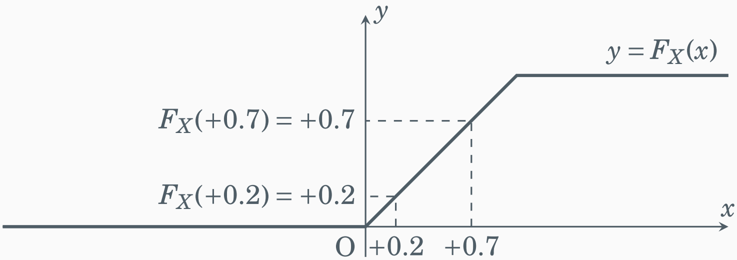


Using the CDF, we can calculate the probability of various events. For example,

$$\begin{aligned}\Pr(0.2 < X \leq 1.4) &= \Pr(X \leq 1.4) - \Pr(X \leq 0.2) \\ &= F_X(1.4) - F_X(0.2) \\ &= 1.0 - 0.2 = 0.8.\end{aligned} \tag{1}$$

Example of CDF for a non-discrete RV

The CDF of a random variable X uniformly distributed in $[0,1]$ is:



Using the CDF, we can calculate the probability of various events. For example,

$$\begin{aligned}
\Pr(0.2 \le X \le 0.7) &= \Pr(X \le 0.7) - \lim_{x \nearrow 0.2} \Pr(x) \\
&= F_X(0.7) - \lim_{x \nearrow 0.2} F_X(x) \\
&= 0.7 - 0.2 = 0.5.
\end{aligned} \tag{1}$$

Why are we not satisfied with the CDF?

However, the CDF is not always welcomed. It is because

- The CDF is not intuitive. At one glance, we do not know around which value the random variable tends to take a value.
- The CDF can be extremely complex even for a practically important distribution.

Although there exists no PMF for a continuous RV in general, we want to indicate which values the RV tends to take frequently as the PMF does for a discrete RV.

The ***probability density function (PDF)*** achieves this objective.

Learning outcomes

By the end of this section, you should be able to:

- Explain what a probability density function represents,
- Explain the relation between the probability density function and cumulative distribution function,
- Calculate the probability of an event using the integral and the probability density function, and
- Calculate summary statistics of continuous random variables.

Notation: sections

In the following, $\mathbb{R}$, $\mathbb{R}_{\geq 0}$, and $\mathbb{R}_{>0}$ are the sets of real numbers, nonnegative real numbers, and positive real numbers, respectively.

Let a and b be real values. By $[a,b]$, (a,b), we denote the closed and open sections defined by

- $[a,b] = \{x \in \mathbb{R} | a \leq x \leq b\}$,
- $(a,b) = \{x \in \mathbb{R} | a < x < b\}$,

respectively. Likewise, by $(a,b]$ and $[a,b)$, we denote the semi-open sets defined by

- $(a,b] = \{x \in \mathbb{R} | a < x \leq b\}$,
- $[a,b) = \{x \in \mathbb{R} | a \leq x < b\}$,

Notation: Napier's constant and the exponential function

The real number constant e, called ***Napier's constant*** or ***Euler's number***, is defined by $\mathrm{e} := \lim_{n \to +\infty} \left(1 + \frac{1}{n}\right)^n$. Note that $\mathrm{e} = 2.718281828...$ and is the only real value that satisfies $\frac{\mathrm{d}}{\mathrm{d}x}\mathrm{e}^x = \mathrm{e}^x$.

We define the ***(natural) exponential function*** $\exp : \mathbb{R} \to \mathbb{R}_{>0}$ by $\exp(x) = \mathrm{e}^x$.

Outline

1. **Continuous Random Variables**
   - Probability density function

Idea of the probability density function

As we have seen in the case of the uniform distribution in the section $[0, 1]$, the probability $\Pr(X = c)$ might be zero for a real value c in many cases. In this case, we cannot say which values the RV tend to take more frequently than others.

Idea of the probability density function

As we have seen in the case of the uniform distribution in the section $[0,1]$, the probability $\Pr(X = c)$ might be zero for a real value c in many cases. In this case, we cannot say which values the RV tend to take more frequently than others.

Hence, we evaluate the probability of the RV taking a value **in a section**. For example, instead of evaluating $\Pr(X = c)$, we evaluate the probability $\Pr(a < X \leq b)$ for real values a, b around c such that $a < b$. If the probability is high and the section length $b - a$ is short, we can say that the RV X takes a value around c frequently.

Idea of the probability density function

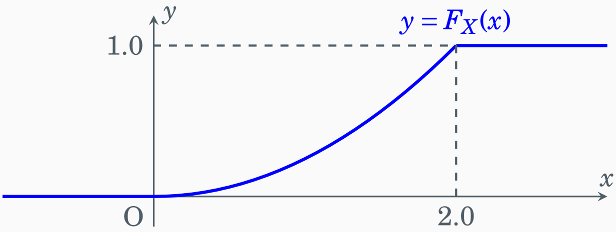
As we have seen in the case of the uniform distribution in the section $[0,1]$, the probability $\Pr(X=c)$ might be zero for a real value c in many cases. In this case, we cannot say which values the RV tend to take more frequently than others.

Hence, we evaluate the probability of the RV taking a value **in a section**. For example, instead of evaluating $\Pr(X=c)$, we evaluate the probability $\Pr(a<X\leq b)$ for real values a,b around c such that $a<b$. If the probability is high and the section length $b-a$ is short, we can say that the RV X takes a value around c frequently.

So, we can regard the probability per the section length as the ***density*** of the probability distribution of the RV X around the section. A high density around a value c indicates that the X tends to take a value around c.

Based on the above idea, we can formulate the ***probability density function (PDF)*** from the cumulative distribution function (CDF) as follows.

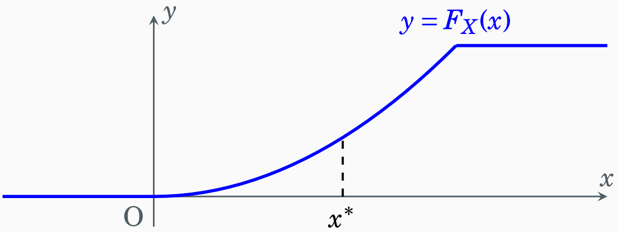
CDF to the density.



The CDF of a RV X.

Figure:

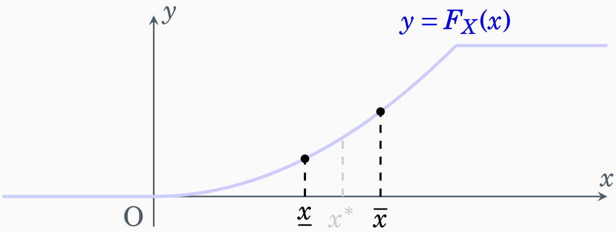
CDF to the density.



Suppose we want to know how frequently the RV X takes a value “around” x^*.

Figure:

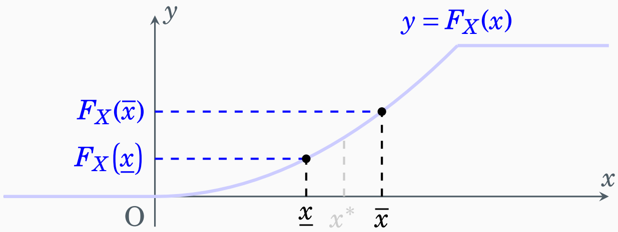
CDF to the density.



We consider an interval $[\underline{x}, \overline{x}]$ including x^*.

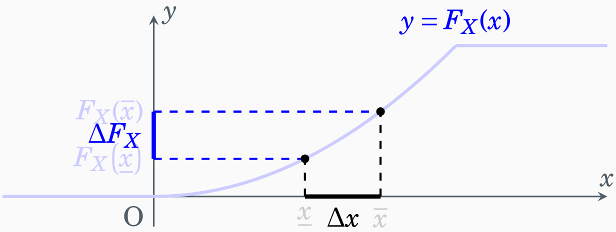
Figure:

CDF to the density.



We find the probability $\Pr(X \in [\underline{x}, \overline{x}])$, given by $F_X(\overline{x}) - F_X(\underline{x})$.

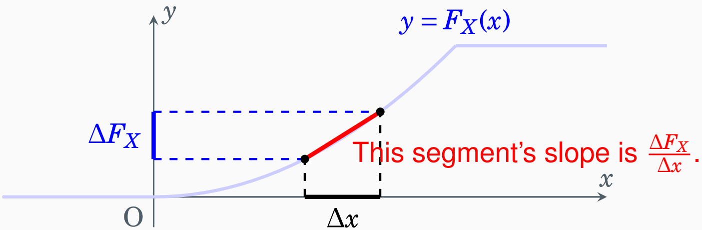
Figure:



Define $\Delta x := \overline{x} - \underline{x}$ and $\Delta F_X := F_X(\overline{x}) - F_X(\underline{x}) = \Pr(X \in [\underline{x}, \overline{x}])$.

Figure:

CDF to the density.



Define $\Delta x := \overline{x} - \underline{x}$ and $\Delta F_X := F_X(\overline{x}) - F_X(\underline{x}) = \Pr(X \in [\underline{x}, \overline{x}])$.

The RV X tends to take a value around x^*
if the probability per length $\frac{\Delta F_X}{\Delta x}$, or the "density" is large.

Figure:

CDF to the density.

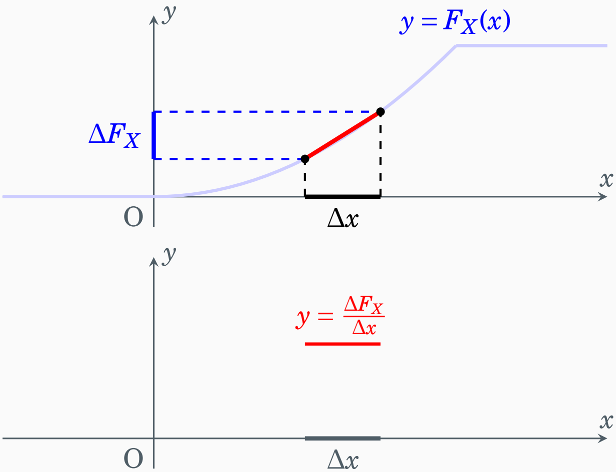


Figure: The graph plot of the probability per length $\frac{\Delta F_X}{\Delta x}$, or the "density"

CDF to the density.

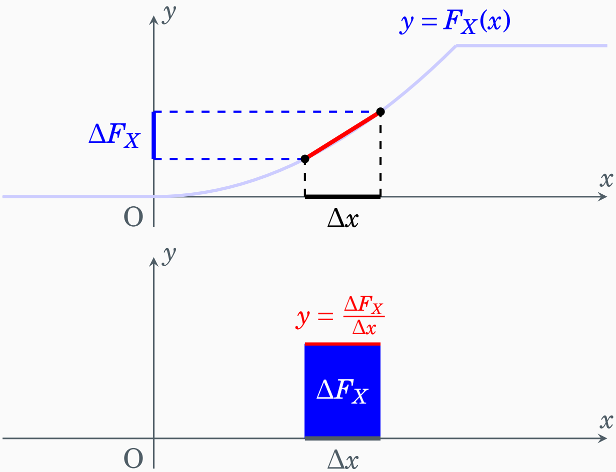


Figure: The area under the graph of $\frac{\Delta F_X}{\Delta x}$ is given by $\Delta x \cdot \frac{\Delta F_X}{\Delta x} = \Delta F_X$.

CDF to the density.

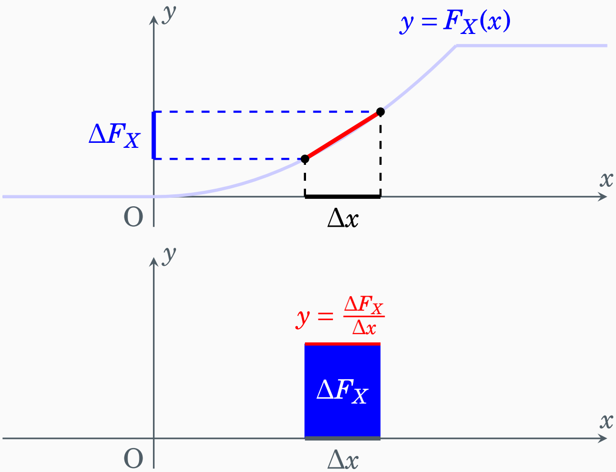


Figure: Let's see how the density varies in other sections.

CDF to the density.

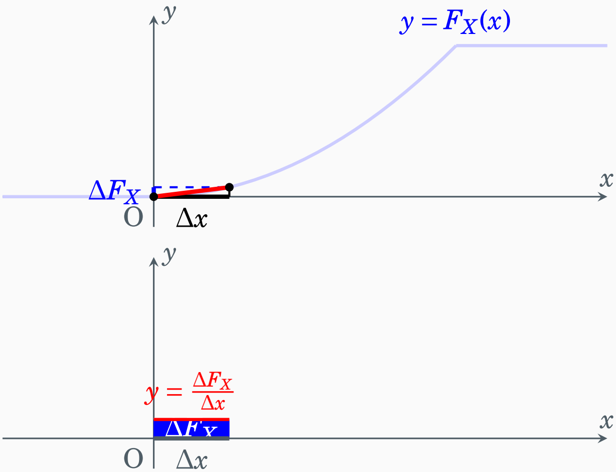


Figure: The density at the section $[0.00, 0.50]$ is small, reflecting the gentle slope of the graph of

CDF to the density.

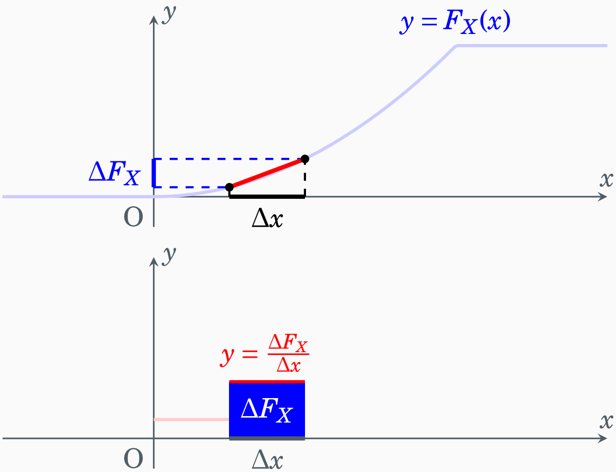


Figure: The density at the section $[0.50, 1.00]$ is larger, reflecting the steeper slope of the graph of

CDF to the density.

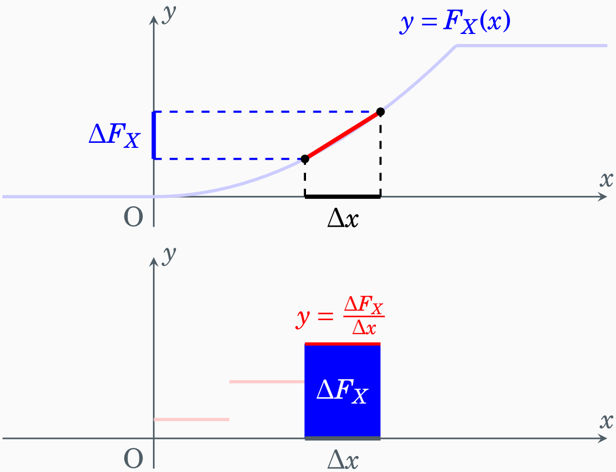


Figure: The density at the section $[1.00, 1.50]$ is even larger, reflecting the steeper slope of the

CDF to the density.

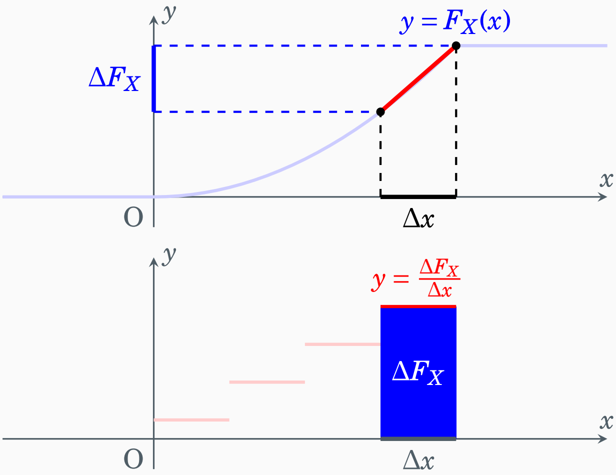


Figure: The density at the section $[1.50, 2.00]$ is even larger, reflecting the steeper slope of the

CDF to the density.

