Mathematics for Informatics 4a

José Figueroa-O'Farrill

Lecture 10
17 February 2012

Public Service Announcement

Next week it is **Innovative Learning Week**.

There are no lectures or tutorials for **mi4**.

The next tutorial will be **Tuesday 28 February**.

Continuous random variables I

Continuous random variables I

- After **discrete random variables**, it is now time to study "continuous" random variables; namely, those taking values in an uncountable set, e.g., $\mathbb{R}$

Continuous random variables I

- After **discrete random variables**, it is now time to study "continuous" random variables; namely, those taking values in an uncountable set, e.g., $\mathbb{R}$
- For example, choose at random a real number between 0 and 1. What is the probability of choosing $\frac{1}{7}$?

Continuous random variables I

- After **discrete random variables**, it is now time to study "continuous" random variables; namely, those taking values in an uncountable set, e.g., $\mathbb{R}$
- For example, choose at random a real number between 0 and 1. What is the probability of choosing $\frac{1}{7}$?
- "At random" means that every number is equally likely, so the probability of choosing $\frac{1}{7}$ is the same as that of any other number. Let's call that proability ϵ. What can ϵ be?

Continuous random variables I

- After **discrete random variables**, it is now time to study "continuous" random variables; namely, those taking values in an uncountable set, e.g., $\mathbb{R}$
- For example, choose at random a real number between 0 and 1. What is the probability of choosing $\frac{1}{7}$?
- "At random" means that every number is equally likely, so the probability of choosing $\frac{1}{7}$ is the same as that of any other number. Let's call that proability ϵ. What can ϵ be?
- We can write the certain event $[0, 1]$ as the disjoint union

$$[0,1] = \bigcup_{x \in [0,1]} \{x\}$$

Continuous random variables I

- After **discrete random variables**, it is now time to study "continuous" random variables; namely, those taking values in an uncountable set, e.g., $\mathbb{R}$
- For example, choose at random a real number between 0 and 1. What is the probability of choosing $\frac{1}{7}$?
- "At random" means that every number is equally likely, so the probability of choosing $\frac{1}{7}$ is the same as that of any other number. Let's call that proability ϵ. What can ϵ be?
- We can write the certain event $[0, 1]$ as the disjoint union

$$[0,1] = \bigcup_{x \in [0,1]} \{x\}$$

- We know that $\mathbb{P}([0, 1]) = 1$, but this is *not* a countable disjoint union.

Continuous random variables II

- So let us break up $[0,1]$ into a countable disjoint union:

$$[0,1] = A_0 \cup \bigcup_{n=1}^{\infty} \{\frac{1}{n}\}$$

where A_0 is simply the complement of $\{1, \frac{1}{2}, \frac{1}{3}, \dots\}$.

Continuous random variables II

- So let us break up $[0,1]$ into a countable disjoint union:

$$[0,1] = A_0 \cup \bigcup_{n=1}^{\infty} \{\frac{1}{n}\}$$

where A_0 is simply the complement of $\{1, \frac{1}{2}, \frac{1}{3}, \ldots\}$.

- Assuming that $\{\frac{1}{n}\}$ is an event for all n, we apply $\mathbb{P}$ to obtain

$$1 = \mathbb{P}(A_0) + \sum_{n=1}^{\infty} \epsilon$$

Continuous random variables II

- So let us break up $[0, 1]$ into a countable disjoint union:

$$[0,1] = A_0 \cup \bigcup_{n=1}^{\infty} \{\frac{1}{n}\}$$

where A_0 is simply the complement of $\{1, \frac{1}{2}, \frac{1}{3}, \ldots\}$.

- Assuming that $\{\frac{1}{n}\}$ is an event for all n, we apply $\mathbb{P}$ to obtain

$$1 = \mathbb{P}(A_0) + \sum_{n=1}^{\infty} \epsilon \qquad \Longrightarrow \quad \epsilon = 0$$

Continuous random variables II

- So let us break up $[0, 1]$ into a countable disjoint union:

$$[0, 1] = A_0 \cup \bigcup_{n=1}^{\infty} \{\frac{1}{n}\}$$

where A_0 is simply the complement of $\{1, \frac{1}{2}, \frac{1}{3}, \dots\}$.

- Assuming that $\{\frac{1}{n}\}$ is an event for all n, we apply $\mathbb{P}$ to obtain

$$1 = \mathbb{P}(A_0) + \sum_{n=1}^{\infty} \epsilon \qquad \Longrightarrow \quad \epsilon = 0$$

- This shows, by the way, why one limits the additivity of $\mathbb{P}$ to *countable* unions; otherwise one would conclude that $\mathbb{P}([0, 1]) = 0$ — a contradiction.

Continuous random variables II

- So let us break up $[0,1]$ into a countable disjoint union:

$$[0,1] = A_0 \cup \bigcup_{n=1}^{\infty} \{\frac{1}{n}\}$$

where A_0 is simply the complement of $\{1, \frac{1}{2}, \frac{1}{3}, \dots\}$.

- Assuming that $\{\frac{1}{n}\}$ is an event for all n, we apply $\mathbb{P}$ to obtain

$$1 = \mathbb{P}(A_0) + \sum_{n=1}^{\infty} \epsilon \qquad \Longrightarrow \quad \epsilon = 0$$

- This shows, by the way, why one limits the additivity of $\mathbb{P}$ to *countable* unions; otherwise one would conclude that $\mathbb{P}([0,1]) = 0$ — a contradiction.
- This argument also shows that any countable subset of $\mathbb{R}$ has zero probability: rationals, algebraic numbers,...

Continuous random variables III

Definition

A **continuous random variable** X on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a function $X : \Omega \to \mathbb{R}$ such that for all $x \in \mathbb{R}$

Continuous random variables III

Definition

A **continuous random variable** X on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a function $X : \Omega \to \mathbb{R}$ such that for all $x \in \mathbb{R}$

1. $\{X \leqslant x\}$ is an event, and

Continuous random variables III

Definition

A **continuous random variable** X on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a function $X : \Omega \to \mathbb{R}$ such that for all $x \in \mathbb{R}$

1. $\{X \leqslant x\}$ is an event, and
2. $\mathbb{P}(X = x) = 0$

Continuous random variables III

Definition

A **continuous random variable** X on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a function $X : \Omega \to \mathbb{R}$ such that for all $x \in \mathbb{R}$

1. $\{X \leqslant x\}$ is an event, and
2. $\mathbb{P}(X = x) = 0$

Remark

The definition requires $\{X = x\}$ to be an event. Let's prove it.

