

Week 4: Continuous Random Variables

Table of contents

- 1 Review of the Last Lesson
- 2 Continuous Random Variables
 - Definitions and Properties
 - Uniform Distribution
 - Normal Distribution
 - Exponential Distribution

Quiz 1

- Week 4 during tutorial
- 50 minutes in total
- 10 MCQ
- Weight: 10%
- Contents tested: Week 1-3 Lecture Notes
- Calculators are allowed.

Academic Integrity

- Academic dishonesty in any form will **not be tolerated**.
- Minimal penalty: **zero** for the quiz.
Other (possible) penalties: a **failing grade** or **expulsion** from DigiPen.
 - First half: Tutorial 4
 - Second half: Quiz 1

Discrete Random Variables

- $X : \Omega \rightarrow \mathbb{R}$ is **discrete** if its set of possible values is **countable** (there is an order to list out all its elements).
- The **probability mass function (PMF)** of X is

$$p(x) = P(X = x), \quad x \in \mathbb{R}.$$

- The **cumulative distribution function (CDF)** of X is

$$F(x) = P(X \leq x), \quad x \in \mathbb{R}.$$

Bernoulli(p)

- $X = \#$ successes in a Bernoulli trial with $P(\text{success}) = p$.
- $X \sim \text{Bernoulli}(p)$.
- Possible values of X are 0 and 1.
- The PMF of X is

$$p(1) = P(X = 1) = p,$$

$$p(0) = P(X = 0) = 1 - p.$$

Binomial Distribution $\text{Binomial}(n, p)$

- $X = \#$ successes in n independent Bernoulli trials with $P(\text{success}) = p$.
- $X \sim \text{Binomial}(n, p)$.
- Possible values of X are $0, 1, \dots, n$.
- The PMF of X is

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x \in \{0, 1, \dots, n\}.$$

Geometric Distribution $\text{Geom}(p)$

- Consider a sequence of Bernoulli trials with $P(\text{success}) = p$.
- $X = \#$ Bernoulli trials needed to get the first success.
- $X \sim \text{Geom}(p)$.
- The set of possible values for X is \mathbb{Z}^+ .
- The PMF of X is

$$p(x) = (1 - p)^{x-1}p, \quad x \in \mathbb{Z}^+$$

Poisson Distribution

- The **Poisson distribution with parameter** λ ($\lambda > 0$) has PMF

$$p(x) = P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda}, \quad x = 0, 1, 2, \dots$$

- Write $X \sim \text{Poisson}(\lambda)$.
- Possible values of X are $0, 1, 2, \dots$

Poisson Distribution vs Binomial Distribution

- Poisson distribution \approx binomial distribution in the case **n is large** and **p is small**.
- Recall: $X \sim \text{Binomial}(n, p) \Rightarrow P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}$.
When n is large and p is small
 - it is hard to compute exactly $\binom{n}{x} p^x (1 - p)^{n-x}$,
 - it is difficult to analyze $\text{Binomial}(n, p)$ for large n .

Poisson Distribution Approximates Binomial Distribution

Lemma 1. If $X \sim \text{Binomial}(n, p)$ and $\lambda = np$ is a constant, then its PMF is *asymptotically equal* to the PMF of $\text{Poisson}(\lambda)$.

Precisely, this means if $\lambda = np$ is a constant, then

$$\lim_{n \rightarrow \infty} \binom{n}{x} p^x (1-p)^{n-x} \div \left(\frac{\lambda^x}{x!} e^{-\lambda} \right) = 1.$$

Proof. Optional.

Rule of Thumb

- Lemma 1 says **Binomial**(n, p) \approx **Poisson**(np) when n is large, p is small and np is a constant.
- **Rule of thumb:** The approximation is acceptable if

$$n \geq 50 \text{ and } np \leq 10.$$

Continuous random variables

- Random variable $X : \Omega \rightarrow \mathbb{R}$ is a **continuous random variable** if the set of its possible values $X(\Omega)$ is **uncountable**.
- Uncountable: there is no sequential to list out all elements of $X(\Omega)$.
- Example: $[0, 1]$ is uncountable, any subset of \mathbb{R} is uncountable.

Probability density function

If X is a continuous random variable, there exists a **Probability Density Function (PDF)** $f(x)$ which satisfies

- (i) $f(x) \geq 0$ for all $x \in \mathbb{R}$,
- (ii) $\int_{-\infty}^{\infty} f(x)dx = 1$;
- (iii) $P(a < X < b) = \int_a^b f(x)dx$ for any $a < b$.

