Lesson 8: Probability ~~distribution~~ density functions (PDFs)

Meike Niederhausen and Nicky Wakim

2025-10-20

Learning Objectives

1. Distinguish between discrete and continuous random variables.
2. Calculate probabilities for continuous random variables. from pdfs
3. Use R to simulate known continuous distributions.

Where are we?

Basics of probability

- Outcomes and events
- Sample space
- Simulations
- Probability axioms
- Probability properties
- Counting
- Rules of probability

Probability for random variables

- Probability functions: pmfs/pdfs
- Cumulative distribution functions (CDFs)
- Important distributions

Two random variables

- Transformations
- Joint distributions
- Independence and conditioning

Expected values

- Means / expected values
- Linear combinations and conditioning
- Variance

Advanced probability

- Central limit theorem
- Functions: moment generating functions

Learning Objectives

1. Distinguish between discrete and continuous random variables.
2. Calculate probabilities for continuous random variables.
3. Use R to simulate known continuous distributions.

Discrete vs. Continuous RVs

- For a **discrete** RV, the set of possible values is either finite or can be put into a countably infinite list.

- **Continuous** RVs take on values from continuous *intervals*, or unions of continuous intervals

$$p_X(x) = P(X = x)$$

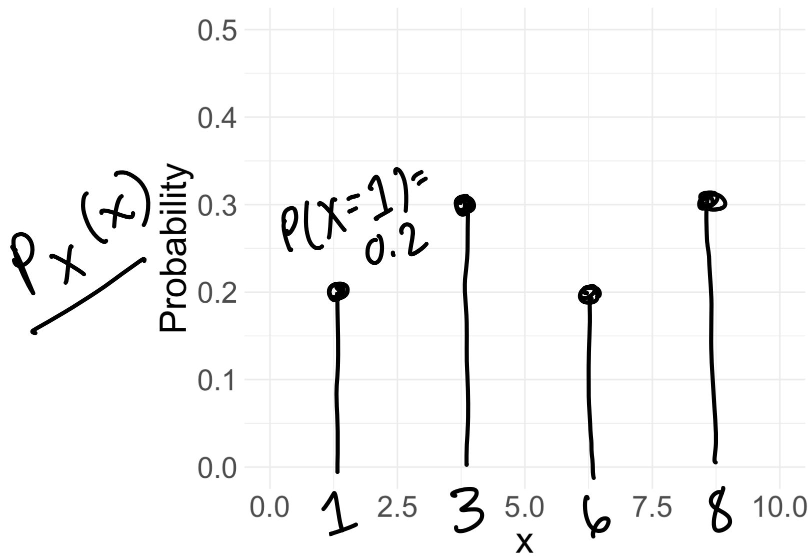
$$f_X(x) \neq P(X = x)$$

	Discrete	Continuous
probability function	mass (probability mass function, PMF)	density (probability density function, PDF)
	$0 \le p_X(x) \le 1$	$0 \le f_X(x)$ (not necessarily ≤ 1)
	$\sum_x p_X(x) = 1$	$\int_{-\infty}^{\infty} f_X(x)\,dx = 1$
	$P(0 \le X \le 2)$ $= P(X=0) + P(X=1) + P(X=2)$ if X is integer valued	$P(0 \le X \le 2)$ $= \int_0^2 f_X(x)\,dx$
	$P(X \le 3) \ne P(X < 3)$ when $P(X=3) \ne 0$	$P(X \le 3) = P(X < 3)$ since $P(X=3) = 0$ always
cumulative distribution function (CDF) $F_X(x)$	$F_X(a) = P(X \le a)$ $= \sum_{x \le a} P(X=a)$ graph of CDF is a step function with jumps of the same size as the mass, from 0 to 1	$F_X(a) = P(X \le a)$ $= \int_{-\infty}^{a} f_X(x)\,dx$ graph of CDF is nonnegative and continuous, rising up from 0 to 1
examples	counting: defects, hits, die values, coin heads/tails, people, card arrangements, trials until success, etc.	lifetimes, waiting times, height, weight, length, proportions, areas, volumes, physical quantities, etc.
named distributions	Bernoulli, Binomial, Geometric, Negative Binomial, Poisson, Hypergeometric, Discrete Uniform	Continuous Uniform, Exponential, Gamma, Beta, Normal
expected value	$\mathbb{E}(X) = \sum_x x p_X(x)$ $\mathbb{E}(g(X)) = \sum_x g(x) p_X(x)$	$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f_X(x)\,dx$ $\mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x)\,dx$
$\mathbb{E}(X^2)$	$\mathbb{E}(X^2) = \sum_x x^2 p_X(x)$	$\mathbb{E}(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x)\,dx$
variance	$\mathrm{Var}(X) =$ $\mathbb{E}(X^2) - (\mathbb{E}(X))^2$	$\mathrm{Var}(X) =$ $\mathbb{E}(X^2) - (\mathbb{E}(X))^2$
std. dev.	$\sigma_X = \sqrt{\mathrm{Var}(X)}$	$\sigma_X = \sqrt{\mathrm{Var}(X)}$