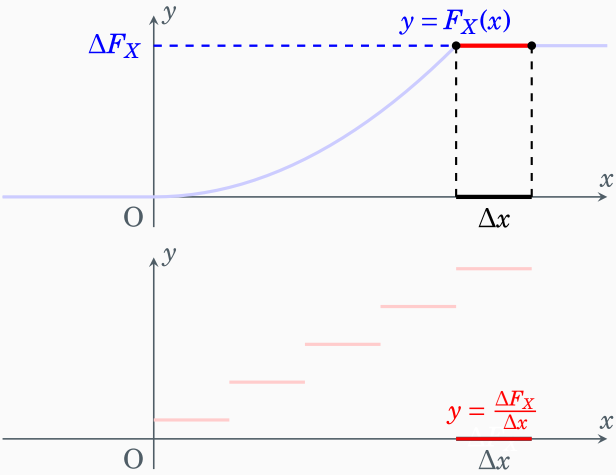


Figure: The density at the section $[2.00, 2.50]$ is zero since the graph of F_X is horizontal.

CDF to the density.

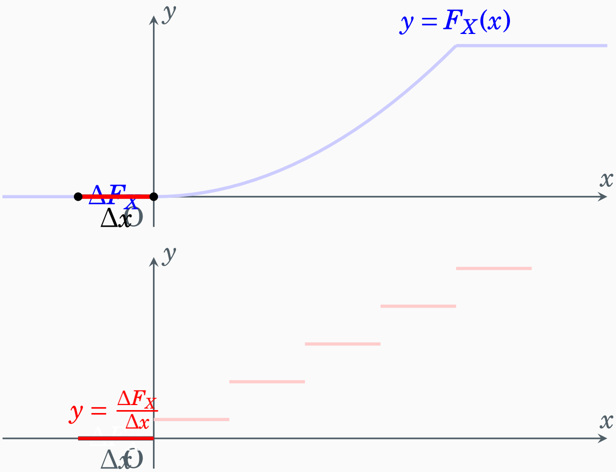


Figure: The density at the section $[-0.50, 0.90]$ is zero since the graph of F_X is horizontal.

CDF to the density.

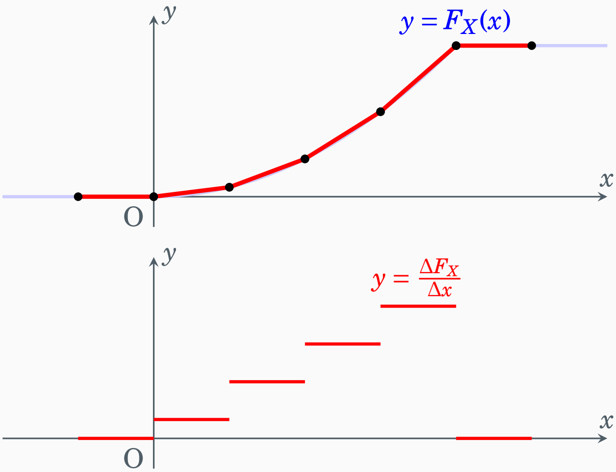


Figure: We have the graph of the piece-wise density of the distribution at each section.

CDF to the density.

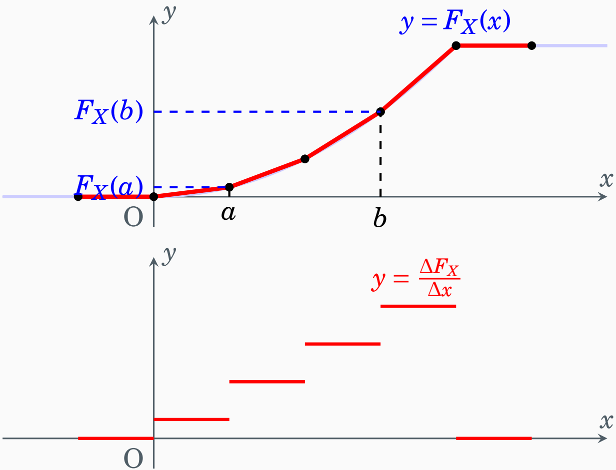


Figure: Using the piece-wise density, we can calculate $\Pr(a < X \leq b)$ if each of a and b is an end

CDF to the density.

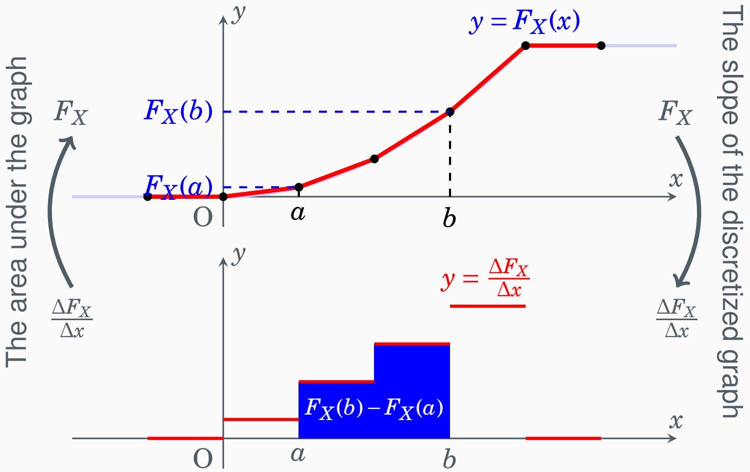


Figure: Actually, the probability $\Pr(a < X \le b) = F_X(b) - F_X(a)$ is the area under the piece-wise

CDF to the density.

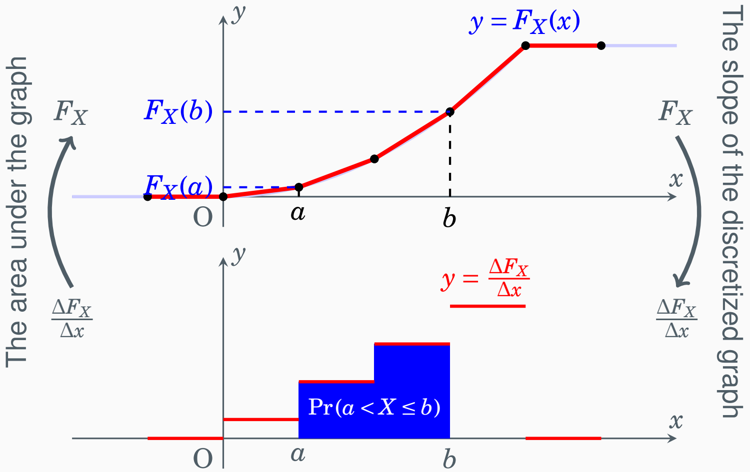


Figure: Actually, the probability $\Pr(a < X \leq b) = F_X(b) - F_X(a)$ is the area under the piece-wise

CDF to the density.

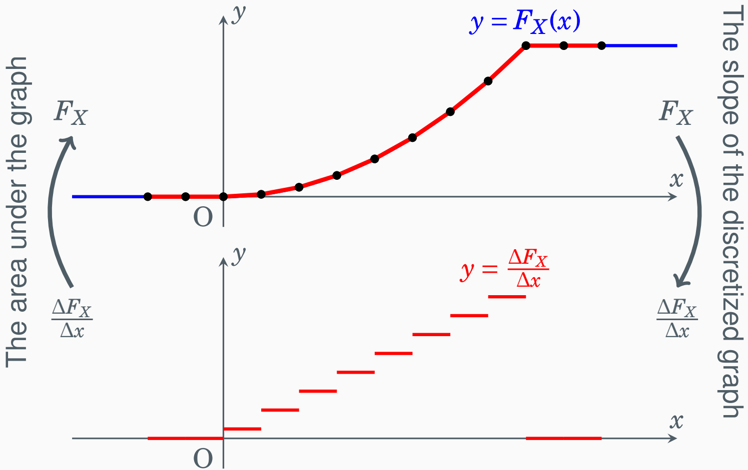


Figure: If the sections are shorter, we can evaluate $\Pr(a < X \leq b)$ for more pairs of a and b.

CDF to the density.

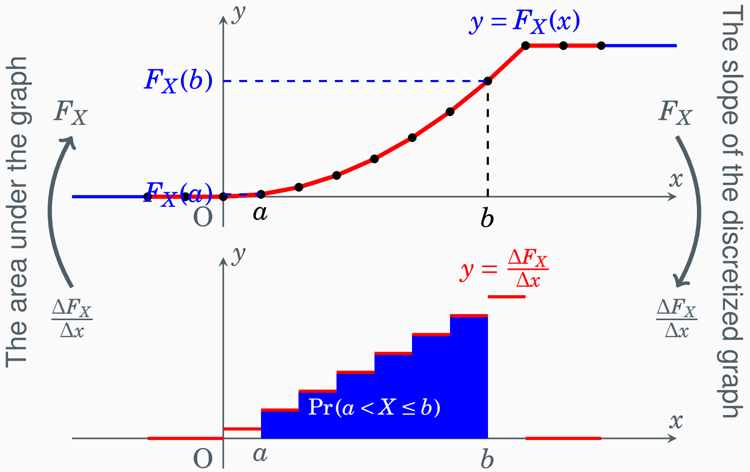


Figure: If the sections are shorter, we can evaluate $\Pr(a < X \leq b)$ for more pairs of a and b.

CDF to the density.

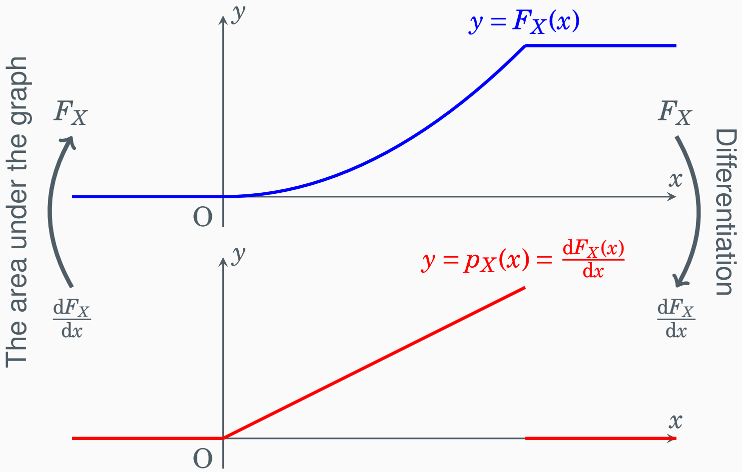


Figure: If we consider infinitely short sections, the density $\frac{\Delta F_X}{\Delta x}$ converges to the derivative $\frac{\mathrm{d}F_X}{\mathrm{d}x}$.

CDF to the density.

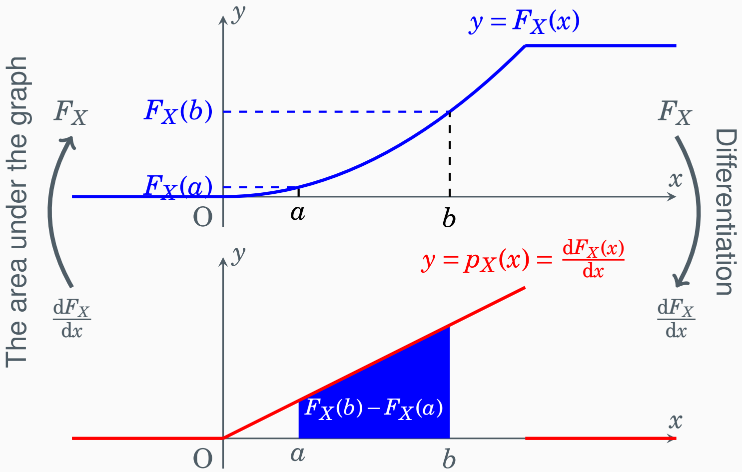


Figure: Actually, the probability $\Pr(a < X \leq b) = F_X(b) - F_X(a)$ is the area under the graph of $\frac{\mathrm{d}F_X}{\mathrm{d}x}$.

CDF to the density.

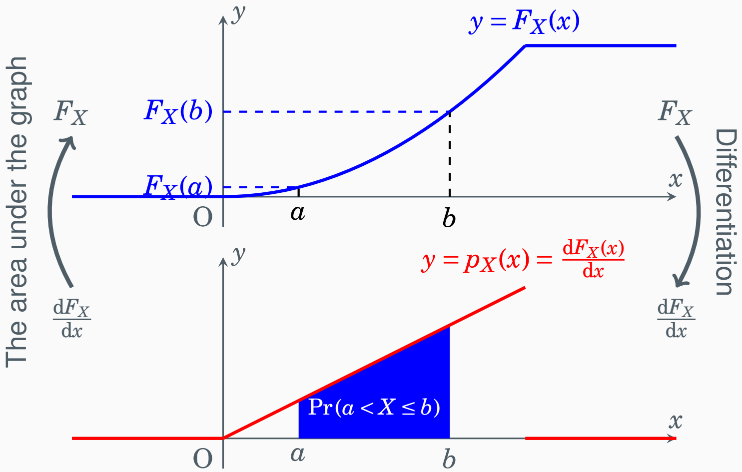


Figure: Actually, the probability $\Pr(a < X \leq b) = F_X(b) - F_X(a)$ is the area under the graph of $\frac{\mathrm{d}F_X}{\mathrm{d}x}$.

Probability density function (PDF)

Definition (Probability density function and continuous random variable)

Let X be a RV. A function $p_X : \mathbb{R} \to \mathbb{R}_{\geq 0}$ is called a ***probability density function (PDF)*** of X if the probability $\Pr(a < X \leq b)$ equals to the area bounded by the graph of $y = p_X(x)$ and $y = 0$ between $x = a$ and $x = b$ for all a and b such that $a \leq b$.
If a RV has at least one PDF, the RV is called a ***continuous random variable***.

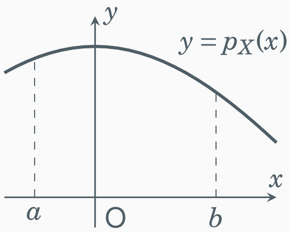


Figure: If p_X is a PDF of X, the probability $\Pr(a < X \leq b)$ is given by the area under the PDF in the domain $(a, b]$.

Probability density function (PDF)

Definition (Probability density function and continuous random variable)

Let X be a RV. A function $p_X : \mathbb{R} \to \mathbb{R}_{\geq 0}$ is called a ***probability density function (PDF)*** of X if the probability $\Pr(a < X \leq b)$ equals to the area bounded by the graph of $y = p_X(x)$ and $y = 0$ between $x = a$ and $x = b$ for all a and b such that $a \leq b$.
If a RV has at least one PDF, the RV is called a ***continuous random variable***.

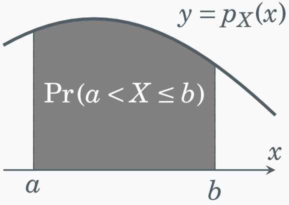


Figure: If p_X is a PDF of X, the probability $\Pr(a < X \leq b)$ is given by the area under the PDF in the domain $(a, b]$.

A continuous RV has nowhere a “mass.”

The area under a curve in a zero-length section is zero. Hence, if a RV is continuous, it has no probability mass anywhere. That is,

Theorem

If X is a continuous RV, the probability $\Pr(X = c)$ is zero for any $c \in \mathbb{R}$.

Hence, when we discuss a continuous RV, we do not need to discuss whether or not a section includes the endpoints. That is,

Corollary

Let X be a continuous RV and a and b be real values such that $a < b$. Then we have,

$$\Pr(a \leq X \leq b) = \Pr(a < X \leq b) = \Pr(a \leq X < b) = \Pr(a < X < b). \tag{2}$$

Hence, we can replace $a < X \leq b$ with $a \leq X \leq b$ or another in the definition of the PDF[1].

Note: the end-points are not ignorable for a discrete RV.

A discrete RV has a probability mass on any value in its support. Hence, for example, $\Pr(a \leq X \leq b) \neq \Pr(a < X \leq b)$ in general.

For example, if X is the value when we roll an ideal six-sided dice, $\Pr(3 \leq X \leq 6) = \frac{4}{6} \neq \Pr(3 < X \leq 6) = \frac{3}{6}$.