Continuous random variables III

Definition

A **continuous random variable** X on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a function $X : \Omega \to \mathbb{R}$ such that for all $x \in \mathbb{R}$

1. $\{X \leqslant x\}$ is an event, and
2. $\mathbb{P}(X = x) = 0$

Remark

The definition requires $\{X = x\}$ to be an event. Let's prove it. Let $B_n = \{X \leqslant x - \frac{1}{n}\}$ for $n = 1, 2, \dots$.

Continuous random variables III

Definition

A **continuous random variable** X on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a function $X : \Omega \to \mathbb{R}$ such that for all $x \in \mathbb{R}$

1. $\{X \leqslant x\}$ is an event, and
2. $\mathbb{P}(X = x) = 0$

Remark

The definition requires $\{X = x\}$ to be an event. Let's prove it. Let $B_n = \{X \leqslant x - \frac{1}{n}\}$ for $n = 1, 2, \ldots$. They are events, whence so are their union $\bigcup_{n=1}^{\infty} B_n = \{X < x\}$,

Continuous random variables III

Definition

A **continuous random variable** X on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a function $X : \Omega \to \mathbb{R}$ such that for all $x \in \mathbb{R}$

1. $\{X \leqslant x\}$ is an event, and
2. $\mathbb{P}(X = x) = 0$

Remark

The definition requires $\{X = x\}$ to be an event. Let's prove it. Let $B_n = \{X \leqslant x - \frac{1}{n}\}$ for $n = 1, 2, \ldots$. They are events, whence so are their union $\bigcup_{n=1}^{\infty} B_n = \{X < x\}$, its complement $\{X \geqslant x\}$,

Continuous random variables III

Definition

A **continuous random variable** X on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a function $X : \Omega \to \mathbb{R}$ such that for all $x \in \mathbb{R}$

1. $\{X \leqslant x\}$ is an event, and
2. $\mathbb{P}(X = x) = 0$

Remark

The definition requires $\{X = x\}$ to be an event. Let's prove it. Let $B_n = \{X \leqslant x - \frac{1}{n}\}$ for $n = 1, 2, \ldots$. They are events, whence so are their union $\bigcup_{n=1}^{\infty} B_n = \{X < x\}$, its complement $\{X \geqslant x\}$, and finally their intersection

$$\{X = x\} = \{X \geqslant x\} \cap \{X \leqslant x\}.$$

Example

The probability space modelling the motivating example of choosing a number at random in $[0, 1]$, is then the triple $(\Omega, \mathcal{F}, \mathbb{P})$, where

Example

The probability space modelling the motivating example of choosing a number at random in $[0, 1]$, is then the triple $(\Omega, \mathcal{F}, \mathbb{P})$, where

1. $\Omega = [0, 1]$;

Example

The probability space modelling the motivating example of choosing a number at random in $[0, 1]$, is then the triple $(\Omega, \mathcal{F}, \mathbb{P})$, where

1. $\Omega = [0, 1]$;
2. $\mathcal{F}$ consists of the intervals $[0, a]$ with $0 \leqslant a \leqslant 1$ together with and any other subsets they generate by iterating complementation and countable unions; and

Example

The probability space modelling the motivating example of choosing a number at random in $[0, 1]$, is then the triple $(\Omega, \mathcal{F}, \mathbb{P})$, where

1. $\Omega = [0, 1]$;
2. $\mathcal{F}$ consists of the intervals $[0, a]$ with $0 \leqslant a \leqslant 1$ together with and any other subsets they generate by iterating complementation and countable unions; and
3. $\mathbb{P}([0, a]) = a$.

Example

The probability space modelling the motivating example of choosing a number at random in $[0, 1]$, is then the triple $(\Omega, \mathcal{F}, \mathbb{P})$, where

1. $\Omega = [0, 1]$;
2. $\mathcal{F}$ consists of the intervals $[0, a]$ with $0 \leqslant a \leqslant 1$ together with and any other subsets they generate by iterating complementation and countable unions; and
3. $\mathbb{P}([0, a]) = a$.

Definition

The **distribution function** F of a continuous random variable X is the function $F(x) = \mathbb{P}(X \leqslant x)$. In the above example, $F(x) = x$.

Probability density functions

Probability density functions

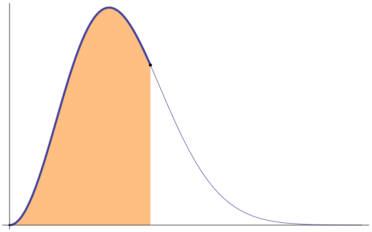
In this course we will be dealing exclusively with continuous random variables whose distribution function F is given by integrating a function f:

$$F(x) = \int_{-\infty}^{x} f(y)\,dy \,.$$

Probability density functions

In this course we will be dealing exclusively with continuous random variables whose distribution function F is given by integrating a function f:

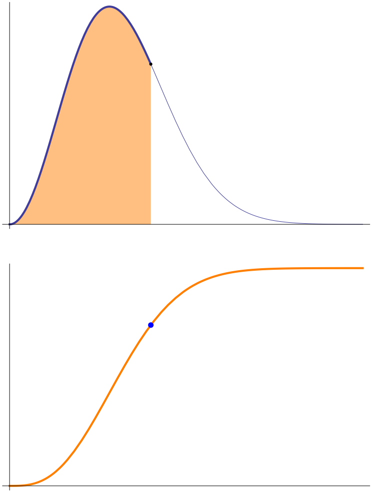
$$F(x) = \int_{-\infty}^{x} f(y)dy \ .$$