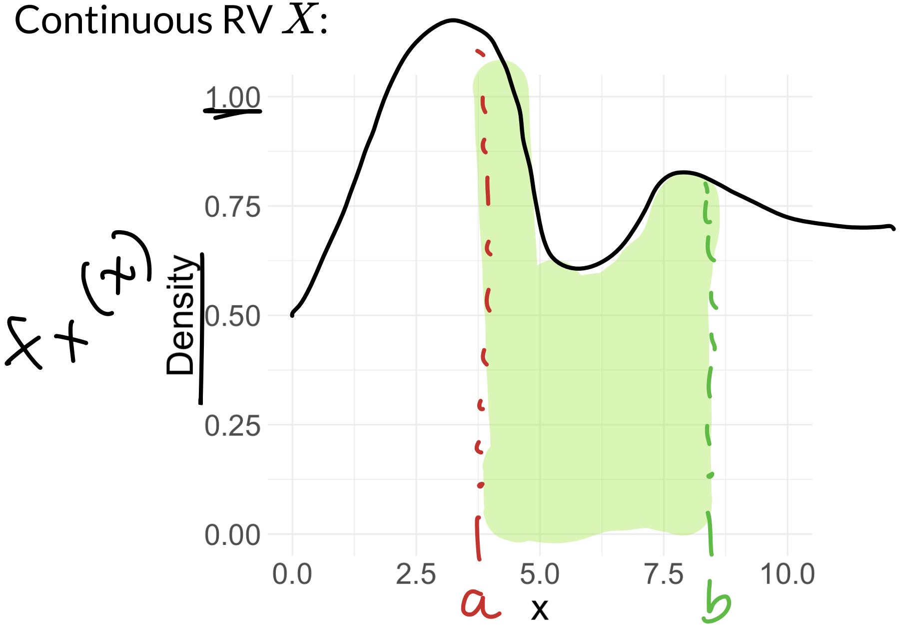
Figure from Introduction to Probability TB (pg. 301)

How to define probabilities for continuous RVs?

Discrete RV X:



Continuous RV X:



- pmf: $p_X(x) = P(X = x)$

- density: $f_X(x)$ $\neq P(X=x)$
- probability: $P(a \leq X \leq b) = \int_a^b f_X(x)dx$

integral b/w
a & b, across $f_X(x)$

What is a probability density function?

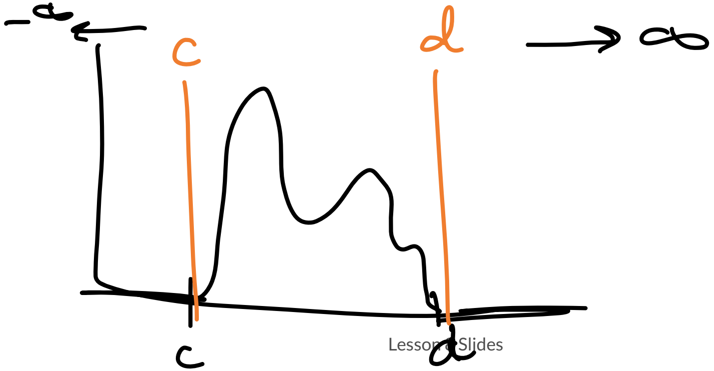
Probability density function

The probability distribution, or **probability density function (pdf)**, of a continuous random variable X is a function $f_X(x)$, such that for all real values a, b with $a \le b$,

$$\mathbb{P}(a \le X \le b) = \int_a^b f_X(x)dx$$

Remarks:

1. Note that $f_X(x) \neq \mathbb{P}(X = x)$!!!
2. In order for $f_X(x)$ to be a pdf, it needs to satisfy the properties
   - $f_X(x) \ge 0$ for all x
   - $\int_{-\infty}^{\infty} f_X(x)dx = 1$



Learning Objectives

1. Distinguish between discrete and continuous random variables.

2. Calculate probabilities for continuous random variables.

3. Use R to simulate known continuous distributions.

Let's demonstrate the PDF with an example (1/5)

Example 1.1

Let $f_X(x) = 2$, for $a \leq x \leq 3$.