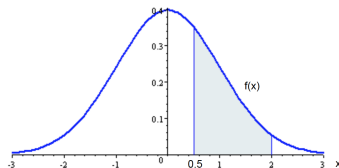
Remarks on PDF

- PDFs are not probability, the equation $P(X = x) = f(x)$ is **wrong**. In fact $P(X = x) = 0$ for any $x \in \mathbb{R}$.
- It only makes sense to talk about the probability that X is in an interval $[a, b]$, i.e., $P(X \in [a, b])$.

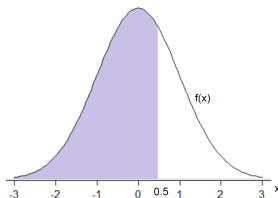
Remarks on PDF (continued)

- The condition (ii) means the area under the curve $f(x)$ is 1.
- The condition (iii) means $P(a < X < b)$ is equal to the area under $f(x)$ taken from $x = a$ to $x = b$.

Examples



$$P(0.5 < X < 2)$$



$$P(X < 0.5)$$

Remarks on PDF (continued)

For continuous random variables, the inequalities \leq and $<$ do not matter. Similarly, the inequalities \geq and $>$ do not matter.

$$P(X < a) = P(X \leq a)$$

$$P(X > b) = P(X \geq b)$$

$$\begin{aligned} P(a < X < b) &= P(a \leq X < b) \\ &= P(a < X \leq b) \\ &= P(a \leq X \leq b). \end{aligned}$$

CDF of Continuous Random Variable

- The **cumulative distribution function (CDF)** of a continuous random variable is defined by the improper integral

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(u)du.$$

- $F(x)$ is the area under the curve $y = f(x)$ from $-\infty$ to x .
- $P(a \leq X \leq b) = \int_a^b f(x)dx = F(b) - F(a)$.

Fundamental Theorem of Calculus

Theorem 1. If the PDF $f(x)$ of a continuous random variable X is continuous at $x = a$, then

$$f(a) = \left. \frac{dF(x)}{dx} \right|_{x=a}.$$

More generally, if $f(x)$ is continuous on \mathbb{R} , then

$$f(x) = \frac{dF(x)}{dx} \text{ for any } x \in \mathbb{R}.$$

Good news: Most PDFs in this course are continuous on \mathbb{R} .

Relationship Between PDF and CDF for a Continuous Random Variable

Let X be a continuous random variable with PDF f and CDF F .

- By definition, the CDF is found by integrating the PDF:

$$F(x) = \int_{-\infty}^x f(t)dt$$

- By the Fundamental Theorem of Calculus, the PDF can be found by differentiating the CDF:

$$f(x) = \frac{d}{dx}[F(x)]$$

Example

The CDF of a continuous random variable X is given as follows.

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

Find the PDF of X .

Solution.

Example

Example

Let the random variable X denote the time a person waits for an elevator to arrive. Suppose the longest one would need to wait for the elevator is 2 minutes, so the possible values of X are given by the interval $[0, 2]$. A possible PDF for X is given by

$$f(x) = \begin{cases} x & \text{if } 0 \leq x \leq 1, \\ 2 - x & \text{if } 1 < x \leq 2, \\ 0 & \text{otherwise.} \end{cases}$$

Show $f(x)$ is a continuous random variable and find the probability that the elevator arrives between 0 and 0.5 minutes.

Continued

Solution.

Continued

Continuing in the context of the previous example, find the CDF at two possible values of X , $x = 0.5$ and $x = 1.5$

Solution.

Example

Let X be a continuous random variable with the following PDF

$$f(x) = \begin{cases} ce^{-x} & \text{if } x \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

where c is a positive constant.

- (a) Find c
- (b) Find the CDF of X , $F(x)$.
- (c) Find $P(1 < X < 3)$.

Continued

Solution.

Continued

Solution.

Exercise

The length of time X , needed by students in a particular course to complete a 1 hour exam is a random variable with PDF given by

$$f(x) = \begin{cases} k(x^2 + x) & \text{if } 0 \leq x \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

For the random variable X .

- 1 Find the value k that makes $f(x)$ a probability density function (PDF)
- 2 Find the cumulative distribution function (CDF)
- 3 Use the CDF to find $P(X \leq 2)$
- 4 Find the probability that a randomly selected student will finish the exam in less than half an hour

Continued

Solution.

Continued

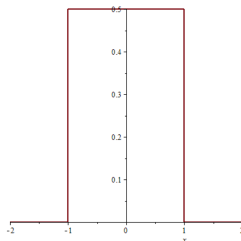
Solution.