Probability density functions

In this course we will be dealing exclusively with continuous random variables whose distribution function F is given by integrating a function f:

$$F(x) = \int_{-\infty}^{x} f(y)\,dy\ .$$

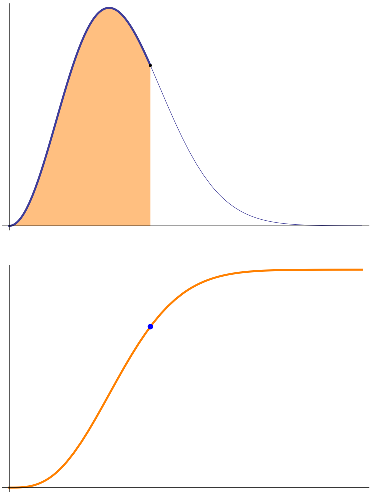


Probability density functions

In this course we will be dealing exclusively with continuous random variables whose distribution function F is given by integrating a function f:

$$F(x) = \int_{-\infty}^{x} f(y)dy \ .$$

The function f is called a "probability density function" (p.d.f.) and the function F is called a "cumulative distribution function" (c.d.f.).



PDFs and CDFs

Definition

A **probability density function** is a function $f(x) \geqslant 0$ normalised such that

$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

PDFs and CDFs

Definition

A **probability density function** is a function $f(x) \geqslant 0$ normalised such that

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

Given f, the non-decreasing function F defined by

$$F(x) = \int_{-\infty}^{x} f(y)dy$$

is called the **cumulative distribution function** of f.

PDFs and CDFs

Definition

A **probability density function** is a function $f(x) \geqslant 0$ normalised such that

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

Given f, the non-decreasing function F defined by

$$F(x) = \int_{-\infty}^{x} f(y)dy$$

is called the **cumulative distribution function** of f.

Discrete random variables have probability **mass** functions

PDFs and CDFs

Definition

A **probability density function** is a function $f(x) \geqslant 0$ normalised such that

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

Given f, the non-decreasing function F defined by

$$F(x) = \int_{-\infty}^{x} f(y)dy$$

is called the **cumulative distribution function** of f.

Discrete random variables have probability **mass** functions, but **continuous** random variables have probability **density** functions.

Continuous random variables and PDFs

As with discrete random variables, we often just say

> *Let* X *be a continuous random variable with probability density function* $f(x)$*...*

without specifying the probability space on which X is defined.

Continuous random variables and PDFs

As with discrete random variables, we often just say

> *Let X be a continuous random variable with probability density function $f(x)$...*

without specifying the probability space on which X is defined. The basic property of the probability density function for a continuous random variable X is that

$$\mathbb{P}(X \in A) = \int_{x \in A} f(x)dx$$

assuming that $\{X \in A\}$ is an event.

Continuous random variables and PDFs

As with discrete random variables, we often just say

> *Let* X *be a continuous random variable with probability density function* $f(x)$*...*

without specifying the probability space on which X is defined.
The basic property of the probability density function for a continuous random variable X is that

$$\mathbb{P}(X \in A) = \int_{x \in A} f(x) dx$$

assuming that $\{X \in A\}$ is an event.
This prompts the following

Question

For which subsets $A \in \mathbb{R}$ is $\{X \in A\}$ an event?

Borel sets

Such subsets are called **Borel sets**.

Borel sets

Such subsets are called **Borel sets**.

- By definition, $(-\infty, x]$ is a Borel set for all $x \in \mathbb{R}$.

Borel sets

Such subsets are called **Borel sets**.

- By definition, $(-\infty, x]$ is a Borel set for all $x \in \mathbb{R}$.
- So are $(-\infty, x) = \bigcup_{n=1}^{\infty} (-\infty, x - \frac{1}{n}]$.

Borel sets

Such subsets are called **Borel sets**.

- By definition, $(-\infty, x]$ is a Borel set for all $x \in \mathbb{R}$.
- So are $(-\infty, x) = \bigcup_{n=1}^{\infty}(-\infty, x - \frac{1}{n}]$.
- By complementation, so are (x, ∞) and $[x, \infty)$

Borel sets

Such subsets are called **Borel sets**.

- By definition, $(-\infty, x]$ is a Borel set for all $x \in \mathbb{R}$.
- So are $(-\infty, x) = \bigcup_{n=1}^{\infty} (-\infty, x - \frac{1}{n}]$.
- By complementation, so are (x, ∞) and $[x, \infty)$
- By intersection, $[x, y] = (-\infty, y] \cap [x, \infty)$

Borel sets

Such subsets are called **Borel sets**.

- By definition, $(-\infty, x]$ is a Borel set for all $x \in \mathbb{R}$.
- So are $(-\infty, x) = \bigcup_{n=1}^{\infty} (-\infty, x - \frac{1}{n}]$.
- By complementation, so are (x, ∞) and $[x, \infty)$
- By intersection, $[x, y] = (-\infty, y] \cap [x, \infty)$
- and similarly (x, y), $[x, y)$, $(x, y]$,...

Borel sets

Such subsets are called **Borel sets**.

- By definition, $(-\infty, x]$ is a Borel set for all $x \in \mathbb{R}$.
- So are $(-\infty, x) = \bigcup_{n=1}^{\infty} (-\infty, x - \frac{1}{n}]$.
- By complementation, so are (x, ∞) and $[x, \infty)$
- By intersection, $[x, y] = (-\infty, y] \cap [x, \infty)$
- and similarly (x, y), $[x, y)$, $(x, y]$,...
- The Borel sets are the smallest σ-field containing the intervals.

Borel sets

Such subsets are called **Borel sets**.

- By definition, $(-\infty, x]$ is a Borel set for all $x \in \mathbb{R}$.
- So are $(-\infty, x) = \bigcup_{n=1}^{\infty}(-\infty, x - \frac{1}{n}]$.
- By complementation, so are (x, ∞) and $[x, \infty)$
- By intersection, $[x, y] = (-\infty, y] \cap [x, \infty)$
- and similarly (x, y), $[x, y)$, $(x, y]$,...
- The Borel sets are the smallest σ-field containing the intervals.
- In fact, all subsets of $\mathbb{R}$ you will ever be likely to meet are Borel.