1. Find the value of a so that $f_X(x)$ is a ~~pdf~~.
   valid

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

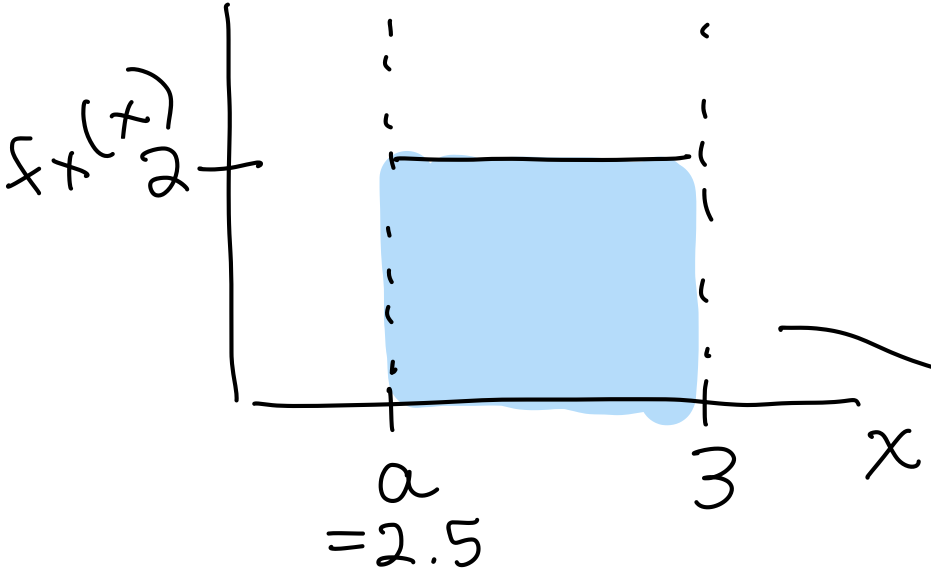
$$\int_a^3 2 = 1$$

$$= 2x \Big|_{x=a}^{x=3} = 2(3) - 2(a)$$

$$= 6 - 2a = 1$$

$$2a = 5$$

$$a = 2.5$$



geom: (Area of rect) $h \cdot w = 1$

$$2(3-a) = 1$$

$$a = 2.5$$

Let’s demonstrate the PDF with an example (2/5)

Example 1.2

Let $f_X(x) = 2$, for $a \leq x \leq 3$. 2.5

2. Find $\mathbb{P}(2.7 \leq X \leq 2.9)$. 3.6

$$P(2.7 \leq X \leq 2.9) = \int_{2.7}^{2.9} 2\,dx$$

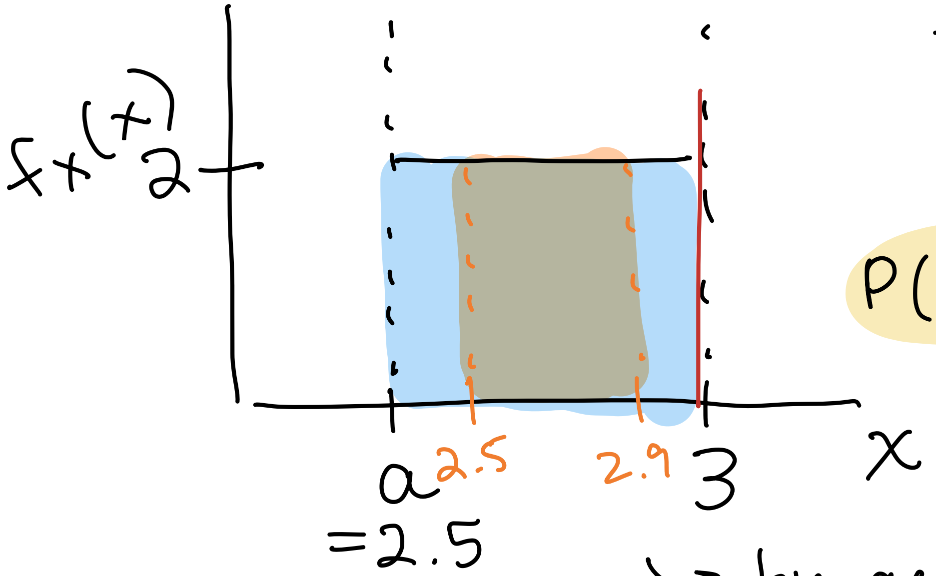
check intergrands are w/in your bounds

$$= 2x \Big|_{x=2.7}^{x=2.9}$$

$$= 2(2.9) - 2(2.7)$$

$$= 0.4$$

$$P(2.7 \leq X \leq 2.9) = 0.4$$



by geom: area of orange / area of blue

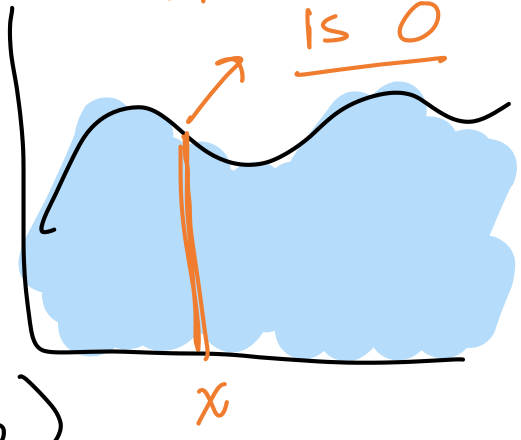
Let's demonstrate the PDF with an example (3/5)

Example 1.3

Let $f_X(x) = 2$, for $a \le x \le 3$.