Uniform distribution

- The **uniform distribution** on $[a, b]$, denoted $X \sim U(a, b)$, has PDF

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \leq x \leq b, \\ 0 & \text{otherwise.} \end{cases}$$

- Example: PDF of $X \sim U(-1, 1)$.



Example 2: CDF of $X \sim U(a, b)$

What is the CDF $F(x)$ of a $X \sim U(a, b)$?

Hint. $F(x) = P(X \leq x) = \int_{-\infty}^x f(u)du.$

CDF of $X \sim U(a, b)$

Normal distribution

- $\mu, \sigma \in \mathbb{R}$ with $\sigma > 0$. The **normal distribution** with parameters μ and σ , denoted $X \sim N(\mu, \sigma^2)$, has PDF

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}.$$

- If $\mu = 0$ and $\sigma = 1$, $X \sim N(0, 1)$ has the **standard normal distribution**.
- μ is called **mean** and σ is called **standard deviation** of the normal distribution.

PDF of $N(\mu, \sigma^2)$

- The PDF of $X \sim N(\mu, \sigma^2)$ is symmetric about $x = \mu$.
- In most cases, we transform $N(\mu, \sigma^2)$ into $N(0, 1)$, where all information on the later distribution is readily available.

CDF Table of $Z \sim N(0, 1)$

z	+ 0.00	+ 0.01	+ 0.02	+ 0.03	+ 0.04	+ 0.05	+ 0.06	+ 0.07	+ 0.08	+ 0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56360	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214

- $F(0) = P(Z \leq 0) = 0.5$.
- $F(0.53) = P(Z \leq 0.53) = 0.70194$.
- $P(0 \leq Z \leq 0.53) = F(0.53) - F(0) = 0.70194 - 0.5 = 0.20194$

Normal Distribution \rightarrow Standard Normal Distribution

Lemma 1. If $X \sim N(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$.

Examples of conversion between normal and standard normal.

- If $X \sim N(1, 4)$, then $Z = \frac{X-1}{2} \sim N(0, 1)$.
- If $X \sim N(2, 3)$, then $Z = \frac{X-2}{\sqrt{3}} \sim N(0, 1)$.
- If $X \sim N(-1, 100)$, then $Z = \frac{X+1}{10} \sim N(0, 1)$.

Example

The IQ scores of a large population follow a normal distribution with $\mu = 100$ and $\sigma = 15$. An individual is selected at random. What is the probability that his score X satisfies $100 < X < 115$?

Solution.

Solution

Exponential Distribution

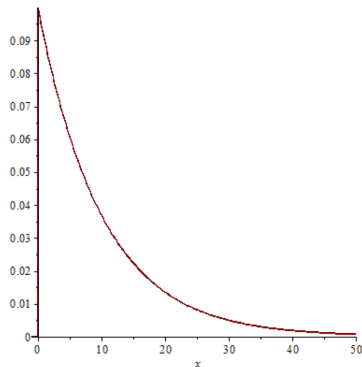
The **exponential distribution** with parameter λ ($\lambda > 0$) has PDF

$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0.$$

Denote $X \sim \text{Exp}(\lambda)$.

- Exponential distribution is usually used to model the occurrence of rare events. (Same as Poisson distribution)
- Example: X = time till occurrence of rare events like radioactive decay, earthquake, tsunami.

Graph of Exponential Distribution



PDF of $\text{Exp}(10)$

Example

Exercise X = time (in years) until the next eruption of a volcano. Assume $X \sim \text{Exp}(0.001)$. What is the probability for an eruption within the next 100 years?

Solution.

Example [Discrete]

- Guessing a Birthday: If you randomly approach a person and try to guess his/her birthday, the probability of his/her birthday falling exactly on the date you have guessed follows a uniform distribution. Reason: every day of the year has equal chances of being his/her birthday. (Each day with probability $1/365$)
- Rolling a Dice: When a fair die is rolled, the probability that the number appearing on the top of the die lies in between one to six follows a uniform distribution.
- Deck of Cards: The total number of cards present in the deck of playing cards is equal to 52.

Example [Continuous]

Imagine you live in a building that has an elevator that will take you to your floor. From experience, once you push the button to call the elevator, it takes between 10 and 30 seconds for you to arrive at your floor. This means the elevator arrival is uniformly distributed between 10 and 30 seconds once you hit the button.

Exercise

Box X contains 2 red and 3 blue balls and box Y contains 4 red and 5 blue balls. One ball is drawn at random from one of the boxes and is found to be blue. Then, the probability that it was from box Y, is

- (a) $25/52$ (b) $21/52$ (c) $15/52$ (d) $7/52$ (e)
None of the above

Solution

Solution