Properties of cumulative distribution functions

Let f be a probability density function with cumulative distribution function F.

Properties of cumulative distribution functions

Let f be a probability density function with cumulative distribution function F. Remember that

$$F(x) = \int_{-\infty}^{x} f(y)\,dy \qquad f(x) \geqslant 0 \,.$$

Properties of cumulative distribution functions

Let f be a probability density function with cumulative distribution function F. Remember that

$$F(x) = \int_{-\infty}^{x} f(y)\,dy \qquad f(x) \geqslant 0\,.$$

Then F satisfies the following properties:

Properties of cumulative distribution functions

Let f be a probability density function with cumulative distribution function F. Remember that

$$F(x) = \int_{-\infty}^{x} f(y)\,dy \qquad f(x) \geqslant 0\,.$$

Then F satisfies the following properties:

- $F(-\infty) = 0$ and $F(\infty) = 1$

Properties of cumulative distribution functions

Let f be a probability density function with cumulative distribution function F. Remember that

$$F(x) = \int_{-\infty}^{x} f(y)\,dy \qquad f(x) \geqslant 0 \,.$$

Then F satisfies the following properties:

- $F(-\infty) = 0$ and $F(\infty) = 1$
- if $x \geqslant y$, then $F(x) \geqslant F(y)$

Properties of cumulative distribution functions

Let f be a probability density function with cumulative distribution function F. Remember that

$$F(x) = \int_{-\infty}^{x} f(y)\,dy \qquad f(x) \geqslant 0 \,.$$

Then F satisfies the following properties:

- $F(-\infty) = 0$ and $F(\infty) = 1$
- if $x \geqslant y$, then $F(x) \geqslant F(y)$
- $F'(x) = f(x)$

Properties of cumulative distribution functions

Let f be a probability density function with cumulative distribution function F. Remember that

$$F(x) = \int_{-\infty}^{x} f(y)\,dy \qquad f(x) \geqslant 0\,.$$

Then F satisfies the following properties:

- $F(-\infty) = 0$ and $F(\infty) = 1$
- if $x \geqslant y$, then $F(x) \geqslant F(y)$
- $F'(x) = f(x)$
- $F(b) - F(a) = \int_a^b f(x)\,dx$

Example (The uniform distribution)

The p.d.f. of the **uniform** distribution on $[a, b]$ is given by

Example (The uniform distribution)

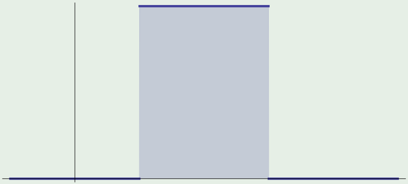
The p.d.f. of the **uniform** distribution on $[a, b]$ is given by

$$f(x) = \begin{cases} 0, & x < a \\ \frac{1}{b-a}, & a \leqslant x \leqslant b \\ 0, & x > b \end{cases}$$

Example (The uniform distribution)

The p.d.f. of the **uniform** distribution on $[a, b]$ is given by

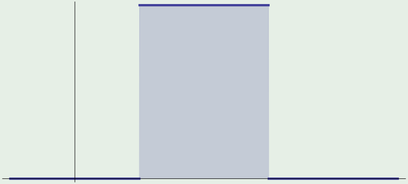
$$f(x) = \begin{cases} 0, & x < a \\ \frac{1}{b-a}, & a \leqslant x \leqslant b \\ 0, & x > b \end{cases}$$



Example (The uniform distribution)

The p.d.f. of the **uniform** distribution on $[a, b]$ is given by

$$f(x) = \begin{cases} 0, & x < a \\ \frac{1}{b-a}, & a \leqslant x \leqslant b \\ 0, & x > b \end{cases}$$



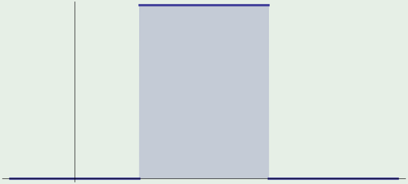
and the c.d.f. is given by

$$F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leqslant x \leqslant b \\ 1, & x > b \end{cases}$$

Example (The uniform distribution)

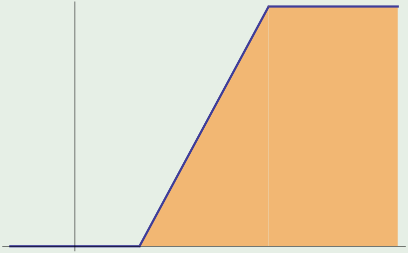
The p.d.f. of the **uniform** distribution on $[a, b]$ is given by

$$f(x) = \begin{cases} 0, & x < a \\ \frac{1}{b-a}, & a \leqslant x \leqslant b \\ 0, & x > b \end{cases}$$



and the c.d.f. is given by

$$F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leqslant x \leqslant b \\ 1, & x > b \end{cases}$$



Example (Waiting for the bus)

Between 4pm and 5pm, buses arrive at your stop at 4pm and then every 15 minutes until 5pm. You arrive at the stop at a random time between 4pm and 5pm. *What is the probability that you will have to wait at least 5 minutes for the bus?*

Example (Waiting for the bus)

Between 4pm and 5pm, buses arrive at your stop at 4pm and then every 15 minutes until 5pm. You arrive at the stop at a random time between 4pm and 5pm. *What is the probability that you will have to wait at least 5 minutes for the bus?*

- Your arrival time at the stop is uniformly distributed between 4pm and 5pm.

Example (Waiting for the bus)

Between 4pm and 5pm, buses arrive at your stop at 4pm and then every 15 minutes until 5pm. You arrive at the stop at a random time between 4pm and 5pm. *What is the probability that you will have to wait at least 5 minutes for the bus?*

- Your arrival time at the stop is uniformly distributed between 4pm and 5pm.
- You will have to wait at least 5 minutes if you arrive between the time a bus arrives and 10 minutes after that.