3. Find $\mathbb{P}(2.7 < X \le 2.9)$.

$P(X = x) = 0$



$P(a \le X \le b) = P(a < X < b)$

$P(2.7 < X \le 2.9) = P(2.7 \le X \le 2.9)$

$= 0.4$

Let's demonstrate the PDF with an example (4/5)

Example 1.4

Let $f_X(x) = 2$, for $a \leq x \leq 3$.

4. Find $\mathbb{P}(X = 2.9)$.

RULE

$$\boxed{P(X=x)=0}$$

$$P(X=2.9) = \int_{2.9}^{2.9} f_X(x)\,dx$$

$$= \int_{2.9}^{2.9} 2\,dx$$

$$= 2x \Big|_{x=2.9}^{x=2.9}$$

$$= 2(2.9) - 2(2.9)$$

$$= 0$$

Let's demonstrate the PDF with an example (5/5)

Example 1.5

Let $f_X(x) = 2$, for $a \leq x \leq 3$.

5. Find $\mathbb{P}(X \leq 2.8)$.

$$P(X \leq 2.8) = \int_{2.5}^{2.8} 2 \, dx$$

$$= 2x \Big|_{x=2.5}^{x=2.8}$$

$$= 2(2.8 - 2.5)$$

$$= 0.6$$

$P(X \leq x)$ is the basis of cumulative distribution function (CDF)

Learning Objectives

1. Distinguish between discrete and continuous random variables.
2. Calculate probabilities for continuous random variables.
3. Use R to simulate known continuous distributions.

Use R to simulate known distributions

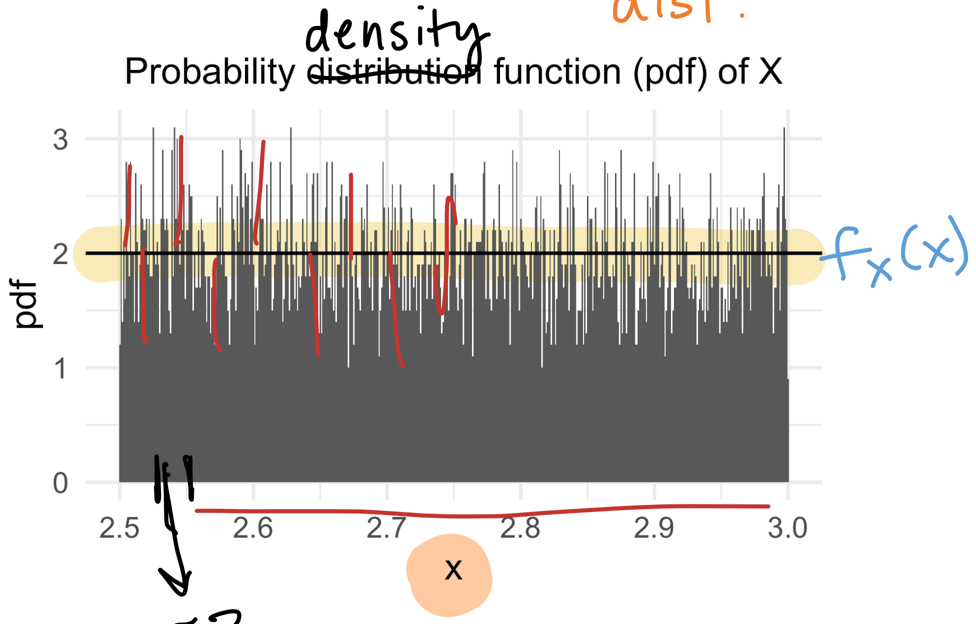
$f_X(x) = 2$, $2.5 \le x \le 3$ is a uniform dist.

- We can use R to simulate continuous random variables and visualize their distributions
- For example, we can simulate a *uniform* distribution between 2.5 and 3

```
uniform = tibble(
  x = runif(n=10000, min=2.5, max=3)
)

ggplot(uniform,
       aes(x = x,
           y = after_stat(density))) +
  geom_histogram( binwidth = 0.001) +
  geom_abline(intercept = 2, slope = 0) +
  labs(
    title = "Probability distribution funct
    x = "x",
    y = "pdf"
  )
```

density



2.53
2.533
2.5401

Use R to simulate any continuous distribution

- We will discuss other ways to simulate continuous distributions once we cover cumulative distribution functions (CDFs) and inverse CDFs