CDF and PDF

Assume that the CDF is differentiable at all the points on the real number line expect for finite points. As we can see in the construction of the PDF from the CDF, we can get the PDF by differentiating the CDF.

In practice, we usually know the PDF in advance but the CDF is unknown. Hence, we need to understand how to evaluate the area bounded by the graph of a general PDF.

Outline

1 **Continuous Random Variables**

Area, integration, and properties of PDF.

How to mathematically calculate the area under the curve?

Let X be a continuous RV and p_X be its PDF. Recall that the probability $\Pr(a \leq X \leq b)$ is given by the area under the graph of PDF p_X in the section $[a, b]$.

Hence, we need a mathematical tool to evaluate the area under the curve of a function in general.

Integration is the area to discuss the area under the graph of a function, (or the volume under the graph of a function in higher-dimensional space). We will learn it in the following.

Definite Integral

Suppose that $a \leq b$.

The (signed) area bounded by the graph of $y = f(x)$ and $y = 0$ between $x = a$ and $x = b$ is called the ***definite integral*** of f between a and b, which is denoted by $\displaystyle\int_a^b f(x)\,\mathrm{d}x$.

We also define $\displaystyle\int_b^a f(x)\,\mathrm{d}x := -\int_a^b f(x)\,\mathrm{d}x$.

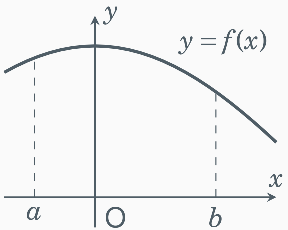


Figure: The definite integral is the area bounded by the graph of the function.

Definite Integral

Suppose that $a \le b$.

The (signed) area bounded by the graph of $y = f(x)$ and $y = 0$ between $x = a$ and $x = b$ is called the ***definite integral*** of f between a and b, which is denoted by $\displaystyle\int_a^b f(x)\,\mathrm{d}x$.

We also define $\displaystyle\int_b^a f(x)\,\mathrm{d}x := -\int_a^b f(x)\,\mathrm{d}x$.

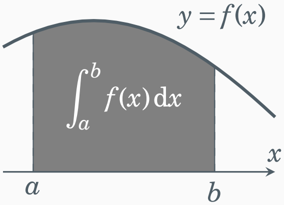


Figure: The definite integral is the area bounded by the graph of the function.

Definite Integral: When the function takes negative values

Areas bounded by the graph taking negative values are counted as negative values.

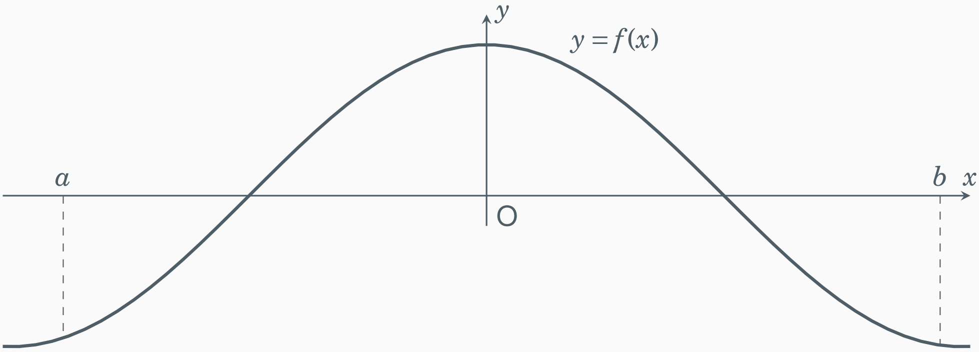


Figure: Areas bounded by the graph taking negative values are counted as negative values.

Definite Integral: When the function takes negative values

Areas bounded by the graph taking negative values are counted as negative values.

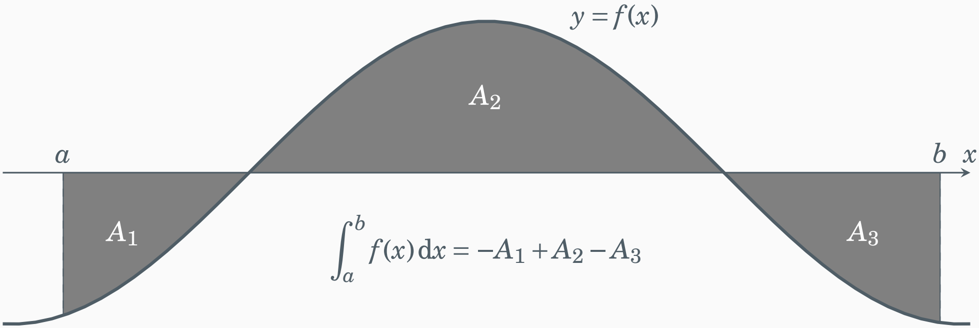


Figure: Areas bounded by the graph taking negative values are counted as negative values.

Any continuous RV has a nonnegative PDF

Assume that X is a continuous RV let p_X be a PDF of X. The probability $\Pr(a < X \le b) = \int_a^b p_X(x)\,\mathrm{d}x$ is always nonnegative, so we expect the PDF p_X to be a nonnegative function.

Strictly speaking, a PDF of a RV is not unique, since the area bounded by the graph does not change even if we change the value of the function at finite or countable points[2].

Nevertheless, if a RV has a PDF, we can assume that it is a nonnegative function without loss of generality.

Theorem

Let X be a continuous RV, i.e., there is a PDF of X. Then, there exists a ***nonegative*** *PDF of X, i.e, a PDF p_X such that $p_X(x) \ge 0$ at any $x \in \mathbb{R}$.*

In the following, we always assume that a PDF is nonnegative.

[2] Strictly speaking, at points in a set with measurement zero.

Probability of a RV being in a complicated shape

If we want to calculate the probability $\Pr(X \in A)$, where $A \subset \mathbb{R}$ has a complicated shape, we can calculate it using the sum rule.

Specifically, suppose that we have a decomposition $A = \bigcup_{i=1}^{n} (a_i, b_i]$, where $(a_i, b_i] \cap (a_j, b_j] = \emptyset$. Then, we have that

$$\Pr(X \in A) = \sum_{i=1}^{n} \Pr(a_i < X \leq b_i). \tag{3}$$

If X is a continuous RV and p_X is its PDF, the above value equals $\sum_{i=1}^{n} \int_{a_i}^{b_i} p_X(x) \mathrm{d}x$.

The same discussion holds even if the decomposition includes open sections like (a_i, b_i) or closed sections like $[a_j, b_j]$.

Note that the above calculation is not always correct if the decomposition includes an uncountably infinite number of sections.

Probability of a RV being in an infinite length section

If we need to evaluate the probability $\Pr(a < X)$, what we do is consider $\Pr(a < X \leq b)$ for an infinitely large b. Hence, we have that $\Pr(a < X) = \lim_{b \to +\infty} \Pr(a < X \leq b)$. The reverse holds for $\Pr(X \leq b)$. In other words, we can evaluate those probabilities by taking the limit of a definite integral as follows.

Theorem

Let X be a continuous RV, whose PDF is p_X, and a and b be real values. Then,

- $\Pr(a < X) = \Pr(a \leq X) = \lim_{b \to +\infty} \int_a^b p_X(x) \mathrm{d}x,$
- $\Pr(X < b) = \Pr(X \leq b) = \lim_{a \to -\infty} \int_a^b p_X(x) \mathrm{d}x.$

The “sum” of the PDF is one.

The section $(a, 0]$ includes all the nonpositive numbers if a is infinitely small and the section $(0, b]$ includes all the positive numbers b is infinitely large. Since a continuous RV X always takes a real value, the sum of the probabilities $\Pr(a < X \leq 0) + \Pr(0 < X \leq b)$ is 1 if a is infinitely small and b is infinitely large. Hence, the following always hold.

Theorem

Let X be a continuous RV whose PDF is p_X. We have that

$$\lim_{a \to -\infty} \int_a^0 p_X(x)\,\mathrm{d}x + \lim_{b \to +\infty} \int_0^b p_X(x)\,\mathrm{d}x = 1 \tag{4}$$

The above property is similar to a property of the probability mass function (PMF) of a discrete RV. To see that, we will introduce the ***improper integral***.

Improper integral

As we have seen, we often want to calculate limits of the definite integral. We call them ***improper integrals***, and use special notations as follows.

Definition (Improper integrals)

Let $f : \mathbb{R} \to \mathbb{R}$ be a function and a and b be real values. We define the value $\displaystyle\int_a^{+\infty} f(x)\,\mathrm{d}x$, $\displaystyle\int_{-\infty}^{b} f(x)\,\mathrm{d}x$, and $\displaystyle\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x$ by the following.

- $\displaystyle\int_a^{+\infty} f(x)\,\mathrm{d}x := \lim_{b\to+\infty}\int_a^b f(x)\,\mathrm{d}x,$
- $\displaystyle\int_{-\infty}^{b} f(x)\,\mathrm{d}x := \lim_{a\to-\infty}\int_a^b f(x)\,\mathrm{d}x,$
- $\displaystyle\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x := \int_{-\infty}^{0} f(x)\,\mathrm{d}x + \int_0^{+\infty} f(x)\,\mathrm{d}x.$

Interpretation of improper integrals

We can regard improper integrals $\int_a^{+\infty} f(x)\,\mathrm{d}x$, $\int_{-\infty}^{b} f(x)\,\mathrm{d}x$, and $\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x$ as the signed areas bounded by a graph of f in the section $(a, +\infty)$, $(-\infty, b)$, and $(-\infty, +\infty)$, respectively.

Rewriting properties of the PDF using improper integrals

We can rewrite the properties of the PDF in previous slides as follows.

Theorem

Let X be a continuous RV, whose PDF is p_X, and a and b be real values. Then,

- $\Pr(a < X) = \Pr(a \leq X) = \int_{a}^{+\infty} p_X(x)\, \mathrm{d}x,$
- $\Pr(X < b) = \Pr(X \leq b) = \int_{-\infty}^{b} p_X(x)\, \mathrm{d}x,$
- $\int_{-\infty}^{+\infty} p_X(x)\, \mathrm{d}x = 1.$

The third property is similar to a property of the PMF: $\sum_{x \in \mathscr{X}} P_X(x) = 1$, where X is a discrete RV, $\mathscr{X}$ is its support and P_X is its PMF.

Properties of the definite integral

Let a, b, c be real numbers and f and g be functions of a real value.

- $\displaystyle\int_b^a f(x)\,\mathrm{d}x := -\int_a^b f(x)\,\mathrm{d}x \qquad \text{(by definition)},$
- $\displaystyle\int_a^a f(x)\,\mathrm{d}x = 0 \qquad \text{(The area is zero in a zero length section.)}.$
- $\displaystyle\int_a^c f(x)\,\mathrm{d}x + \int_c^b f(x)\,\mathrm{d}x = \int_a^b f(x)\,\mathrm{d}x \qquad \text{(horizontal concatenation)}.$

Outline

1. **Continuous Random Variables**

   Calculating integral

Calculating definite integrals

Let X be a continuous RV and p_X be its PDF. Since the probability $\Pr(a < X \leq b)$ is given by the definite integral $\int_a^b p(x)\,\mathrm{d}x$, we need to know **how to calculate definite integrals** to understand the behavior of the continuous RV X.

[3] e.g., the trapezoidal rule, the Gauss-Legendre quadrature rule, the double exponential formula

Calculating definite integrals

Let X be a continuous RV and p_X be its PDF. Since the probability $\Pr(a < X \leq b)$ is given by the definite integral $\int_a^b p(x)\,\mathrm{d}x$, we need to know **how to calculate definite integrals** to understand the behavior of the continuous RV X.

There are two directions to calculate definite integrals.

- **Numerical integration** by approximating the area by shapes of which we can calculate the area easier.
- **Analytical integration** by conducting integration as the inverse operation of differentiation.

[3] e.g., the trapezoidal rule, the Gauss-Legendre quadrature rule, the double exponential formula

Calculating definite integrals

Let X be a continuous RV and p_X be its PDF. Since the probability $\Pr(a < X \leq b)$ is given by the definite integral $\int_a^b p(x)\,\mathrm{d}x$, we need to know **how to calculate definite integrals** to understand the behavior of the continuous RV X.

There are two directions to calculate definite integrals.

- **Numerical integration** by approximating the area by shapes of which we can calculate the area easier.
- **Analytical integration** by conducting integration as the inverse operation of differentiation.

In general, numerical integration methods[3] can apply to a variety of cases but cause an approximation error. The analytical integration methods can give us the exact value but have limited applications. In practice, we combine them depending on the situation. In this lecture, **we focus on analytical methods**. It also helps us learn numerical integration.

[3] e.g., the trapezoidal rule, the Gauss-Legendre quadrature rule, the double exponential formula

Basic idea of calculating an integral

When we constructed the probability density function (PDF) p_X, we differentiated the cumulative distribution function (CDF) F_X. Specifically, $p_X(x) = \frac{\mathrm{d}}{\mathrm{d}x} F_X(x)$. Conversely, we observed that the area $\int_a^b p_X(x)\,\mathrm{d}x$ under the graph of the PDF corresponds to the difference $F_X(b) - F_X(a)$.

Basic idea of calculating an integral

When we constructed the probability density function (PDF) p_X, we differentiated the cumulative distribution function (CDF) F_X. Specifically, $p_X(x) = \dfrac{\mathrm{d}}{\mathrm{d}x} F_X(x)$. Conversely, we observed that the area $\displaystyle\int_a^b p_X(x)\,\mathrm{d}x$ under the graph of the PDF corresponds to the difference $F_X(b) - F_X(a)$.

To wrap up, to calculate the definite integral $\displaystyle\int_a^b p_X(x)\,\mathrm{d}x$, we can use a function whose derivative is p_X.

Basic idea of calculating an integral

When we constructed the probability density function (PDF) p_X, we differentiated the cumulative distribution function (CDF) F_X. Specifically, $p_X(x) = \frac{\mathrm{d}}{\mathrm{d}x} F_X(x)$. Conversely, we observed that the area $\int_a^b p_X(x)\,\mathrm{d}x$ under the graph of the PDF corresponds to the difference $F_X(b) - F_X(a)$.

To wrap up, to calculate the definite integral $\int_a^b p_X(x)\,\mathrm{d}x$, we can use a function whose derivative is p_X.

According to the ***fundamental theorem of calculus (FTC)***, this relation between the derivative and the definite integral applies to a general function. We can use this relation to calculate a definite integral.

Integral is the “inverse” of differentiation

Definition (Primitive function)

Let a and b be real numbers such that $a < b$ and $f : [a,b] \to \mathbb{R}$. If $F : [a,b] \to \mathbb{R}$ satisfies $F' = f$, i.e., $\frac{\mathrm{d}}{\mathrm{d}x} F(x) = f(x)$ for all $x \in [a,b]$, then F is called a ***primitive function*** or an ***antiderivative function*** of f.

Theorem (The fundamental theorem of calculus (FTC))

Let a and b be real numbers such that $a < b$ and $f : [a,b] \to \mathbb{R}$ be integrable. Suppose that there exists a primitive function $F : [a,b] \to \mathbb{R}$ of f, then we have that

$$\int_a^b f(t)\,\mathrm{d}t = F(b) - F(a). \tag{5}$$

We often denote $F(b) - F(a)$ by $[F(x)]_a^b$.

According to the FTC, we can **calculate an integral using a primitive function**!

Calculating the definite integral

To calculate the definite integral

$$\int_a^b f(x)\,\mathrm{d}x\,, \tag{6}$$

the following steps suffice.

- **Step 1**: Find a primitive (antiderivative) function $F : [a,b] \to \mathbb{R}$, which satisfies $F' = f$.
- **Step 2**: Evaluate the value of $[F(x)]_a^b := F(b) - F(a)$.

Examples of calculating a definite integral

Example

Let $f(x) = x$.

We can calculate the definite integral $\displaystyle\int_{-4}^{5} f(x)\,\mathrm{d}x = \int_{-4}^{5} x\,\mathrm{d}x$ as follows.

- **Step 1**:

- **Step 2**:

Examples of calculating a definite integral

Example

Let $f(x) = x$.

We can calculate the definite integral $\displaystyle\int_{-4}^{5} f(x)\,\mathrm{d}x = \int_{-4}^{5} x\,\mathrm{d}x$ as follows.