Example (Waiting for the bus)

Between 4pm and 5pm, buses arrive at your stop at 4pm and then every 15 minutes until 5pm. You arrive at the stop at a random time between 4pm and 5pm. *What is the probability that you will have to wait at least 5 minutes for the bus?*

- Your arrival time at the stop is uniformly distributed between 4pm and 5pm.
- You will have to wait at least 5 minutes if you arrive between the time a bus arrives and 10 minutes after that.
- That's 10 minutes in every 15 minutes, so the probability is $\frac{10}{15} = \frac{2}{3}$.

Example (The standard normal distribution)

The p.d.f. of the **standard normal** distribution is

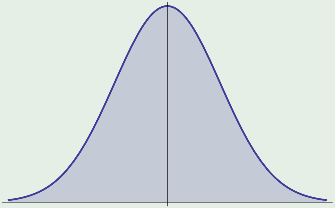
$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

Example (The standard normal distribution)

The p.d.f. of the **standard normal** distribution is

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

It is also called a **gaussian** distribution.



Example (The standard normal distribution – continued)

The proof that $f(x)$ is a probability density function follows from a standard trick.

Example (The standard normal distribution – continued)

The proof that $f(x)$ is a probability density function follows from a standard trick. Let

$$I = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} dx$$

which we have to show to be equal to 1.

Example (The standard normal distribution – continued)

The proof that $f(x)$ is a probability density function follows from a standard trick. Let

$$I = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} dx$$

which we have to show to be equal to 1. We compute I^2:

$$\begin{aligned} I^2 &= \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} dx \right)^2 \\ &= \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} dx \right) \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2} dy \right) \\ &= \frac{1}{2\pi} \iint e^{-(x^2+y^2)/2} dx\, dy \end{aligned}$$

where the integral is over the whole (x, y)-plane.

Example (The standard normal distribution – continued)

The proof that $f(x)$ is a probability density function follows from a standard trick. Let

$$I = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} dx$$

which we have to show to be equal to 1. We compute I^2:

$$\begin{aligned} I^2 &= \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} dx \right)^2 \\ &= \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} dx \right) \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2} dy \right) \\ &= \frac{1}{2\pi} \iint e^{-(x^2+y^2)/2} dx\, dy \end{aligned}$$

where the integral is over the whole (x, y)-plane. We now change to polar coordinates.

Example (The standard normal distribution – continued)

$$x = r\cos\theta \qquad y = r\sin\theta \qquad r > 0 \qquad 0 \leqslant \theta < 2\pi$$

Example (The standard normal distribution – continued)

$$x = r\cos\theta \qquad y = r\sin\theta \qquad r > 0 \qquad 0 \leqslant \theta < 2\pi$$

so that

$$x^2 + y^2 = r^2 \qquad dx\,dy = r\,dr\,d\theta$$

Example (The standard normal distribution – continued)

$$x = r\cos\theta \qquad y = r\sin\theta \qquad r > 0 \qquad 0 \leqslant \theta < 2\pi$$

so that

$$x^2 + y^2 = r^2 \qquad dx\,dy = r\,dr\,d\theta$$

Into I^2,

$$\begin{aligned} I^2 &= \frac{1}{2\pi}\iint e^{-(x^2+y^2)/2} dx\,dy \\ &= \frac{1}{2\pi}\int_{\theta=0}^{2\pi}\int_{r=0}^{\infty} e^{-r^2/2} r\,dr\,d\theta \\ &= \int_0^\infty e^{-r^2/2} d(\tfrac{1}{2}r^2) \\ &= \int_0^\infty e^{-u} du = 1 \end{aligned}$$

Example (The normal distribution)

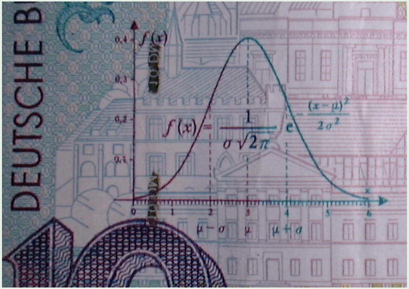
The **normal** distribution with parameters μ and σ^2 has probability density function

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Example (The normal distribution)

The **normal** distribution with parameters μ and σ^2 has probability density function

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



The standard normal distribution has $\mu = 0$ and $\sigma = 1$. We will see that μ and σ^2 are the mean and variance, respectively.

Example (The error function)

The cumulative distribution function of the normal distribution is

$$F(x) = \tfrac{1}{2} + \tfrac{1}{2}\,\mathrm{erf}\left(\frac{x-\mu}{\sqrt{2}\sigma}\right)$$

Example (The error function)

The cumulative distribution function of the normal distribution is

$$F(x) = \tfrac{1}{2} + \tfrac{1}{2}\,\mathrm{erf}\left(\frac{x-\mu}{\sqrt{2}\sigma}\right)$$

where erf is the **error function**, defined by

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\,dt$$

Example (The exponential distribution)

The p.d.f. of the **exponential** distribution with parameter λ is

Example (The exponential distribution)

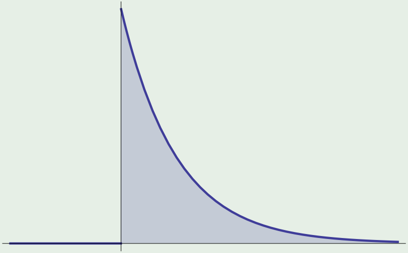
The p.d.f. of the **exponential** distribution with parameter λ is

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geqslant 0 \\ 0, & x < 0 \end{cases}$$

Example (The exponential distribution)

The p.d.f. of the **exponential** distribution with parameter λ is

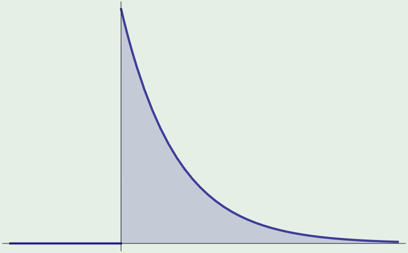
$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geqslant 0 \\ 0, & x < 0 \end{cases}$$



Example (The exponential distribution)

The p.d.f. of the **exponential** distribution with parameter λ is

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geqslant 0 \\ 0, & x < 0 \end{cases}$$



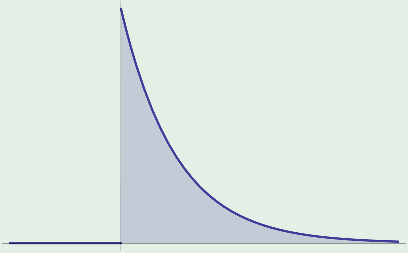
and the c.d.f. is given by

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geqslant 0 \\ 0, & x < 0 \end{cases}$$

Example (The exponential distribution)

The p.d.f. of the **exponential** distribution with parameter λ is

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geqslant 0 \\ 0, & x < 0 \end{cases}$$



and the c.d.f. is given by

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geqslant 0 \\ 0, & x < 0 \end{cases}$$

The exponential distribution has no memory

Let X be exponentially distributed with parameter λ.

The exponential distribution has no memory

Let X be exponentially distributed with parameter λ.Then $\mathbb{P}(X \leqslant x) = F(x) = 1 - e^{-\lambda x}$ for $x \geqslant 0$, whence

$$\mathbb{P}(X > x) = 1 - \mathbb{P}(X \leqslant x) = e^{-\lambda x}$$

The exponential distribution has no memory

Let X be exponentially distributed with parameter λ.Then $\mathbb{P}(X \leqslant x) = F(x) = 1 - e^{-\lambda x}$ for $x \geqslant 0$, whence

$$\mathbb{P}(X > x) = 1 - \mathbb{P}(X \leqslant x) = e^{-\lambda x}$$

If $x, y \geqslant 0$, from $e^{-\lambda(x+y)} = e^{-\lambda x} e^{-\lambda y}$ we have that

$$\mathbb{P}(X > x + y) = \mathbb{P}(X > x)\mathbb{P}(X > y)$$

The exponential distribution has no memory

Let X be exponentially distributed with parameter λ.Then $\mathbb{P}(X \leqslant x) = F(x) = 1 - e^{-\lambda x}$ for $x \geqslant 0$, whence

$$\mathbb{P}(X > x) = 1 - \mathbb{P}(X \leqslant x) = e^{-\lambda x}$$

If $x, y \geqslant 0$, from $e^{-\lambda(x+y)} = e^{-\lambda x} e^{-\lambda y}$ we have that

$$\mathbb{P}(X > x + y) = \mathbb{P}(X > x)\mathbb{P}(X > y)$$

or equivalently, partitioning $\Omega = \{X > x\} \cup \{X \leqslant x\}$,

$$\begin{aligned}\mathbb{P}(X > x + y) &= \mathbb{P}(X > x + y \mid X > x)\mathbb{P}(X > x) \\ &\quad + \mathbb{P}(X > x + y \mid X \leqslant x)\mathbb{P}(X \leqslant x)\end{aligned}$$

The exponential distribution has no memory

Let X be exponentially distributed with parameter λ.Then $\mathbb{P}(X \leqslant x) = F(x) = 1 - e^{-\lambda x}$ for $x \geqslant 0$, whence

$$\mathbb{P}(X > x) = 1 - \mathbb{P}(X \leqslant x) = e^{-\lambda x}$$

If $x, y \geqslant 0$, from $e^{-\lambda(x+y)} = e^{-\lambda x}e^{-\lambda y}$ we have that

$$\mathbb{P}(X > x + y) = \mathbb{P}(X > x)\mathbb{P}(X > y)$$

or equivalently, partitioning $\Omega = \{X > x\} \cup \{X \leqslant x\}$,

$$\begin{aligned}\mathbb{P}(X > x + y) &= \mathbb{P}(X > x + y \mid X > x)\mathbb{P}(X > x) \\ &\quad + \mathbb{P}(X > x + y \mid X \leqslant x)\mathbb{P}(X \leqslant x)\end{aligned}$$

whence cancelling the $\mathbb{P}(X > x)$ from both sides,

$$\mathbb{P}(X > x + y \mid X > x) = \mathbb{P}(X > y)$$

Example

- Let X denote the time it takes for a computer programme to crash

Example

- Let X denote the time it takes for a computer programme to crash
- It is sensible to assume that X is exponentially distributed

Example

- Let X denote the time it takes for a computer programme to crash
- It is sensible to assume that X is exponentially distributed
- The conditional probability of it not crashing after a time $x+y$ given that it didn't crash after a time x is $\mathbb{P}(X > x+y \mid X > x)$

Example

- Let X denote the time it takes for a computer programme to crash
- It is sensible to assume that X is exponentially distributed
- The conditional probability of it not crashing after a time $x + y$ given that it didn't crash after a time x is $\mathbb{P}(X > x + y \mid X > x)$
- which was shown to equal $\mathbb{P}(X > y)$

Example

- Let X denote the time it takes for a computer programme to crash
- It is sensible to assume that X is exponentially distributed
- The conditional probability of it not crashing after a time $x + y$ given that it didn't crash after a time x is $\mathbb{P}(X > x + y \mid X > x)$
- which was shown to equal $\mathbb{P}(X > y)$
- which is the probability of not crashing after a time y

Example

- Let X denote the time it takes for a computer programme to crash
- It is sensible to assume that X is exponentially distributed
- The conditional probability of it not crashing after a time $x+y$ given that it didn't crash after a time x is $\mathbb{P}(X > x+y \mid X > x)$
- which was shown to equal $\mathbb{P}(X > y)$
- which is the probability of not crashing after a time y
- so the fact that it didn't crash for a time x is of no relevance

Example

- Let X denote the time it takes for a computer programme to crash
- It is sensible to assume that X is exponentially distributed
- The conditional probability of it not crashing after a time $x + y$ given that it didn't crash after a time x is $\mathbb{P}(X > x + y \mid X > x)$
- which was shown to equal $\mathbb{P}(X > y)$
- which is the probability of not crashing after a time y
- so the fact that it didn't crash for a time x is of no relevance
- i.e., the exponential distribution simply does not remember that fact.

Expectations

Let X be a continuous random variable with probability density function $f(x)$.