- **Step 1**: Find a primitive (antiderivative) function F, which satisfies $F' = f$. In this example case, we can use a function $F(x) = \frac{1}{2}x^2$ as a primitive function since $\frac{\mathrm{d}}{\mathrm{d}x}\frac{1}{2}x^2 = x$.
- **Step 2**:

Examples of calculating a definite integral

Example

Let $f(x) = x$.

We can calculate the definite integral $\int_{-4}^{5} f(x)\,\mathrm{d}x = \int_{-4}^{5} x\,\mathrm{d}x$ as follows.

- **Step 1**: Find a primitive (antiderivative) function F, which satisfies $F' = f$. In this example case, we can use a function $F(x) = \frac{1}{2}x^2$ as a primitive function since $\frac{\mathrm{d}}{\mathrm{d}x}\frac{1}{2}x^2 = x$.
- **Step 2**: Evaluate the value of $[F(x)]_{-4}^{5} := F(5) - F(-4)$. In this example case, $F(5) - F(-4) = \frac{1}{2}(5)^2 - \frac{1}{2}(-4)^2 = \frac{25}{2} - 8 = \frac{9}{2}$.

Examples of calculating a definite integral

Example

Let $f(x) = x$.

We can calculate the definite integral $\int_{-4}^{5} f(x)\,\mathrm{d}x = \int_{-4}^{5} x\,\mathrm{d}x$ as follows.

- **Step 1**: Find a primitive (antiderivative) function F, which satisfies $F' = f$. In this example case, we can use a function $F(x) = \frac{1}{2}x^2$ as a primitive function since $\frac{\mathrm{d}}{\mathrm{d}x}\frac{1}{2}x^2 = x$.
- **Step 2**: Evaluate the value of $[F(x)]_{-4}^{5} := F(5) - F(-4)$. In this example case, $F(5) - F(-4) = \frac{1}{2}(5)^2 - \frac{1}{2}(-4)^2 = \frac{25}{2} - 8 = \frac{9}{2}$.

Hence, we have that

$$\int_{-4}^{5} f(x)\,\mathrm{d}x = \frac{9}{2}. \tag{7}$$

A primitive function is not unique.

As we have seen, finding a primitive function is essential to calculate the definite integral. Here, we must note that a primitive function is not unique.

If a function $F_1 : [a,b] \to \mathbb{R}$ is a primitive function of $f : [a,b] \to \mathbb{R}$, then $F_2 : [a,b] \to \mathbb{R}$ defined by $F_2(x) = F_1(x) + C$ is also a primitive function, where $C \in \mathbb{R}$ is a constant.

Example

Both $F_1(x) = \frac{1}{2}x^2$ and $F_2(x) = \frac{1}{2}x^2 + 5$ are primitive functions of $f(x) = x$.

The primitive function is unique up to an additive constant.

A primitive function is not unique. **However**, it is unique **up to an additive constant** in the following sense.

The primitive function is unique up to an additive constant.

A primitive function is not unique. **However**, it is unique **up to an additive constant** in the following sense.

Theorem (The primitive function is unique up to an additive constant)

Let a and b be real values such that $a < b$. If both $F_1 : [a,b] \to \mathbb{R}$ and $F_2 : [a,b] \to \mathbb{R}$ are primitive functions of f, the difference between F_1 and F_2 is a constant function. In other words, there exists a constant $C \in \mathbb{R}$ such that $F_2(x) - F_1(x) = C$.

The primitive function is unique up to an additive constant.

A primitive function is not unique. **However**, it is unique **up to an additive constant** in the following sense.

Theorem (The primitive function is unique up to an additive constant)

Let a and b be real values such that $a < b$. If both $F_1 : [a,b] \to \mathbb{R}$ and $F_2 : [a,b] \to \mathbb{R}$ are primitive functions of f, the difference between F_1 and F_2 is a constant function. In other words, there exists a constant $C \in \mathbb{R}$ such that $F_2(x) - F_1(x) = C$.

To wrap up, if F is a primitive function of f, then, for any constant C, the function given by $F(x) + C$ is also a primitive function of f, and conversely, all the primitive functions are written in this form. We write this fact as follows.

$$\int f(x)\,\mathrm{d}x = F(x) + C, \tag{8}$$

Here, the symbol $\int f(x)\,\mathrm{d}x$ in the LHS denotes all the primitive functions of f. Here, the constant C in the RHS is called the ***constant of integration***.

Examples of primitive functions

Example

The function $F(x) = \frac{1}{2}x^2$ is a primitive function of $f(x) = x$ since $F'(x) = f(x)$. Hence, $\displaystyle\int f(x)\,\mathrm{d}x = \frac{1}{2}x^2 + C$. Here, C is the constant of integration.

Note about the proof

We can prove the uniqueness of the primitive function up to an additive constant by the ***mean value theorem***.

Indefinite integral

Let f be a function and a be a real value. The function defined by the following form is called an ***indefinite integral*** of f.

$$\int_a^x f(t)\,\mathrm{d}t\,. \tag{9}$$

It is known that if f be continuous, then an indefinite integral is a primitive function of f. Note that some literature use the term “indefinite integral” to refer to a primitive function for this reason, while not all primitive functions are written in the above form.

Linearity of the antidifferentiation and integral

Since the derivation is a linear operator, the antidifferentiation, the operation to find a primitive function, is linear as well in the following sense.

Theorem (Linearity of antidifferentiation)

Let $f, g : \mathbb{R} \to \mathbb{R}$ be functions and F and G be the primitive functions of f and g, respectively. Also, let α and β be real values.
Then, $\alpha F + \beta G$ is a primitive function of $\alpha f + \beta g$.
In other words,

$$\int \left(\alpha f(x) + \beta g(x)\right) \mathrm{d}x = \alpha \int f(x) \, \mathrm{d}x + \beta \int g(x) \, \mathrm{d}x . \tag{10}$$

We can easily prove the above by taking the derivatives of both sides[4].

[4] Strictly speaking, we should consider the uniqueness of the primitive function up to an additive constant

Linearity of the definite integral

By combining the linearity of the antidifferentiation and the FTC, we can immediately get the linearity of the definite integral, which is a useful formula.

Corollary (Linearity of the definite integral)

Let $f, g : \mathbb{R} \to \mathbb{R}$ *be functions and* a, b, α *and* β *be real values. Then,*

$$\int_a^b \big(\alpha f(x) + \beta g(x)\big)\,\mathrm{d}x = \alpha \int_a^b f(x)\,\mathrm{d}x + \beta \int_a^b g(x)\,\mathrm{d}x\,. \tag{11}$$

Example of the linearity of antidifferentiation

Example

In the following, C is the constant of integration.

- $\displaystyle\int (\cos x + x^2)\,\mathrm{d}x = \int \cos x\,\mathrm{d}x + \int x^2\,\mathrm{d}x = \sin x + \frac{1}{3}x^3 + C$. Hence,
  $\displaystyle\int_0^{\pi} (\cos x + x^2)\,\mathrm{d}x = \left(\sin \pi + \frac{1}{3}\pi^3\right) - \left(\sin 0 + \frac{1}{3}\cdot 0^3\right) = \frac{1}{3}\pi^3.$
- $\displaystyle\int 5\exp(x)\,\mathrm{d}x = 5\int \exp(x)\,\mathrm{d}x = 5\exp(x) + C$. Hence,
  $\displaystyle\int_1^3 5\exp(x)\,\mathrm{d}x = (5\exp(3)) - (5\exp(1)) = 5\mathrm{e}(\mathrm{e}^2 - 1).$

Finding the primitive function is not always easy.

To calculate the derivative, we had many useful formulae. Let $f, g : \mathbb{R} \to \mathbb{R}$ be differentiable functions, then, e.g.,

- $(fg)' = f'g + fg'$ for the product,
- $(g \circ f)' = \left(g' \circ f\right) f'$ for the composition.

Recall that the composition $g \circ f$ is defined by $(g \circ f)(x) = g(f(x))$.

However, generally speaking, antidifferentiation is more difficult than differentiation. Specifically, we have no formulae to find a primitive function of a general product or composition like in differentiation. Nevertheless, we have some techniques to make such calculation more feasible for some cases, called ***integration by parts*** and ***integration by substitution***.

Integration by parts

Let f and g be real functions and F and G be those primitive functions. While we cannot generally write the primitive function of the product fg only by F and G, the technique, called ***integration by parts***, based on the following equation might help.

$$\int f(x)g(x)\,\mathrm{d}x = f(x)G(x) - \int f'(x)G(x)\,\mathrm{d}x\,. \tag{12}$$

Note that we assume that f is differentiable in the above.

By the above equation, we can find the primitive function of fg as long as we know that of $f'G$.

The proof of the above equation is easy if we differentiate the RHS.

Integration by parts

Let f and g be real functions and F and G be those primitive functions. While we cannot generally write the primitive function of the product fg only by F and G, the technique, called ***integration by parts***, based on the following equation might help.

$$\int f(x)g(x)\,\mathrm{d}x = f(x)G(x) - \int f'(x)G(x)\,\mathrm{d}x\,. \tag{12}$$

Example

$$\begin{aligned}\int x\cos(x)\,\mathrm{d}x &= x\sin(x) - \int (x)'\sin(x)\,\mathrm{d}x \\ &= x\sin(x) - \int 1\cdot\sin(x)\,\mathrm{d}x \\ &= x\sin(x) - (-\cos x) + C.\end{aligned} \tag{13}$$

Integration by parts

Let f and g be real functions and F and G be those primitive functions. While we cannot generally write the primitive function of the product fg only by F and G, the technique, called ***integration by parts***, based on the following equation might help.

$$\int f(x)g(x)\,\mathrm{d}x = f(x)G(x) - \int f'(x)G(x)\,\mathrm{d}x\,. \tag{12}$$

Example

$$\begin{aligned}\int \log(x)\,\mathrm{d}x = \int \log(x)\cdot 1\,\mathrm{d}x = \log(x)\cdot x - \int (\log(x))'\cdot x\,\mathrm{d}x \\ = \log(x)\cdot x - \int \frac{1}{x}\cdot x\,\mathrm{d}x \\ = x\log(x) - x + C.\end{aligned} \tag{13}$$

Integration by substitution

Let f and g be real functions and assume f be differentiable. If the integrand includes the composition $g \circ f$, we cannot generally write the primitive function only by the primitive functions of f and g. However, we may find it by the following technique, called ***integration by substitution***.

Theorem (Integration by substitution for indefinite integral)

$$\int g(f(t))f'(t)\,\mathrm{d}t = \int g(x)\,\mathrm{d}x \bigg|_{x=f(t)}, \tag{14}$$

where the RHS means the function we obtain by substituting $x = f(t)$ to a primitive function of g.

Both directions of the above equation are useful.

Integration by substitution

Let f and g be real functions and assume f be differentiable. If the integrand includes the composition $g \circ f$, we cannot generally write the primitive function only by the primitive functions of f and g. However, we may find it by the following technique, called ***integration by substitution***.

Theorem (Integration by substitution for definite integral)

$$\int_a^b g(f(t))f'(t)\,\mathrm{d}t = \int_{f(a)}^{f(b)} g(x)\,\mathrm{d}x\,, \tag{14}$$

where the RHS means the function we obtain by substituting $x = f(t)$ for a primitive function of g.

Both directions of the above equation are useful.

Why do we call it integration by substitution?

The previous page's formula is called integration by substitution because the formula is informally given by substituting $x = f(t)$ as follows.

$$\begin{aligned}
\int_a^b g(f(t))f'(t)\,\mathrm{d}t &= \int_{t=a}^{t=b} g(f(t))\frac{\mathrm{d}f(t)}{\mathrm{d}t}\,\mathrm{d}t \\
&= \int_{t=a}^{t=b} g(x)\frac{\mathrm{d}x}{\mathrm{d}t}\,\mathrm{d}t \\
&= \int_{t=a}^{t=b} g(x)\,\mathrm{d}x \\
&= \int_{x=f(a)}^{x=f(b)} g(x)\,\mathrm{d}x\,.
\end{aligned} \tag{15}$$

Note that the above discussion is mathematically inaccurate (especially where we used $\frac{\mathrm{d}x}{\mathrm{d}t}\,\mathrm{d}t = \mathrm{d}x$). If we want to formally prove the formula, we should simply differentiate both sides of the formula for indefinite integral.

Examples of integration by substitution.

Recall the formula.

$$\int_a^b g(f(t))f'(t)\,\mathrm{d}t = \int_{f(a)}^{f(b)} g(x)\,\mathrm{d}x\,, \tag{16}$$

Example (integration by substitution: from left to right)

$$\begin{aligned}
\int_0^{+2} t\exp\left(-t^2\right)\mathrm{d}t &= -\frac{1}{2}\int_0^{+2}\exp\left(-t^2\right)\cdot(-2t)\,\mathrm{d}t \\
&= -\frac{1}{2}\int_0^{+2}\exp\left(-t^2\right)\cdot\left(-t^2\right)'\,\mathrm{d}t \\
&= -\frac{1}{2}\int_{-0^2}^{-2^2}\exp(x)\,\mathrm{d}x \\
&= -\frac{1}{2}[\exp(x)]_{-0^2}^{-2^2} = -\frac{1}{2}[\exp(-4)-\exp(0)] = \frac{1}{2}[1-\exp(-4)].
\end{aligned} \tag{17}$$

Examples of integration by substitution.

Recall the formula.

$$\int_a^b g(f(t))f'(t)\,\mathrm{d}t = \int_{f(a)}^{f(b)} g(x)\,\mathrm{d}x\,, \tag{16}$$

Example (integration by substitution: from right to left)

$$\begin{aligned}
\int_0^1 \sqrt{1-x^2}\,\mathrm{d}x &= \int_{\frac{\pi}{2}}^0 \sqrt{1-\cos^2(t)}(\cos(t))'\,\mathrm{d}t \quad \text{since } \cos\left(\frac{\pi}{2}\right)=0, \cos(0)=1,\\
&= \int_{\frac{\pi}{2}}^0 \sqrt{1-\cos^2(t)}(-\sin(t))\,\mathrm{d}t \\
&= \int_0^{\frac{\pi}{2}} \sin^2(t)\,\mathrm{d}t = \int_0^{\frac{\pi}{2}} \frac{1-\cos(2t)}{2}\,\mathrm{d}t = \left[\frac{1}{2}t - \frac{1}{4}\sin(2t)\right]_0^{\frac{\pi}{2}} = \frac{1}{4}\pi.
\end{aligned} \tag{17}$$

Example of definite integral calculation in probability theory

Example (Exponential distribution)

The distribution of a RV X is called the ***exponential distribution*** with mean μ if it has a PDF p_X given by

$$p_X(x) := \begin{cases} 0 & \text{if } x < 0, \\ \frac{1}{\mu}\exp\left(-\frac{x}{\mu}\right) & \text{if } x \geq 0. \end{cases} \tag{18}$$

For nonnegative numbers a and b, the probability $\Pr(a < X \leq b)$ is given by

$$\begin{aligned} \Pr(a < X \leq b) &= \int_a^b p_X(x)\,\mathrm{d}x = \int_a^b \frac{1}{\mu}\exp\left(-\frac{x}{\mu}\right)\mathrm{d}x \\ &= \left[-\exp\left(-\frac{x}{\mu}\right)\right]_a^b = \exp\left(-\frac{a}{\mu}\right) - \exp\left(-\frac{b}{\mu}\right). \end{aligned} \tag{19}$$

A primitive function of the product/composition is not easily found.

We know that the primitive functions of $\frac{1}{x}$ and $\sin$, or $\exp$ and $-x^2$. Indeed,

$$\int \frac{1}{x}\,\mathrm{d}x = \log|x| + C, \int \sin x\,\mathrm{d}x = -\cos + C, \int \left(-x^2\right)\mathrm{d}x = -\frac{1}{3}x^3 + C, \int \exp(x)\,\mathrm{d}x = \exp(x) + C. \tag{20}$$

However, it is known that the primitive functions of $\frac{1}{x}\sin x$ and $\exp\left(-x^2\right)$ are not ***elementary***, although $\frac{1}{x}\sin x$ and $\exp\left(-x^2\right)$ themselves are elementary.