Expectations

Let X be a continuous random variable with probability density function $f(x)$. Then we define its **expectation** (or **mean**) by

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f(x) dx$$

(provided the integral exists)

Expectations

Let X be a continuous random variable with probability density function $f(x)$. Then we define its **expectation** (or **mean**) by

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f(x) dx$$

(provided the integral exists)

Notice that if f is **symmetric**, so that $f(-x) = f(x)$, then $\mathbb{E}(X) = 0$.

Expectations

Let X be a continuous random variable with probability density function $f(x)$. Then we define its **expectation** (or **mean**) by

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f(x) dx$$

(provided the integral exists)

Notice that if f is **symmetric**, so that $f(-x) = f(x)$, then $\mathbb{E}(X) = 0$.

Example (The mean of the exponential distribution)

Let X be exponentially distributed with parameter λ.

Expectations

Let X be a continuous random variable with probability density function $f(x)$. Then we define its **expectation** (or **mean**) by

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f(x)\,dx$$

(provided the integral exists)

Notice that if f is **symmetric**, so that $f(-x) = f(x)$, then $\mathbb{E}(X) = 0$.

Example (The mean of the exponential distribution)

Let X be exponentially distributed with parameter λ. Then

$$\begin{aligned}\mathbb{E}(X) &= \int_0^{\infty} x\lambda e^{-\lambda x}\,dx = \lambda \int_0^{\infty} x e^{-\lambda x}\,dx \\ &= -\lambda \frac{d}{d\lambda} \int_0^{\infty} e^{-\lambda x}\,dx = -\lambda \frac{d}{d\lambda} \frac{1}{\lambda} = \frac{1}{\lambda}\end{aligned}$$

Example (The mean of the uniform distribution)

Let X be uniformly distributed in $[a, b]$.

Example (The mean of the uniform distribution)

Let X be uniformly distributed in $[a, b]$. Then

$$\mathbb{E}(X) = \int_a^b \frac{x}{b-a} dx = \frac{1}{2(b-a)} x^2 \Big|_a^b = \frac{b^2 - a^2}{2(b-a)} = \frac{b+a}{2}$$

Example (The mean of the uniform distribution)

Let X be uniformly distributed in $[a, b]$. Then

$$\mathbb{E}(X) = \int_a^b \frac{x}{b-a} dx = \tfrac{1}{2(b-a)}\, x^2 \Big|_a^b = \frac{b^2 - a^2}{2(b-a)} = \frac{b+a}{2}$$

In other words, the mean is the midpoint in the interval.

Example (The mean of the uniform distribution)

Let X be uniformly distributed in $[a, b]$. Then

$$\mathbb{E}(X) = \int_a^b \frac{x}{b-a} dx = \tfrac{1}{2(b-a)} \; x^2 \Big|_a^b = \frac{b^2 - a^2}{2(b-a)} = \frac{b+a}{2}$$

In other words, the mean is the midpoint in the interval.

Example (Waiting for the bus – continued)

In the example about ▸ waiting for the bus, *what is your expected waiting time?*

Example (The mean of the uniform distribution)

Let X be uniformly distributed in $[a,b]$. Then

$$\mathbb{E}(X) = \int_a^b \frac{x}{b-a} dx = \frac{1}{2(b-a)} x^2 \Big|_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{b+a}{2}$$

In other words, the mean is the midpoint in the interval.

Example (Waiting for the bus – continued)

In the example about waiting for the bus, *what is your expected waiting time?*

Your expected arrival time is uniformly distributed, but you are interested in the expectatation of the waiting time.

Example (The mean of the uniform distribution)

Let X be uniformly distributed in $[a, b]$. Then

$$\mathbb{E}(X) = \int_a^b \frac{x}{b-a} dx = \frac{1}{2(b-a)} x^2 \Big|_a^b = \frac{b^2 - a^2}{2(b-a)} = \frac{b+a}{2}$$

In other words, the mean is the midpoint in the interval.

Example (Waiting for the bus – continued)

In the example about ▸ waiting for the bus, *what is your expected waiting time?*

Your expected arrival time is uniformly distributed, but you are interested in the expectatation of the waiting time. Because of the periodicity of the buses, it is enough to consider a 15-minute interval: 4pm-4:15pm, say.

Example (The mean of the uniform distribution)

Let X be uniformly distributed in $[a, b]$. Then

$$\mathbb{E}(X) = \int_a^b \frac{x}{b-a} dx = \tfrac{1}{2(b-a)}\, x^2\Big|_a^b = \frac{b^2 - a^2}{2(b-a)} = \frac{b+a}{2}$$

In other words, the mean is the midpoint in the interval.

Example (Waiting for the bus – continued)

In the example about ▸ waiting for the bus, *what is your expected waiting time?*

Your expected arrival time is uniformly distributed, but you are interested in the expectatation of the waiting time. Because of the periodicity of the buses, it is enough to consider a 15-minute interval: 4pm-4:15pm, say. Then the waiting time is the same as the arrival time, and hence the expectation is the midpoint of the interval, so $7\frac{1}{2}$ minutes.

Example (The mean of the normal distribution)

Let X be normally distributed with parameters μ and σ^2.

Example (The mean of the normal distribution)

Let X be normally distributed with parameters μ and σ^2. Then

$$\mathbb{E}(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-(x-\mu)^2/2\sigma^2} dx$$

Example (The mean of the normal distribution)

Let X be normally distributed with parameters μ and σ^2. Then

$$\mathbb{E}(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-(x-\mu)^2/2\sigma^2} dx$$

We change coordinates to $y = x - \mu$, so that

$$\begin{aligned}
\mathbb{E}(X) &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (y+\mu) e^{-y^2/2\sigma^2} dy \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} y e^{-y^2/2\sigma^2} dy + \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \mu e^{-y^2/2\sigma^2} dy \\
&= 0 + \mu = \mu
\end{aligned}$$

Example (The mean of the normal distribution)

Let X be normally distributed with parameters μ and σ^2. Then

$$\mathbb{E}(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-(x-\mu)^2/2\sigma^2} dx$$

We change coordinates to $y = x - \mu$, so that

$$\begin{aligned}
\mathbb{E}(X) &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (y+\mu) e^{-y^2/2\sigma^2} dy \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} y e^{-y^2/2\sigma^2} dy + \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \mu e^{-y^2/2\sigma^2} dy \\
&= 0 + \mu = \mu
\end{aligned}$$

where the first term vanishes because of symmetric integration and the second equals μ by using the normalisation of the normal probability density function.