Here, we call a function ***elementary*** if we can write the function as a composition of finitely many

- algebraic functions, functions represented as a root of polynomial-function-coefficient polynomial equations, including polynomial, rational functions and fractional powers, e.g., $5x^2 + x - 3$, $\sqrt{3}x + 5$, $\frac{3x+1}{-2x^2+x+5}$, etc.
- trigonometric functions, e.g., $\sin x$, $\cos x$ etc.,
- exponential function $\exp x$,
- logarithmic function $\log x$.

A primitive function of the product/composition is not easily found.

We know that the primitive functions of $\frac{1}{x}$ and $\sin$, or $\exp$ and $-x^2$. Indeed,

$$\int \frac{1}{x}\,\mathrm{d}x = \log|x| + C, \int \sin x\,\mathrm{d}x = -\cos + C, \int \left(-x^2\right)\mathrm{d}x = -\frac{1}{3}x^3 + C, \int \exp(x)\,\mathrm{d}x = \exp(x) + C. \tag{20}$$

However, it is known that the primitive functions of $\frac{1}{x}\sin x$ and $\exp\left(-x^2\right)$ are not ***elementary***, although $\frac{1}{x}\sin x$ and $\exp\left(-x^2\right)$ themselves are elementary.

Roughly speaking, most functions we can imagine without the inverse function and the primitive function are elementary.

The fact that the primitive functions of $\frac{1}{x}\sin x$ and $\exp\left(-x^2\right)$ are not elementary means we have no way to write those primitive functions.

From the computer science viewpoint, the above fact means that we cannot easily find the exact value of the integrals of those functions. Some non-elementary primitive functions might be implemented by some libraries if they are famous. If they are not implemented, you might need to calculate the definite integral using a numerical method.

A primitive function of the product/composition is not easily found.

We know that the primitive functions of $\frac{1}{x}$ and $\sin$, or $\exp$ and $-x^2$. Indeed,

$$\int \frac{1}{x}\,\mathrm{d}x = \log|x| + C, \int \sin x\,\mathrm{d}x = -\cos + C, \int \left(-x^2\right)\mathrm{d}x = -\frac{1}{3}x^3 + C, \int \exp(x)\,\mathrm{d}x = \exp(x) + C. \tag{20}$$

However, it is known that the primitive functions of $\frac{1}{x}\sin x$ and $\exp\left(-x^2\right)$ are not ***elementary***, although $\frac{1}{x}\sin x$ and $\exp\left(-x^2\right)$ themselves are elementary.

In fact, these functions are important in many areas.

- The PDF of the normal distribution is proportional to $\exp\left(-x^2\right)$. The normal distribution is the most important distribution in probability theory, owing to the central limit theorem.
- The sine cardinal function $\frac{\sin x}{x}$ appears in many application areas, including physics, probability theory, signal processing, optics, etc., because it is the Fourier transform of the rectangle function.

Outline

1. **Continuous Random Variables**

   - Summary statistics of continuous RV and integral

Expectation (mean) of a continuous random variable

The expectation of a continuous RV is defined similarly to that of a discrete RV. Specifically, we get the definition for a continuous RV by replacing the PMF and the sum with the PDF and the integration in the definition for a discrete RV.

Expectation (mean) of a continuous random variable

The expectation of a continuous RV is defined similarly to that of a discrete RV. Specifically, we get the definition for a continuous RV by replacing the PMF and the sum with the PDF and the integration in the definition for a discrete RV.

Definition (Expectation of a continuous RV)

Let X be a continuous RV and p_X be its probability density function (PDF). Then, the expectation $\mathbb{E}X$ of X is defined by

$$\mathbb{E}X := \int_{-\infty}^{+\infty} x p(x) \, \mathrm{d}x \,. \tag{21}$$

Cf.) The expectation of a discrete RV X is given by $\sum_{x \in \mathcal{X}} x P_X(x)$, where P_X is the probability mass function.

Example: expectation of exponential distribution

Example (Expectation of the exponential distribution)

The PDF p_X of a RV X following the exponential distribution with mean μ is given by

$$p_X(x) := \begin{cases} 0 & \text{if } x < 0, \\ \frac{1}{\mu}\exp\left(-\frac{x}{\mu}\right) & \text{if } x \geq 0. \end{cases} \tag{22}$$

Noting that the density is zero for the negative domain, we can calculate the expectation $\mathbb{E}X$ using integration by parts as follows.

$$\begin{aligned} \mathbb{E}X &= \int_{-\infty}^{+\infty} x p_X(x)\,\mathrm{d}x = \int_0^{+\infty} x \cdot \frac{1}{\mu}\exp\left(-\frac{x}{\mu}\right)\mathrm{d}x = \int_0^{+\infty} x\left(-\exp\left(-\frac{x}{\mu}\right)\right)'\mathrm{d}x \\ &= \left[x \cdot \left(-\exp\left(-\frac{x}{\mu}\right)\right)\right]_0^{+\infty} - \int_0^{+\infty} (x)' \cdot \left(-\exp\left(-\frac{x}{\mu}\right)\right)\mathrm{d}x = -\int_0^{+\infty}\left(-\exp\left(-\frac{x}{\mu}\right)\right)\mathrm{d}x \\ &= -\left[\mu\exp\left(-\frac{x}{\mu}\right)\right]_0^{+\infty} = \mu. \end{aligned} \tag{23}$$

The expectation of a function of a continuous RV

A function of a discrete RV is always a discrete RV. However, a function of a continuous RV is not always a continuous RV.

The expectation of a function of a continuous RV

A function of a discrete RV is always a discrete RV. However, a function of a continuous RV is not always a continuous RV.

For example, if f is the sign function defined by

$$f(x) := \begin{cases} -1 & \text{if } x < 0, \\ 0 & \text{if } x = 0, \\ +1 & \text{if } x > 0, \end{cases} \tag{24}$$

and X is a continuous RV whose PDF p_X is given by

$$p_X(x) = \begin{cases} +1 & \text{if } -\frac{1}{2} \leq x \leq +\frac{1}{2}, \\ 0 & \text{otherwise}. \end{cases} \tag{25}$$

Then, the RV $f(X)$ takes values -1 and $+1$ with equal probability. In particular, it is a discrete RV, whose support is $\{-1, +1\}$.

The expectation of a function of a continuous RV

A function of a discrete RV is always a discrete RV. However, a function of a continuous RV is not always a continuous RV.

Even though a function of a continuous RV may not be a continuous RV, its expectation can always be calculated by the following formula, which is similar to the formula for a discrete RV.

Theorem

Let X be a continuous RV and its PDF be p_X. Also, let $f : \mathbb{R} \to \mathbb{R}$ be a real-valued function taking a real value as an input. The expectation $\mathbb{E} f(X)$ of the random variable X is given as follows.

$$\mathbb{E} f(X) = \int_{-\infty}^{+\infty} f(x) p_X(x) \, \mathrm{d}x \,. \tag{24}$$

Cf.) For a discrete RV whose support and PMF are $\mathscr{X}$ and P_X, respectively, we have that $\mathbb{E} f(X) = \sum_{x \in \mathscr{X}} f(x) P_X(x)$.

The linearity of the expectation on continuous RVs

The following theorem, which holds for a discrete RV, also holds for a continuous RV.

Theorem (The linearity of the expectation)

Let X be a random variable, $a, b \in \mathbb{R}$ be real numbers, and $f, g : \mathbb{R} \to \mathbb{R}$ be real-valued functions taking a real variable. Then, we have that

$$\mathbb{E}[af(X) + bg(X)] = a\,\mathbb{E}f(X) + b\,\mathbb{E}g(X). \tag{25}$$

Variance and standard deviation of a continuous random variable

The definitions of the variance and standard deviation are the same for a continuous RV. Specifically, for a continuous RV X, whose expectation is μ_X, its variance $\mathbb{V}(X)$ is defined by $\mathbb{V}(X) := \mathbb{E}\left(X - \mu_X\right)^2$. The standard deviation is defined by $\sigma_X := \sqrt{\mathbb{V}(X)}$.

When we know the explicit form of the PDF, we can use the following formulae.

Theorem

Let X be a continuous RV and its PDF be p_X. Suppose that the expectation $\mathbb{E}X = \int_{-\infty}^{+\infty} x p_X(x)\,\mathrm{d}x$ exists and denote it by μ_X. The variance $\mathbb{V}(X)$ is given by the following formula.

$$\mathbb{V}(X) = \int_{-\infty}^{+\infty} \left(x - \mu_X\right)^2 p_X(x)\,\mathrm{d}x = \int_{-\infty}^{+\infty} x^2 p_X(x)\,\mathrm{d}x - \left(\mu_X\right)^2. \tag{26}$$

Outline

1 Continuous Random Variables

- Jointly continuous random variables and multiple integral

Handling multiple non-discrete random variables

Similar to the univariate random variable case, we can define the cumulative distribution function (CDF) for multiple RV even if they are not discrete.

Definition (The CDF of two RVs)

Let X and Y be random variables. The ***cumulative distribution function (CDF)*** $F_{X,Y} : \mathbb{R}^2 \to [0,1]$ of X and Y is defined by

$$F_{X,Y}(x,y) := \Pr(X \leq x \wedge Y \leq y), \tag{27}$$

where $\wedge$ indicates the logical "and" statement.

Using the CDF, we can calculate the probability $\Pr(a_1 < X \leq b_1 \wedge a_2 < Y \leq b_2)$ by

$$\Pr(a_1 < X \leq b_1 \wedge a_2 < Y \leq b_2) = F_{X,Y}(b_1,b_2) - F_{X,Y}(a_1,b_2) - F_{X,Y}(b_1,a_2) + F_{X,Y}(b_1,b_2) \tag{28}$$

To define the probability density function for a multivariate RV

Let $X_1, X_2, \ldots, X_m$ be random variables. As in the univariate random variable case, the CDF may not be easy to interpret or not be elementary even in practical cases. Hence, we want to define the probability density function (PDF) for multiple RV cases.

To define the probability density function for a multivariate RV

Let $X_1, X_2, \ldots, X_m$ be random variables. As in the univariate random variable case, the CDF may not be easy to interpret or not be elementary even in practical cases. Hence, we want to define the probability density function (PDF) for multiple RV cases.

The univariate continuous RV theory allows us to define the PDF for each RV, but they are not sufficient to understand the behavior of a multiple RVs completely, as we saw in multiple discrete RV cases. To tackle this issue, for discrete RV cases, we evaluated the Joint PMF, which returns the probability mass of the event $(X_1, X_2, \ldots, X_m) = (x_1, x_2, \ldots, x_m)$. Similarly, we want to define the function that returns the probability density at $(X_1, X_2, \ldots, X_m) = (x_1, x_2, \ldots, x_m)$. Since the PDF for a univariate continuous RV was defined using the area under the graph, let us define the graph of a multivariate function and the high-dimensional area (volume) in the following.

The graph of a multivariate function and multiple integral

Let $D = [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_m, b_m]$ be a m-dimensional hyper-rectangle. Let $f : D \to \mathbb{R}$ be a function of a m-dimensional variable. Similar to one-dimensional function cases, we call the set of points

$$\{(x_1, x_2, \ldots, x_m, f(x_1, x_2, \ldots, x_m)) | (x_1, x_2, \ldots, x_m) \in D\} \tag{29}$$

the ***graph*** of a function f. The (signed) volume in the domain D bounded by the graph of $y = f(\boldsymbol{x})$ and $y = 0$ is called the ***multiple integral*** of f on D, denoted by $\int_D f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$.

Joint PDF

Based on the definition of multiple integration, we can define the joint probability density function (joint PDF) of a multivariate random variable.

Definition

Let $X_1, X_2, \ldots, X_m$ be random variables. If $p_{X_1,X_2,\ldots,X_m} : \mathbb{R}^m \to \mathbb{R}_{\geq 0}$ satisfies

$$\Pr((X_1, X_2, \ldots, X_m) \in D) = \int_D p_{X_1,X_2,\ldots,X_m}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \tag{30}$$

for any m-dimensional hyper-rectangle D, then the function $p_{X_1,X_2,\ldots,X_m}$ is called the ***joint probability density function (joint PDF)*** of $X_1, X_2, \ldots, X_m$.
If $(X_1, X_2, \ldots, X_m)$ have a joint PDF, we call them ***jointly continuous random variables (jointly continuous RVs)***.

Multiple continuous RVs are not always jointly continuous

Let $X_1, X_2, \ldots, X_m$ be continuous RVs. In other words, suppose that there exist PDFs $p_{X_1}, p_{X_2}, \ldots, p_{X_m}$ for $X_1, X_2, \ldots, X_m$, respectively.

Even under this assumption, **it is possible that $X_1, X_2, \ldots, X_m$ have no joint PDF**.

Multiple continuous RVs are not always jointly continuous

Let $X_1, X_2, \ldots, X_m$ be continuous RVs. In other words, suppose that there exist PDFs $p_{X_1}, p_{X_2}, \ldots, p_{X_m}$ for $X_1, X_2, \ldots, X_m$, respectively.

Even under this assumption, **it is possible that $X_1, X_2, \ldots, X_m$ have no joint PDF**.

Example

For example, let X and Y be a continuous RV following the uniform distribution on $[0,1]$ and suppose that $X = Y$ always hold. Then, both X and Y have the same PDF $p_X(z) = p_Y(z) \begin{cases} 1 & \text{if } 0 \leq z \leq 1, \\ 0 & \text{otherwise.} \end{cases}$, so both X and Y are continuous RVs. However, the probability mass concentrates on the segment from the origin $(0,0)$ to the point $(1,1)$ in the xy space. The segment has zero area, but if there existed the joint PDF, the volume bounded by the graph of the joint PDF on the segment would be 1. This is a contradiction, so X, Y have no joint PDF. Hence, X, Y are not jointly continuous.

Multiple continuous RVs are not always jointly continuous

Let $X_1, X_2, \ldots, X_m$ be continuous RVs. In other words, suppose that there exist PDFs $p_{X_1}, p_{X_2}, \ldots, p_{X_m}$ for $X_1, X_2, \ldots, X_m$, respectively.

Even under this assumption, **it is possible that $X_1, X_2, \ldots, X_m$ have no joint PDF**.

To wrap up, **even if both X and Y are continuous RVs, it does not follow that the X, Y are jointly continuous!** Conversely, jointly continuous RVs are always multiple continuous RVs.

For this reason, we need to **distinguish multiple continuous RVs and jointly continuous RVs**. The former is the broader concept, but we focus on the latter since we have many mathematical tools based on the multiple integration to analyze them.

Multiple integral on a complicated shape

In high-dimensional space, we might want to consider the volume bounded by a function in a complicated shape, say $\mathscr{A}$, that cannot be represented as a union of hyper-rectangles.

Multiple integral on a complicated shape

In high-dimensional space, we might want to consider the volume bounded by a function in a complicated shape, say $\mathscr{A}$, that cannot be represented as a union of hyper-rectangles.

For example, we might want to consider the probability $\Pr((X, Y) \in \mathscr{A})$, where $\mathscr{A} := \left\{(x, y) \in \mathbb{R}^2 \middle| x^2 + y^2 \leq 1\right\}$ is defined as the unit disk centered at the origin. We cannot decompose the disk into rectangles, so we cannot evaluate the probability by the sum rule if we can only define the probability of the multivariate RV being in a rectangle.

Multiple integral on a complicated shape

In high-dimensional space, we might want to consider the volume bounded by a function in a complicated shape, say $\mathscr{A}$, that cannot be represented as a union of hyper-rectangles.

Hence, we want to define the volume bounded by a function on a general set $\mathscr{A}$. We can do it by multiplying the value of the function by zero everywhere outside of $\mathscr{A}$ as follows.

Definition

Let $D = [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_m, b_m]$ be a m-dimensional hyper-rectangle.
For a general subset $\mathscr{A} \subset D$, we define the multiple integral of f on $\mathscr{A}$ by

$$\int_{\mathscr{A}} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} := \int_D 1_{\mathscr{A}}(\boldsymbol{x}) f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}, \tag{31}$$

where the indicator function $1_{\mathscr{A}}$ is defined by $1_{\mathscr{A}}(\boldsymbol{x}) := \begin{cases} 1 & \text{if } \boldsymbol{x} \in \mathscr{A}, \\ 0 & \text{if } \boldsymbol{x} \notin \mathscr{A}. \end{cases}$

Probability on a complicated shape

The probability density function can be applied to a complicated shape.