Summary

- X a **continuous random variable**: for all x, $\{X \leqslant x\}$ is an event and $\mathbb{P}(X = x) = 0$

Summary

- X a **continuous random variable**: for all x, $\{X \leqslant x\}$ is an event and $\mathbb{P}(X = x) = 0$
- Continuous random variables have continuous **distribution functions** $F(x) = \mathbb{P}(X \leqslant x)$

Summary

- X a **continuous random variable**: for all x, $\{X \leqslant x\}$ is an event and $\mathbb{P}(X = x) = 0$
- Continuous random variables have continuous **distribution functions** $F(x) = \mathbb{P}(X \leqslant x)$
- F is often defined by a **probability density function** f:

$$F(x) = \int_{-\infty}^{x} f(y)dy \qquad f(x) \geqslant 0 \qquad \int_{-\infty}^{\infty} f(x)dx = 1$$

and is called a **cumulative distribution function**

Summary

- X a **continuous random variable**: for all x, $\{X \leqslant x\}$ is an event and $\mathbb{P}(X = x) = 0$
- Continuous random variables have continuous **distribution functions** $F(x) = \mathbb{P}(X \leqslant x)$
- F is often defined by a **probability density function** f:

$$F(x) = \int_{-\infty}^{x} f(y)dy \qquad f(x) \geqslant 0 \qquad \int_{-\infty}^{\infty} f(x)dx = 1$$

  and is called a **cumulative distribution function**
- We have met several probability density functions:
  - **uniform**: $f(x) = \frac{1}{b-a}$ for $x \in [a, b]$

Summary

- X a **continuous random variable**: for all x, $\{X \leqslant x\}$ is an event and $\mathbb{P}(X = x) = 0$
- Continuous random variables have continuous **distribution functions** $F(x) = \mathbb{P}(X \leqslant x)$
- F is often defined by a **probability density function** f:

$$F(x) = \int_{-\infty}^{x} f(y)dy \qquad f(x) \geqslant 0 \qquad \int_{-\infty}^{\infty} f(x)dx = 1$$

  and is called a **cumulative distribution function**
- We have met several probability density functions:
  - **uniform**: $f(x) = \frac{1}{b-a}$ for $x \in [a, b]$
  - **normal**: $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$

Summary

- X a **continuous random variable**: for all x, $\{X \leqslant x\}$ is an event and $\mathbb{P}(X = x) = 0$
- Continuous random variables have continuous **distribution functions** $F(x) = \mathbb{P}(X \leqslant x)$
- F is often defined by a **probability density function** f:

$$F(x) = \int_{-\infty}^{x} f(y)dy \qquad f(x) \geqslant 0 \qquad \int_{-\infty}^{\infty} f(x)dx = 1$$

  and is called a **cumulative distribution function**
- We have met several probability density functions:
  - **uniform**: $f(x) = \frac{1}{b-a}$ for $x \in [a, b]$
  - **normal**: $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$
  - **exponential**: $f(x) = \lambda e^{-\lambda x}$ for $x \geqslant 0$

Summary

- X a **continuous random variable**: for all x, $\{X \leqslant x\}$ is an event and $\mathbb{P}(X = x) = 0$
- Continuous random variables have continuous **distribution functions** $F(x) = \mathbb{P}(X \leqslant x)$
- F is often defined by a **probability density function** f:

$$F(x) = \int_{-\infty}^{x} f(y)dy \qquad f(x) \geqslant 0 \qquad \int_{-\infty}^{\infty} f(x)dx = 1$$

  and is called a **cumulative distribution function**
- We have met several probability density functions:
  - **uniform**: $f(x) = \frac{1}{b-a}$ for $x \in [a, b]$
  - **normal**: $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$
  - **exponential**: $f(x) = \lambda e^{-\lambda x}$ for $x \geqslant 0$ (has no memory!)

Summary

- X a **continuous random variable**: for all x, $\{X \leqslant x\}$ is an event and $\mathbb{P}(X = x) = 0$
- Continuous random variables have continuous **distribution functions** $F(x) = \mathbb{P}(X \leqslant x)$
- F is often defined by a **probability density function** f:

$$F(x) = \int_{-\infty}^{x} f(y)dy \qquad f(x) \geqslant 0 \qquad \int_{-\infty}^{\infty} f(x)dx = 1$$

  and is called a **cumulative distribution function**
- We have met several probability density functions:
  - **uniform**: $f(x) = \frac{1}{b-a}$ for $x \in [a, b]$
  - **normal**: $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$
  - **exponential**: $f(x) = \lambda e^{-\lambda x}$ for $x \geqslant 0$ (has no memory!)
- The **mean** $\mu = \int_{-\infty}^{\infty} x f(x) dx$

Summary

- X a **continuous random variable**: for all x, $\{X \leqslant x\}$ is an event and $\mathbb{P}(X = x) = 0$
- Continuous random variables have continuous **distribution functions** $F(x) = \mathbb{P}(X \leqslant x)$
- F is often defined by a **probability density function** f:

$$F(x) = \int_{-\infty}^{x} f(y)dy \qquad f(x) \geqslant 0 \qquad \int_{-\infty}^{\infty} f(x)dx = 1$$

  and is called a **cumulative distribution function**
- We have met several probability density functions:
  - **uniform**: $f(x) = \frac{1}{b-a}$ for $x \in [a, b]$
  - **normal**: $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$
  - **exponential**: $f(x) = \lambda e^{-\lambda x}$ for $x \geqslant 0$ (has no memory!)
- The **mean** $\mu = \int_{-\infty}^{\infty} x f(x) dx$, and equals $\frac{a+b}{2}$, μ and $\frac{1}{\lambda}$ for the above p.d.f.s, respectively.