Theorem

Let $X_1, X_2, \ldots, X_m$ be jointly continuous RVs, and let $p_{X_1, X_2, \ldots, X_m}$ be the joint PDF. For $\mathscr{A} \in \mathbb{R}^m$, assume that it is bounded, i.e., there exists hyper-rectangle $D = [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_m, b_m]$ such that $\mathscr{A} \in D$. Then, we have that

$$\Pr((X_1, X_2, \ldots, X_m) \in \mathscr{A}) = \int_{\mathscr{A}} p_{X_1, X_2, \ldots, X_m}(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \tag{32}$$

The assumption about the boundedness of $\mathscr{A}$ will be removed later using improper integrations.

We can calculating a multiple integral by the iterated integral

We need to calculate a multiple integral to evaluate the probability of an event related to jointly continuous RVs. How can we do that?

Actually, we can calculate a multiple integral by the ***iterated integral***, according to Fubini-Tonelli Theorem.

Theorem (Fubini-Tonelli Theorem)

Let $f : \mathbb{R}^m \to \mathbb{R}$ be a function and $\mathscr{A} \subset \mathbb{R}^m$ be a subset of $\mathbb{R}^m$ and suppose that there exists a bounded m-dimensional hyper-rectangle $D = [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_m, b_m]$.
Then, under a certain loose conditions, we have that

$$\int_{\mathscr{A}} f(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} = \int_{a_m}^{b_m} \cdots \left(\int_{a_2}^{b_2} \left(\int_{a_1}^{b_1} 1_{\mathscr{A}}(\boldsymbol{x}) f(\boldsymbol{x}) \, \mathrm{d}x_1 \right) \mathrm{d}x_2 \right) \cdots \mathrm{d}x_m . \tag{33}$$

Note that the order of the indices is exchangeable.

Special case: calculating a double integral

A bivariable multiple integral is called a ***double integral***. The formula for a double integral is given as follows.

Corollary (Fubini-Tonelli theorem on a double integral)

Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a function and $\mathscr{A} \subset \mathbb{R}^2$ be a subset of $\mathbb{R}^2$ and suppose that there exists a bounded 2-dimensional hyper-rectangle $D = [a_1, b_1] \times [a_2, b_2]$. Then, under a certain loose conditions, we have that

$$\begin{aligned} \iint_{\mathscr{A}} f(x_1, x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 &= \int_{a_2}^{b_2} \left[\int_{a_1}^{b_1} 1_{\mathscr{A}}(x_1, x_2) f(x_1, x_2)\,\mathrm{d}x_1 \right] \mathrm{d}x_2 \\ &= \int_{a_1}^{b_1} \left[\int_{a_2}^{b_2} 1_{\mathscr{A}}(x_1, x_2) f(x_1, x_2)\,\mathrm{d}x_2 \right] \mathrm{d}x_1 \,. \end{aligned} \tag{34}$$

Strictly speaking, the following condition must be satisfied for the Fubini-Tonelli theorem to hold, i.e., for the iterated integral to give the correct value of the multiple integral.

Condition: The following limit converges (note the absolute value operation).

$$\int_{a_m}^{b_m} \cdots \left(\int_{a_2}^{b_2} \left(\int_{a_1}^{b_1} 1_{\mathscr{A}}(\boldsymbol{x}) |f(\boldsymbol{x})| \, \mathrm{d}x_1 \right) \mathrm{d}x_2 \right) \cdots \mathrm{d}x_m \, . \tag{35}$$

However, the above is rarely an issue in engineering or computer science.

Improper multiple integral

We might want to evaluate the volume bounded by a function's graph in a unbounded domain $\mathscr{A}$. In that case, we define the ***improper multiple integral*** as follows.

Definition (Improper multiple integral)

Let $\mathscr{A} \subset \mathbb{R}^m$ and $f : \mathbb{R}^m \to \mathbb{R}$ be a function defined on $\mathbb{R}^m$. Denote by $\lim_{\substack{a \to -\infty \\ b \to +\infty}}$ the iterated limit operator $\lim_{b \to +\infty} \lim_{a \to -\infty}$. Assume that a certain loose condition is satisfied. Then, we define $\int_{\mathscr{A}} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$ by

$$\int_{\mathscr{A}} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} := \lim_{\substack{a_m \to -\infty \\ b_m \to +\infty}} \cdots \lim_{\substack{a_2 \to -\infty \\ b_2 \to +\infty}} \lim_{\substack{a_1 \to -\infty \\ b_1 \to +\infty}} \int_D 1_{\mathscr{A}}(\boldsymbol{x}) f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}, \tag{36}$$

where $D = [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_m, b_m]$.

Advanced: the loose condition

We assumed some condition in the previous slide. This is because, in fact, the definition in the previous slide is not standard and we usually use another definition for the integral of a function on $\mathbb{R}^m$.

However, if a condition is satisfied, the two definitions are consistent, which is why we imposed the condition. The condition is as follows:

Condition: The following limit converges (note the absolute value operation).

$$\lim_{\substack{a_m \to -\infty \\ b_m \to +\infty}} \int_{a_m}^{b_m} \cdots \left(\lim_{\substack{a_2 \to -\infty \\ b_2 \to +\infty}} \int_{a_2}^{b_2} \left(\lim_{\substack{a_1 \to -\infty \\ b_1 \to +\infty}} \int_{a_1}^{b_1} 1_{\mathscr{A}}(\boldsymbol{x}) |f(\boldsymbol{x})| \, \mathrm{d}x_1 \right) \mathrm{d}x_2 \right) \cdots \mathrm{d}x_m \,. \tag{37}$$

To prove that these two are equivalent, first we define it in the standard way based on the Lebesgue integral, and use the dominant convergence theorem and Fubini-Tonelli's theorem iteratively.

Calculating an improper multiple integral

We can calculate a improper multiple integral by an iterated improper integral.

Theorem (Calculating an improper multiple integral)

Let $\mathscr{A} \subset \mathbb{R}^m$ and $f : \mathbb{R}^m \to \mathbb{R}$ be a function defined on $\mathbb{R}^m$. Assume that the loose condition in the previous slide is satisfied. Then, we have that

$$\int_{\mathscr{A}} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = \int_{-\infty}^{+\infty} \cdots \left(\int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} 1_{\mathscr{A}}(\boldsymbol{x}) f(\boldsymbol{x})\,\mathrm{d}x_1 \right) \mathrm{d}x_2 \right) \cdots \mathrm{d}x_m . \tag{38}$$

Special case: an improper multiple integral on the whole space

By substituting $\mathscr{A}$ with $\mathbb{R}^m$, we can define and calculate the improper multiple integral on the whole space $\mathbb{R}^m$. Here, what we need to do is to substitute $1_{\mathbb{R}^m}(\boldsymbol{x}) = 1$ for any $\boldsymbol{x} \in \mathbb{R}^m$.

Corollary (Calculating an improper multiple integral on the whole space)

Let $f : \mathbb{R}^m \to \mathbb{R}$ be a function defined on $\mathbb{R}^m$. Assume that the loose condition is satisfied. Then, we have that

$$\begin{aligned}\int_{\mathbb{R}^m} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} &:= \lim_{\substack{a_m \to -\infty \\ b_m \to +\infty}} \cdots \lim_{\substack{a_2 \to -\infty \\ b_2 \to +\infty}} \lim_{\substack{a_1 \to -\infty \\ b_1 \to +\infty}} \int_D 1_{\mathbb{R}^m}(\boldsymbol{x}) f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \\ &= \int_{-\infty}^{+\infty} \cdots \left(\int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} f(\boldsymbol{x})\,\mathrm{d}x_1 \right) \mathrm{d}x_2 \right) \cdots \mathrm{d}x_m \,.\end{aligned} \tag{39}$$

Special case: an improper double integral

Corollary (Calculating an improper double integral)

Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a function and $\mathscr{A} \subset \mathbb{R}^2$ be a subset of $\mathbb{R}^2$. Then, under the loose condition, we have that

$$\begin{aligned}\iint_{\mathscr{A}} f(x_1, x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 &= \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} 1_{\mathscr{A}}(x_1, x_2) f(x_1, x_2)\,\mathrm{d}x_1 \right) \mathrm{d}x_2 \\ &= \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} 1_{\mathscr{A}}(x_1, x_2) f(x_1, x_2)\,\mathrm{d}x_2 \right) \mathrm{d}x_1 .\end{aligned} \tag{40}$$

Steps to calculate a double integral

Recall that we have

$$\iint_{\mathscr{A}} f(x_1,x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 = \int_{-\infty}^{+\infty}\left(\int_{-\infty}^{+\infty} 1_{\mathscr{A}}(x_1,x_2) f(x_1,x_2)\,\mathrm{d}x_1\right)\mathrm{d}x_2 = \int_{-\infty}^{+\infty}\left(\int_{-\infty}^{+\infty} 1_{\mathscr{A}}(x_1,x_2) f(x_1,x_2)\,\mathrm{d}x_2\right)\mathrm{d}x_1\,. \tag{41}$$

Hence, we can calculate the double integral $\iint_{\mathscr{A}} f(x_1,x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2$ as follows.

- **Step 1.**

- **Step 2.**

Steps to calculate a double integral

Recall that we have

$$\iint_{\mathscr{A}} f(x_1,x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 = \int_{-\infty}^{+\infty}\left(\int_{-\infty}^{+\infty} 1_{\mathscr{A}}(x_1,x_2)f(x_1,x_2)\,\mathrm{d}x_1\right)\mathrm{d}x_2 \\ = \int_{-\infty}^{+\infty}\left(\int_{-\infty}^{+\infty} 1_{\mathscr{A}}(x_1,x_2)f(x_1,x_2)\,\mathrm{d}x_2\right)\mathrm{d}x_1\,. \tag{41}$$

Hence, we can calculate the double integral $\iint_{\mathscr{A}} f(x_1,x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2$ as follows.

- **Step 1.** Find the function $g(x_2) := \int_{-\infty}^{+\infty} 1_{\mathscr{A}}(x_1,x_2)f(x_1,x_2)\,\mathrm{d}x_1$ by the improper integral with respect to x_1.
- **Step 2.**

Steps to calculate a double integral

Recall that we have

$$\begin{aligned}\iint_{\mathscr{A}} f(x_1,x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 &= \int_{-\infty}^{+\infty}\left(\int_{-\infty}^{+\infty} 1_{\mathscr{A}}(x_1,x_2) f(x_1,x_2)\,\mathrm{d}x_1\right)\mathrm{d}x_2 \\ &= \int_{-\infty}^{+\infty}\left(\int_{-\infty}^{+\infty} 1_{\mathscr{A}}(x_1,x_2) f(x_1,x_2)\,\mathrm{d}x_2\right)\mathrm{d}x_1\,.\end{aligned} \tag{41}$$

Hence, we can calculate the double integral $\iint_{\mathscr{A}} f(x_1,x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2$ as follows.

- **Step 1.** Find the function $g(x_2) := \int_{-\infty}^{+\infty} 1_{\mathscr{A}}(x_1,x_2) f(x_1,x_2)\,\mathrm{d}x_1$ by the improper integral with respect to x_1.
- **Step 2.** Evaluate the improper integral $\int_{-\infty}^{+\infty} g(x_2)\,\mathrm{d}x_2$ with respect to x_2.

Joint PDF example 1: Uniform distribution

Example (Uniform distribution)

Let X and Y be RVs following the bivariate uniform distribution with the support $[0,3]\times[-1,+1]$. The RVs X and Y have has the joint PDF

$$p_{X,Y}(x,y) = \begin{cases} \frac{1}{6} & \text{if } (x,y) \in [0,3]\times[-1,+1], \\ 0 & \text{if } (x,y) \notin [0,3]\times[-1,+1]. \end{cases} \tag{42}$$

For example, the probability $\Pr\left((X,Y) \in [0,\frac{1}{2}]\times[0,\frac{1}{4}]\right)$ is given by

$$\int_0^{\frac{1}{4}} \int_0^{\frac{1}{2}} \frac{1}{6}\,\mathrm{d}x\,\mathrm{d}y = \int_0^{\frac{1}{4}} \left[\frac{1}{6}x\right]_0^{\frac{1}{2}} \mathrm{d}y = \int_0^{\frac{1}{4}} \frac{1}{12}\,\mathrm{d}y = \left[\frac{1}{12}y\right]_0^{\frac{1}{4}} = \frac{1}{48} \tag{43}$$

Joint PDF example 2

Example

Let X_1, X_2 be jointly continuous RVs and assume that its joint PDF p_{X_1,X_2} is given by

$$p_{X_1,X_2}(x_1,x_2) = 1_{\mathscr{A}}(x_1,x_2) f(x_1,x_2), \tag{44}$$

where $f(x_1,x_2) = 3 - 3x_1 - \frac{3}{2}x_2$, and $\mathscr{A} = \left\{(x_1,x_2) \in \mathbb{R}^2 \middle| x_1 \geq 0, x_2 \geq 0, \frac{x_1}{1} + \frac{x_2}{2} \leq 1\right\}$.

In the following, we will first confirm that $\iint_{\mathbb{R}^2} p_{X_1,X_2}(x_1,x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 = 1$, then calculate the probability $\Pr((X,Y) \in \mathscr{B})$, where $\mathscr{B} = \left\{(x_1,x_2) \in \mathbb{R}^2 \middle| x_1 \geq 0, x_2 \geq 0, x_1 + x_2 \leq 1\right\}$.

Joint PDF example 2: (i) Confirming the integral on $\mathbb{R}^2$ is 1

Let's calculate $\int_{\mathbb{R}^2} p_{X_1,X_2}(x_1,x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 = 1$ by the iterated integration. We have that $\int_{\mathbb{R}^2} p_{X_1,X_2}(x_1,x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 = \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} p_{X_1,X_2}(x_1,x_2)\,\mathrm{d}x_2 \right) \mathrm{d}x_1$. We first evaluate the integral $\int_{-\infty}^{+\infty} p_{X_1,X_2}(x_1,x_2)\,\mathrm{d}x_2$.

Joint PDF example 2: (i) Confirming the integral on $\mathbb{R}^2$ is 1

Let's calculate $\int_{\mathbb{R}^2} p_{X_1,X_2}(x_1,x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 = 1$ by the iterated integration. We have that $\int_{\mathbb{R}^2} p_{X_1,X_2}(x_1,x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 = \int_{-\infty}^{+\infty}\left(\int_{-\infty}^{+\infty} p_{X_1,X_2}(x_1,x_2)\,\mathrm{d}x_2\right)\mathrm{d}x_1$. We first evaluate the integral $\int_{-\infty}^{+\infty} p_{X_1,X_2}(x_1,x_2)\,\mathrm{d}x_2$.

Since $p_{X_1,X_2}(x_1,x_2) = 1_{\mathscr{A}}(x_1,x_2) f(x_1,x_2)$, we have that

$$\int_{-\infty}^{+\infty} p_{X_1,X_2}(x_1,x_2)\,\mathrm{d}x_2 = \begin{cases} \int_0^{2-2x_1} f(x_1,x_2)\,\mathrm{d}x_2 & \text{if } 0 \le x_1 \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{45}$$

Joint PDF example 2: (i) Confirming the integral on $\mathbb{R}^2$ is 1

Let's calculate $\int_{\mathbb{R}^2} p_{X_1,X_2}(x_1,x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 = 1$ by the iterated integration. We have that $\int_{\mathbb{R}^2} p_{X_1,X_2}(x_1,x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 = \int_{-\infty}^{+\infty}\left(\int_{-\infty}^{+\infty} p_{X_1,X_2}(x_1,x_2)\,\mathrm{d}x_2\right)\mathrm{d}x_1$. We first evaluate the integral $\int_{-\infty}^{+\infty} p_{X_1,X_2}(x_1,x_2)\,\mathrm{d}x_2$.

Since $p_{X_1,X_2}(x_1,x_2) = 1_{\mathscr{A}}(x_1,x_2)f(x_1,x_2)$, we have that

$$\int_{-\infty}^{+\infty} p_{X_1,X_2}(x_1,x_2)\,\mathrm{d}x_2 = \begin{cases} \int_0^{2-2x_1} f(x_1,x_2)\,\mathrm{d}x_2 & \text{if } 0 \le x_1 \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{45}$$

Hence, we have that

$$\int_{\mathbb{R}^2} p_{X_1,X_2}(x_1,x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 = \int_0^1\left(\int_0^{2-2x_1} f(x_1,x_2)\,\mathrm{d}x_2\right)\mathrm{d}x_1 = 1. \tag{46}$$

Joint PDF example 2 (ii) Probability in Region $\mathscr{B}$

Let's calculate $\Pr((x_1, x_2) \in \mathscr{B}) = \int_{\mathbb{R}^2} 1_{\mathscr{B}}(x_1, x_2) p_{X_1,X_2}(x_1, x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 = 1$ by the iterated integration. We have that $\int_{\mathbb{R}^2} 1_{\mathscr{B}}(x_1, x_2) p_{X_1,X_2}(x_1, x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 = \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} 1_{\mathscr{B}}(x_1, x_2) p_{X_1,X_2}(x_1, x_2)\,\mathrm{d}x_2 \right) \mathrm{d}x_1$. We first evaluate the integral $\int_{-\infty}^{+\infty} 1_{\mathscr{B}}(x_1, x_2) p_{X_1,X_2}(x_1, x_2)\,\mathrm{d}x_2$.

Joint PDF example 2 (ii) Probability in Region $\mathscr{B}$

Let's calculate $\Pr((x_1,x_2) \in \mathscr{B}) = \int_{\mathbb{R}^2} 1_{\mathscr{B}}(x_1,x_2)p_{X_1,X_2}(x_1,x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 = 1$ by the iterated integration. We have that $\int_{\mathbb{R}^2} 1_{\mathscr{B}}(x_1,x_2)p_{X_1,X_2}(x_1,x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 = \int_{-\infty}^{+\infty}\left(\int_{-\infty}^{+\infty} 1_{\mathscr{B}}(x_1,x_2)p_{X_1,X_2}(x_1,x_2)\,\mathrm{d}x_2\right)\mathrm{d}x_1$. We first evaluate the integral $\int_{-\infty}^{+\infty} 1_{\mathscr{B}}(x_1,x_2)p_{X_1,X_2}(x_1,x_2)\,\mathrm{d}x_2$.

As $\mathscr{B} \subset \mathscr{A}$, we have that $1_{\mathscr{B}}(x_1,x_2)p_{X_1,X_2}(x_1,x_2) = 1_{\mathscr{B}}(x_1,x_2)f(x_1,x_2)$.

Joint PDF example 2 (ii) Probability in Region $\mathscr{B}$

Let's calculate $\Pr((x_1, x_2) \in \mathscr{B}) = \int_{\mathbb{R}^2} 1_{\mathscr{B}}(x_1, x_2) p_{X_1,X_2}(x_1, x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 = 1$ by the iterated integration. We have that $\int_{\mathbb{R}^2} 1_{\mathscr{B}}(x_1, x_2) p_{X_1,X_2}(x_1, x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 = \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} 1_{\mathscr{B}}(x_1, x_2) p_{X_1,X_2}(x_1, x_2)\,\mathrm{d}x_2 \right) \mathrm{d}x_1$. We first evaluate the integral $\int_{-\infty}^{+\infty} 1_{\mathscr{B}}(x_1, x_2) p_{X_1,X_2}(x_1, x_2)\,\mathrm{d}x_2$.

As $\mathscr{B} \subset \mathscr{A}$, we have that $1_{\mathscr{B}}(x_1, x_2) p_{X_1,X_2}(x_1, x_2) = 1_{\mathscr{B}}(x_1, x_2) f(x_1, x_2)$. Hence

$$\int_{-\infty}^{+\infty} p_{X_1,X_2}(x_1, x_2)\,\mathrm{d}x_2 = \begin{cases} \int_0^{1-x_1} f(x_1, x_2)\,\mathrm{d}x_2 & \text{if } 0 \le x_1 \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{47}$$

Hence, we have that

$$\int_{\mathbb{R}^2} p_{X_1,X_2}(x_1, x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 = \int_0^1 \left(\int_0^{1-x_1} f(x_1, x_2)\,\mathrm{d}x_2 \right) \mathrm{d}x_1 = \frac{3}{4}. \tag{48}$$

Integration by substitution for a multiple integral

When we want to evaluate a multiple integral of a complicatedly composed function, an integration by substitution might help, as it does for univariate case.

Theorem (Integration by substitution for a multiple integral)

Let $f : \mathbb{R}^m \to \mathbb{R}$ be a m-variable real-valued function and $\boldsymbol{\varphi} : \mathbb{R}^m \to \mathbb{R}^m$ be a bijective differentiable m-variable m-dimensional-vector-valued function. Also, let U be a subset of $\mathbb{R}^m$. Then we have the following.

$$\int_U f(\boldsymbol{\varphi}(\boldsymbol{u})) \left| \det\left(\frac{\partial \boldsymbol{\varphi}}{\partial \boldsymbol{u}}(\boldsymbol{u}) \right) \right| \mathrm{d}\boldsymbol{u} = \int_{\boldsymbol{\varphi}(U)} f(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}. \tag{49}$$

Here, $\det$ indicates the determinant, and $\frac{\partial \boldsymbol{\varphi}}{\partial \boldsymbol{u}}$ is the Jacobian of $\boldsymbol{\varphi}$.

Difference between a univariable integral and a multiple integral

Strictly, the previous slide's formula is not a strict extension of the univariable case since we have the absolute value operator outside the determinant of the Jacobian. This difference comes because we do not care the direction of the integral as we did in a univariable case.

Specifically, we distinguished $\int_a^b$ and $\int_b^a$ in the univariable case, but we do not care such differences in a multiple integral.

If you want to distinguish them in multiple integral, you can learn a ***differential form*** or ***volume form***.

Integration by substitution for a double integral

To see the formula in detail, let us consider the bivariable case.

Corollary (Integration by substitution for a double integral)

$$
\begin{aligned}
&\int_U f(\varphi_1(u_1,u_2),\varphi_2(u_1,u_2))\left|\det\left(\begin{bmatrix} \frac{\partial\varphi_1}{\partial u_1}(u_1,u_2) & \frac{\partial\varphi_1}{\partial u_2}(u_1,u_2) \\ \frac{\partial\varphi_2}{\partial u_1}(u_1,u_2) & \frac{\partial\varphi_2}{\partial u_2}(u_1,u_2)\end{bmatrix}\right)\right| \mathrm{d}u_1\,\mathrm{d}u_2 \\
&= \int_{\boldsymbol{\varphi}(U)} f(x_1,x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2
\end{aligned} \tag{50}
$$

Here, recall that

$$
\det\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = ad - bc. \tag{51}
$$

An example of integration by substitution

Most practical substitutions are given by the polar coordinate: $x = r\cos\theta, y = r\sin\theta$.

By this substitution, we have that $\sqrt{x^2 + y} = r$.

Also, the determinant of the Jacobian of the coordinate transform is given by

$$\det\left(\begin{bmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{bmatrix}\right) = \det\left(\begin{bmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{bmatrix}\right) = r\cos^2\theta - \left(-r\sin^2\right) = r. \tag{52}$$

Using the above results, we can calculate, for example,

$$\begin{aligned} \iint_{x^2+y^2\le 1} \left(1 - \sqrt{x^2+y^2}\right) \mathrm{d}x\,\mathrm{d}y &= \int_0^{2\pi}\int_0^1 (1-r)\left|\det\left(\begin{bmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{bmatrix}\right)\right| \mathrm{d}r\,\mathrm{d}\theta \\ &= \int_0^{2\pi}\int_0^1 (1-r)|r|\,\mathrm{d}r\,\mathrm{d}\theta \\ &= \int_0^{2\pi}\left[\int_0^1 \left(r - r^2\right)\mathrm{d}r\right]\mathrm{d}\theta = \int_0^{2\pi}\frac{1}{6}\,\mathrm{d}\theta = \frac{1}{3}\pi. \end{aligned} \tag{53}$$

Outline

1 **Continuous Random Variables**

- Relation among jointly continuous RVs

Note

In the following, we focus on two variable cases to make the discussion easier.
Nonetheless, the same discussion holds for general cases.

Marginal PDF (bivariable cases)

We first discuss the PDF of each RV, which helps us see the conditional distribution later, as we did in discrete cases.

Marginal PDF (bivariable cases)

We first discuss the PDF of each RV, which helps us see the conditional distribution later, as we did in discrete cases.

When we have the joint PDF of (X, Y), each of X and Y also has a PDF. To distinguish it from the joint PDF, we call each ***marginal probability density function (marginal PDF)***. We can obtain the explicit form of each by the integral as follows.

Marginal PDF (bivariable cases)

We first discuss the PDF of each RV, which helps us see the conditional distribution later, as we did in discrete cases.

When we have the joint PDF of (X, Y), each of X and Y also has a PDF. To distinguish it from the joint PDF, we call each ***marginal probability density function (marginal PDF)***. We can obtain the explicit form of each by the integral as follows.

Theorem

Suppose that (X, Y) is a bivariate continuous RV and its joint PDF is $p_{X,Y}$. Then, the ***marginal probability density functions (marginal PDFs)*** p_X *and* p_Y *are given by*

$$\begin{aligned} p_X(x) &= \int_{-\infty}^{+\infty} p_{X,Y}(x,y)\mathrm{d}y, \\ p_Y(y) &= \int_{-\infty}^{+\infty} p_{X,Y}(x,y)\mathrm{d}x. \end{aligned} \tag{54}$$

respectively.

Conditional PDF (bivariate cases)

Similar to the conditional PMF, we can consider the PDF of a RV updated by knowing the value of the other RV. The updated PDF is called the ***conditional probability distribution function (conditional PDF)***. As in the conditional PMF, the conditional PDF is proportional to the joint PDF. Since the integral of the conditional PDF on the whole real number line must be 1, the conditional PDF is defined as the conditional PDF over the marginal PDF.

Definition

Suppose that (X,Y) is a bivariate continuous RV and its joint PDF is $p_{X,Y}$. Then, for all y such that $p_Y(y) \neq 0$, the ***conditional probability distribution function (conditional PDF)*** $p_{X|Y}$ of X given $Y = y$ is defined by

$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}. \tag{55}$$

Note: if $p_Y(y) = 0$, the above fraction diverges. However, we do not care it since Y cannot such a value y.

Independence

Similar to discrete RV cases, if the conditional PDF is always the same as the marginal PDF, we say that the two RVs are ***independent***, that is, not related.

Definition (Independence of continuous RVs)

Let X and Y be RVs and assume that they have a joint PDF $p_{X,Y}$ and let their marginal PDFs be p_X and p_Y. Also, denote the conditional PDF of X given Y and that of Y given X by $p_{X|Y}$ and $p_{Y|X}$, respectively.
We say that the RVs X and Y are (mutually) ***independent*** if one of the following equivalent conditions holds

- $p_{X,Y}(x,y) = p_X(x)p_Y(y)$ for all (x,y).
- $p_{X|Y}(x|y) = p_X(x)$ for all (x,y) such that $p_Y(y) \neq 0$.
- $p_{Y|X}(y|x) = p_Y(y)$ for all (x,y) such that $p_X(x) \neq 0$.

Calculating the expectation of a function from joint PDF

When we quantify the relation between RVs, we often calculate the expectation of a function, as we do to evaluate the covariance. We can calculate it using the joint PDF as follows.

Theorem (Expectation of a function of jointly continuous RVs)

Let $(X_1, X_2, \ldots, X_m)$ *be a multivariate RV and* $p_{X_1, X_2, \ldots, X_m}$ *be the joint PDF. Let* $f : \mathbb{R}^m \to \mathbb{R}$ *be a function. The expectation of the random variable* $f(X_1, X_2, \ldots, X_m)$ *is given by*

$$\int_{\mathbb{R}^m} f(\boldsymbol{x}) p_{X_1, X_2, \ldots, X_m}(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}. \tag{56}$$

Covariance

For two RVs X and Y, the covariance $\mathrm{Cov}(X,Y)$ is defined by $\mathrm{Cov}(X,Y) := \mathbb{E}\left(X-\mu_X\right)\left(Y-\mu_Y\right)$. We can calculate it using the joint PDF.

Theorem

Let X and Y are random variables and μ_X and μ_Y be the expectation of X and Y, respectively. Suppose that $p_{X,Y}$ is a joint PDF of X and Y. Then, the covariance $\mathrm{Cov}(X,Y)$ is given by

$$\mathrm{Cov}(X,Y) = \iint_{\mathbb{R}^2} \left(x-\mu_X\right)\left(y-\mu_Y\right) p_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y. \tag{57}$$

Example: Multivariate normal distribution

Example (Multivariate normal distribution)

Let $\boldsymbol{\mu}$ be a real m-dimensional vector and $\boldsymbol{\Sigma}$ be a real $m \times m$ positive definite matrix, i.e., a $m \times m$ matrix such that $\boldsymbol{x}^\top \boldsymbol{\Sigma} \boldsymbol{x} > 0$ for any non-zero m-dimensional vector $\boldsymbol{x}$. We call the distribution of a m-tuple $(X_1, X_2, \ldots, X_m)$ of RVs a ***multivariate normal distribution*** if it has the following joint PDF $p_{X,Y}$.

$$p_{X_1,X_2,\ldots,X_m}(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^m \det(\boldsymbol{\Sigma})}} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu})\right) \tag{58}$$

Example: Multivariate normal distribution

Example (Multivariate normal distribution)

Recall that the joint PDF of a multivariate normal distribution is given as follows.

$$p_{X_1,X_2,\dots,X_m}(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^m \det(\boldsymbol{\Sigma})}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right) \tag{58}$$

For a bivariable case $m = 2$, the joint PDF is given by

$$p_{X_1,X_2}(x_1,x_2) = \frac{1}{\sqrt{(2\pi)^2 \det(\boldsymbol{\Sigma})}} \exp\left(-\frac{1}{2}\left(\begin{bmatrix} x_1 & x_2 \end{bmatrix} - \begin{bmatrix} \mu_1 & \mu_2 \end{bmatrix}\right)\begin{bmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{bmatrix}^{-1}\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}\right)\right), \tag{59}$$

where $\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$ is a 2-dimensional vector and $\boldsymbol{\Sigma} = \begin{bmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{bmatrix}$ is a real 2×2 positive definite matrix.

Example: Multivariate normal distribution

Example (Multivariate normal distribution)

Recall that the joint PDF of a multivariate normal distribution is given as follows.

$$p_{X_1,X_2,\ldots,X_m}(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^m \det(\boldsymbol{\Sigma})}} \exp\left(-\frac{1}{2}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)^\top \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)\right) \tag{58}$$

We can see that $p_{X,Y}(x,y)$ takes its maximum if $s = \boldsymbol{\mu}$ since $\left(\boldsymbol{x}-\boldsymbol{\mu}\right)^\top \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)$ is zero if $s = \boldsymbol{\mu}$ and positive otherwise, according to the positive definite assumption on $\boldsymbol{\Sigma}$.

Example: Multivariate normal distribution

Example (Multivariate normal distribution)

Recall that the joint PDF of a multivariate normal distribution is given as follows.

$$p_{X_1,X_2,\ldots,X_m}(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^m \det(\boldsymbol{\Sigma})}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right) \tag{58}$$

We can see that $p_{X,Y}(x,y)$ takes its maximum if $s = \boldsymbol{\mu}$ since $(\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$ is zero if $s = \boldsymbol{\mu}$ and positive otherwise, according to the positive definite assumption on $\boldsymbol{\Sigma}$. Unfortunately, we cannot calculate the probability $\Pr((X,Y) \in \mathscr{A})$ analytically for general $\mathscr{A}$.

Example: Multivariate normal distribution

Example (Multivariate normal distribution)

Recall that the joint PDF of a multivariate normal distribution is given as follows.

$$p_{X_1,X_2,\dots,X_m}(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^m \det(\boldsymbol{\Sigma})}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right) \tag{58}$$

We can see that $p_{X,Y}(x,y)$ takes its maximum if $s = \boldsymbol{\mu}$ since $(\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$ is zero if $s = \boldsymbol{\mu}$ and positive otherwise, according to the positive definite assumption on $\boldsymbol{\Sigma}$. Unfortunately, we cannot calculate the probability $\Pr((X,Y) \in \mathscr{A})$ analytically for general $\mathscr{A}$. Nevertheless, we can prove that the mean $\mathbb{E}X_i$ of the ith RV X_i is μ_i, the ith element of the vector $\boldsymbol{\mu}$. Also, the covariance matrix is given by $\boldsymbol{\Sigma}$. In other words, the covariance between $\mathrm{Cov}(X_i, X_j)$ is given by the entry s_{ij} in the ith row and the jth column of the matrix $\boldsymbol{\Sigma}$. In particular, the variance $\mathbb{V}(X_i) = s_{ii}$.

Example: Multivariate normal distribution

Example (Multivariate normal distribution)

Recall that the joint PDF of a multivariate normal distribution is given as follows.

$$p_{X_1,X_2,\ldots,X_m}(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^m \det(\boldsymbol{\Sigma})}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right) \tag{58}$$

Suppose that $\boldsymbol{\Sigma}$ is a diagonal matrix, i.e., $s_{ij} = 0$ if $i \neq j$. Then, $\det(\boldsymbol{\Sigma}) = \prod_{i=1}^{m} s_i$ and $-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) = -\frac{1}{2}\sum_{i=1}^{m}\frac{\left(x_i-\mu_i\right)^2}{s_{ii}}$. Therefore, we have the decomposition: $p_{X_1,X_2,\ldots,X_m}(\boldsymbol{x}) = \prod_{i=1}^{m}\frac{1}{\sqrt{2\pi s_{ii}}}\exp\left(-\frac{\left(x_i-\mu_i\right)^2}{2s_{ii}}\right)$. Hence, if $\boldsymbol{\Sigma}$ is diagonal, then $X_1, X_2, \ldots, X_m$ are mutually independent.

Outline

1 Continuous Random Variables

Exercises

Exercise (Continuous random variable)

Let X be a random variable, and let F_X be the cumulative distribution function (CDF) of X, given by

$$F_X(x) = \begin{cases} 0 & \text{if } x \leq 0, \\ \frac{1}{4}x^2 & \text{if } 0 \leq x \leq 2, \\ 1 & \text{if } x \geq 2. \end{cases}$$

(1) Evaluate the probability $\Pr(0.25 \leq X \leq 0.75)$.

(2) F_X is differentiable at all but a finite number of points, and its derivative is X's probability density function (p_X), which can be arbitrary at points of non-differentiability. Evaluate $p_X(0.5)$ and $p_X(3)$.

Example answer:

(1) From the definition of the cumulative distribution function,
$\Pr(0.25 \leq X \leq 0.75) = F_X(0.75) - F_X(x) \lim_{x \to 0.25}(x)$. Since F_X is a continuous function,
$\lim_{x \to 0.25} F_X(x) = F_X(0.25)$.

Therefore, $\Pr(0.25 \leq X \leq 0.75) = F_X(0.75) - F_X(0.25) = \frac{1}{4}0.75^2 - \frac{1}{4}0.25^2 = \frac{1}{4}$.

Example answer:

(2) Except at the two points $x = 0, 2$, on all of the real line, the derivative of F_X can be simply calculated using the formula for the derivative of a polynomial $\dfrac{d}{dx}x^n = nx^{n-1}$ as follows:

$$\begin{cases} 0 & \text{if } x < 0, \\ \frac{1}{2}x & \text{if } 0 < x < 2, \\ 0 & \text{if } x > 2. \end{cases}$$

Hence, $p_X(0.5) = \frac{1}{2} \cdot (0.5) = 0.25, p_X(3) = 0.$

Example answer:

Note: F_X is, in fact, differentiable at $x = 0$. This can be shown since the value of the left derivative $\lim_{h \nearrow 0} \frac{F_X(0+h) - F_X(0)}{h}$ matches the value of the right derivative $\lim_{h \searrow 0} \frac{F_X(0+h) - F_X(0)}{h}$, both being 0, thus the derivative $\dfrac{d}{dx} F_X(0) = 0$.

On the other hand, F_X is not differentiable at $x = 2$. This is because the value of the left derivative $\lim_{h \nearrow 0} \frac{F_X(2+h) - F_X(2)}{h}$ is 1, and the value of the right derivative $\lim_{h \searrow 0} \frac{F_X(2+h) - F_X(2)}{h}$ is 0, and the two do not match.

Exercise (Multivariate normal distribution)

(1) Define integral $K(R) = \int_0^R r \exp\left(-\frac{r^2}{2}\right) dr$. Find $K(2)$.

(2) By the change of variables $\begin{cases} x = r\cos\theta, \\ y = r\sin\theta, \end{cases}$ where $r \geq 0$, compute the absolute value of the Jacobian determinant $|\det(\frac{\partial(x,y)}{(r,\theta)})| =$ at $r = 0.5, \theta = \pi$.

Exercise (Multivariate normal distribution)

(3) For $R \geq 0$, evaluate the double integral $I(R) = \iint_{x^2+y^2 \leq R^2} \exp\left(-\frac{x^2+y^2}{2}\right) dxdy$ for $R = 2$.

(4) Evaluate the value of the improper double integral $\int_{\mathbb{R}^2} \exp\left(-\frac{x^2+y^2}{2}\right) dxdy$ as R approaches infinity, i.e., $\lim_{R \to +\infty} I(R)$.

(5) The bivariate improper integral discussed in the above (4) can be decomposed into the product of univariate improper integrals as follows:

$$\int_{\mathbb{R}^2} \exp\left(-\frac{x^2+y^2}{2}\right) dxdy = \int_{\mathbb{R}^2} \exp\left(-\frac{x^2}{2}\right) \exp\left(-\frac{y^2}{2}\right) dxdy = \left(\int_{-\infty}^{+\infty} \exp\left(-\frac{x^2}{2}\right) dx\right) \left(\int_{-\infty}^{+\infty} \exp\left(-\frac{y^2}{2}\right) dy\right) = \left(\int_{-\infty}^{+\infty} \exp\left(-\frac{x^2}{2}\right) dx\right)^2$$

Evaluate the improper integral $\int_{-\infty}^{+\infty} \exp\left(-\frac{x^2}{2}\right) dx$.

Exercise (Multivariate normal distribution)

For (6) - (10), let X, Y be random variables with the joint probability density function $p_{X,Y}$ specified by
$p_{X,Y}(x,y) = c\exp\left(-\frac{x^2+y^2}{2}\right)$,
where c is a constant.
(6) Given that $p_{X,Y}$ is a joint probability density function, determine the constant c.
(7) Calculate the probability $\Pr(X^2+Y^2 \le 2^2)$.
(8) The marginal probability density function for X, $p_X(x)$, is found by $p_X(x) = \int_{-\infty}^{+\infty} p_{X,Y}(x,y)dy$. Evaluate $p_X(-2)$.
(9) The conditional probability density function for Y given X, $p_{Y|X}(y|x)$, is calculated by $p_{Y|X}(y|x) = \frac{p_{X,Y}(x,y)}{p_X(x)}$. Evaluate $p_{Y|X}(0|-2)$.
(10) Evaluate the expected value of $(X^2+Y^2)^2$, $\mathbb{E}[(X^2+Y^2)^2]$.

Exercise (Multivariate normal distribution)

For (6) - (10), let X, Y be random variables with the joint probability density function $p_{X,Y}$ specified by

$p_{X,Y}(x,y) = c\exp\left(-\frac{x^2+y^2}{2}\right)$,

where c is a constant.

(11) Select the **ONE correct statement** from the above:

- X and Y are independent, and the covariance of X and Y is 0. (correct)
- X and Y are independent, and the covariance of X and Y is non-zero.
- X and Y are not independent, and the covariance of X and Y is 0.
- X and Y are not independent, and the covariance of X and Y is non-zero.

Example answer:

(1) $K(R)=\int_0^R r\exp\left(-\frac{r^2}{2}\right)dr$ can be calculated using a substitution of variables with $s=\frac{r^2}{2}$, leading to $\int_0^R r\exp\left(-\frac{r^2}{2}\right)dr=\int_0^{\frac{R^2}{2}}\exp(-s)\frac{ds}{dr}dr=\int_0^{\frac{R^2}{2}}\exp(-s)ds$. Since $\int_0^{\frac{R^2}{2}}\exp(-s)ds=[-\exp(-s)]_0^{R^2}=1-\exp\left(-R^2\right)$, for $R=2$, we have $K(2)=1-\exp\left(-\frac{2^2}{2}\right)=1-\frac{1}{e^2}$.

Example answer:

(2) By definition, the Jacobian is given by $\frac{\partial(x,y)}{\partial(r,\theta)} = \begin{bmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{bmatrix}$. The first column vector $\begin{bmatrix} \frac{\partial x}{\partial r} \\ \frac{\partial y}{\partial r} \end{bmatrix}$ represents the velocity vector of the (x,y) coordinates moving at unit speed in the positive direction of r with θ held fixed, and the second column vector $\begin{bmatrix} \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial \theta} \end{bmatrix}$ represents the velocity vector of the (x,y) coordinates when θ is moved at unit speed with r held fixed. Calculating the Jacobian, we find $\begin{bmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{bmatrix}$, and thus its determinant is $r(\cos^2\theta + \sin^2\theta) = r$, which is always positive, meaning the absolute value of the determinant of the Jacobian is r (independent of the value of θ). Therefore, for $r = 0.5$, $\theta = \pi$, we have $|\det(\frac{\partial(x,y)}{\partial(r,\theta)})| = 0.5$. This coordinate transformation is known as polar coordinate transformation, where r represents the distance from the point (x,y) to the origin, and θ represents the angle formed by the line segment $(0,0)$ — (x,y) with the positive direction of the x-axis.

Example answer:

(3) The formula for variable substitution (substitution integration) in double integrals is given by $\int_{A'} f(x,y)dxdy = \int_A f(x(r,\theta),y(r,\theta))|\frac{\partial(x,y)}{\partial(r,\theta)}|drd\theta$, where the right side uses notation loosely, with r and θ representing the functions for x and y values respectively, and A' and A are the regions corresponding through the variable transformation. In this problem, the region A corresponding to $x^2+y^2 \leq R^2$ translates in the (r,θ) coordinate system to a region A' satisfying $0 < r \leq R$ and $0 \leq \theta < 2\pi$, the original domain of θ. Paying attention to the calculated $|\frac{\partial(x,y)}{\partial(r,\theta)}| = r$, we can compute $\int_{x^2+y^2\leq R^2} \exp\left(-\frac{x^2+y^2}{2}\right)dxdy = \int_0^{2\pi}\int_0^R \exp\left(-\frac{r^2}{2}\right)rdrd\theta$. Evaluating the double integral by computing the integral over r first, as done in (1) where $K(R) = 1-\exp\left(-\frac{R^2}{2}\right)$, we find $I(R) = \int_0^{2\pi}\left(1-\exp\left(-\frac{R^2}{2}\right)\right)d\theta = 2\pi(1-\exp\left(-\frac{R^2}{2}\right))$. Therefore, $I(2) = 2\pi\left(1-\frac{1}{e^2}\right)$.

(4) $\int_{\mathbb{R}^2}\exp\left(-\frac{x^2+y^2}{2}\right)dxdy = \lim_{R\to+\infty} I(R) = \lim_{R\to+\infty} 2\pi\left(1-\exp\left(-\frac{R^2}{2}\right)\right) = 2\pi.$

Example answer:

(5) $2\pi = \int_{\mathbb{R}^2} \exp\left(-\frac{x^2+y^2}{2}\right) dxdy = \left(\int_{-\infty}^{+\infty} \exp\left(-\frac{x^2}{2}\right) dx\right)^2$. Since $\exp\left(-\frac{x^2}{2}\right)$ is always positive, $\int_{-\infty}^{+\infty} \exp\left(-\frac{x^2}{2}\right) dx$ is non-negative. Hence, $\int_{-\infty}^{+\infty} \exp\left(-\frac{x^2}{2}\right) dx = \sqrt{2\pi}$. It is known that the antiderivative of $\exp\left(-\frac{x^2}{2}\right)$ is not an elementary function, making it difficult to directly compute this improper integral as the limit of a definite integral. This problem approached the double integral and polar coordinate transformation, an idea dating back to Poisson in the 19th century, and this broad integral is known as the Gaussian integral or Euler-Poisson integral.

Example answer:

(6) Since $p_{X,Y}$ is the joint probability density function, $\int_{\mathbb{R}^2} p_{X,Y}(x,y)dxdy = 1$. Using the result from (4), we can compute $\int_{\mathbb{R}^2} p_{X,Y}(x,y)dxdy = \int_{\mathbb{R}^2} c \exp\left(-\frac{x^2+y^2}{2}\right)dxdy = c2\pi$. Therefore, $c = \frac{1}{2\pi}$.

(7) From the definition of the joint probability density function, $\Pr(X^2+Y^2 \leq R^2) = \int_{x^2+y^2\leq R^2} p_{X,Y}(x,y)dxdy$. Using the definition of $p_{X,Y}$ and the value of c found in (6), we have $\int_{x^2+y^2\leq R^2} p_{X,Y}(x,y)dxdy = \int_{x^2+y^2\leq R^2} \frac{1}{2\pi} \exp\left(-\frac{x^2+y^2}{2}\right)dxdy = \frac{1}{2\pi}I(R)$. Thus, $\Pr(X^2+Y^2 \leq 2^2) = \frac{1}{2\pi}I(2) = \frac{1}{2\pi}2\pi\left(1-\frac{1}{e^2}\right)$.

Example answer:

(8) The integrand $p_{X,Y}(x,y) = \frac{1}{2\pi}\exp\left(-\frac{x^2+y^2}{2}\right) = \frac{1}{2\pi}\exp\left(-\frac{x^2}{2}\right)\exp\left(-\frac{y^2}{2}\right)$ can be decomposed into functions of x and y. Utilizing this,

$p_X(x) = \int_{-\infty}^{+\infty} p_{X,Y}(x,y)dy = \int_{-\infty}^{+\infty} \frac{1}{2\pi}\exp\left(-\frac{x^2}{2}\right)\exp\left(-\frac{y^2}{2}\right)dy = \frac{1}{2\pi}\exp\left(-\frac{x^2}{2}\right)\int_{-\infty}^{+\infty}\exp\left(-\frac{y^2}{2}\right)dy.$

From (5), $\int_{-\infty}^{+\infty}\exp\left(-\frac{y^2}{2}\right)dy = \sqrt{2\pi}$ hence, $p_X(x) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)$. Therefore,

$p_X(-2) = \frac{1}{\sqrt{2\pi e^2}}$.

(9) From $p_{Y|X}(y|x) = \dfrac{p_{X,Y}(x,y)}{p_X(x)}$, we have $p_{Y|X}(0|-2) = \dfrac{\frac{1}{2\pi e^2}}{\frac{1}{\sqrt{2\pi e^2}}} = \dfrac{1}{\sqrt{2\pi}}$.

Example answer:

(10) $\mathbb{E}[(X^2+Y^2)^2] = \int_{\mathbb{R}^2} (x^2+y^2)^2 p_{X,Y}(x,y)dxdy$ can be calculated using the joint probability density function. Utilizing polar coordinate transformation similarly to (3) and (4), where $(x^2+y^2)^2 = r^4$, we compute $\int_{\mathbb{R}^2} (x^2+y^2)^2 p_{X,Y}(x,y)dxdy = \lim_{R\to+\infty} \int_0^{2\pi} L(R)d\theta$, where $L(R) = \int_0^R r^4 \frac{1}{2\pi} \exp\left(-\frac{r^2}{2}\right) \cdot rdr$. Using the substitution $s = \frac{r^2}{2}$, we find it equals $\int_0^{\frac{R^2}{2}} 4s^2 \frac{1}{2\pi} \exp(-s)ds$. Evaluating the antiderivative of $s^2\exp(-s)$ through integration by parts twice, we find $\int s^2 \exp(-s) = -s^2\exp(-s) - 2s\exp(-s) - 2\exp(-s) + C$, where C is the integration constant. Thus, $L(R) = \frac{2}{\pi}\left(2 - 2\cdot\left(\frac{R^4}{4} + 2\cdot\frac{R^2}{2} + 2\right)\exp\left(-\frac{R^2}{2}\right)\right)$. Therefore, $\int_{\mathbb{R}^2} (x^2+y^2)^2 p_{X,Y}(x,y)dxdy = \lim_{R\to+\infty} \int_0^{2\pi} L(R)d\theta = 8$.

Example answer:

(11) From the result of (8), $p_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$, and similarly, $p_Y(y) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right)$.
Thus, $p_{X,Y}(x, y) = p_X(x) p_Y(y)$, and the random variables X and Y are independent.
Therefore, the covariance and correlation coefficient between X and Y are 0.