Lesson 9: Cumulative distribution functions (CDFs)
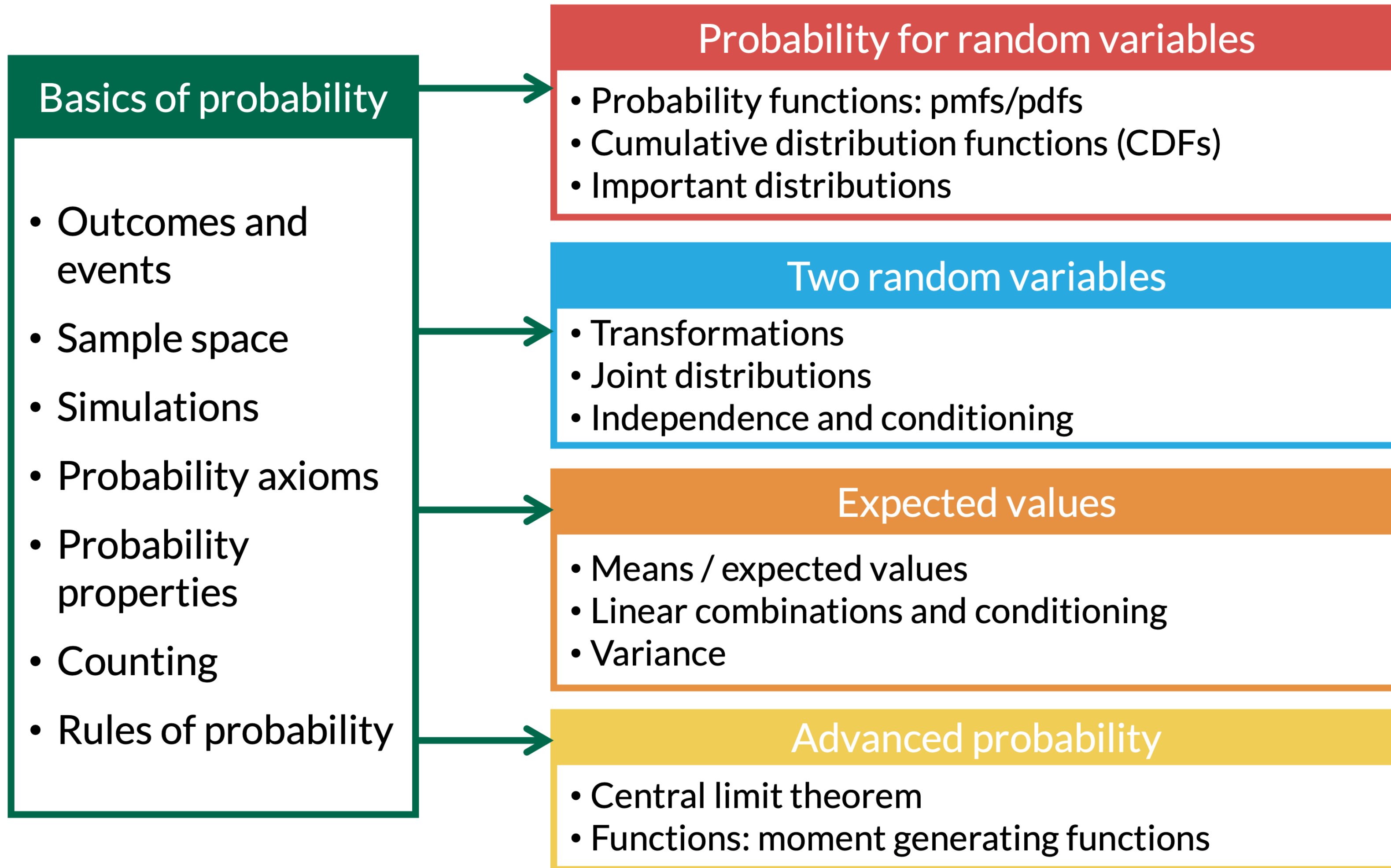
Nicky Wakim

2025-10-22

Learning Objectives

1. Understand the definition of cumulative distribution functions (CDFs) for discrete and continuous random variables.

2. Compute CDFs for given probability mass functions (pmfs).

3. Compute CDFs for given probability density functions (pdfs) and pdfs from CDFs.

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. Understand the definition of cumulative distribution functions (CDFs) for discrete and continuous random variables.

2. Compute CDFs for given probability mass functions (pmfs).

3. Compute CDFs for given probability density functions (pdfs) and pdfs from CDFs.

What is a cumulative distribution function?

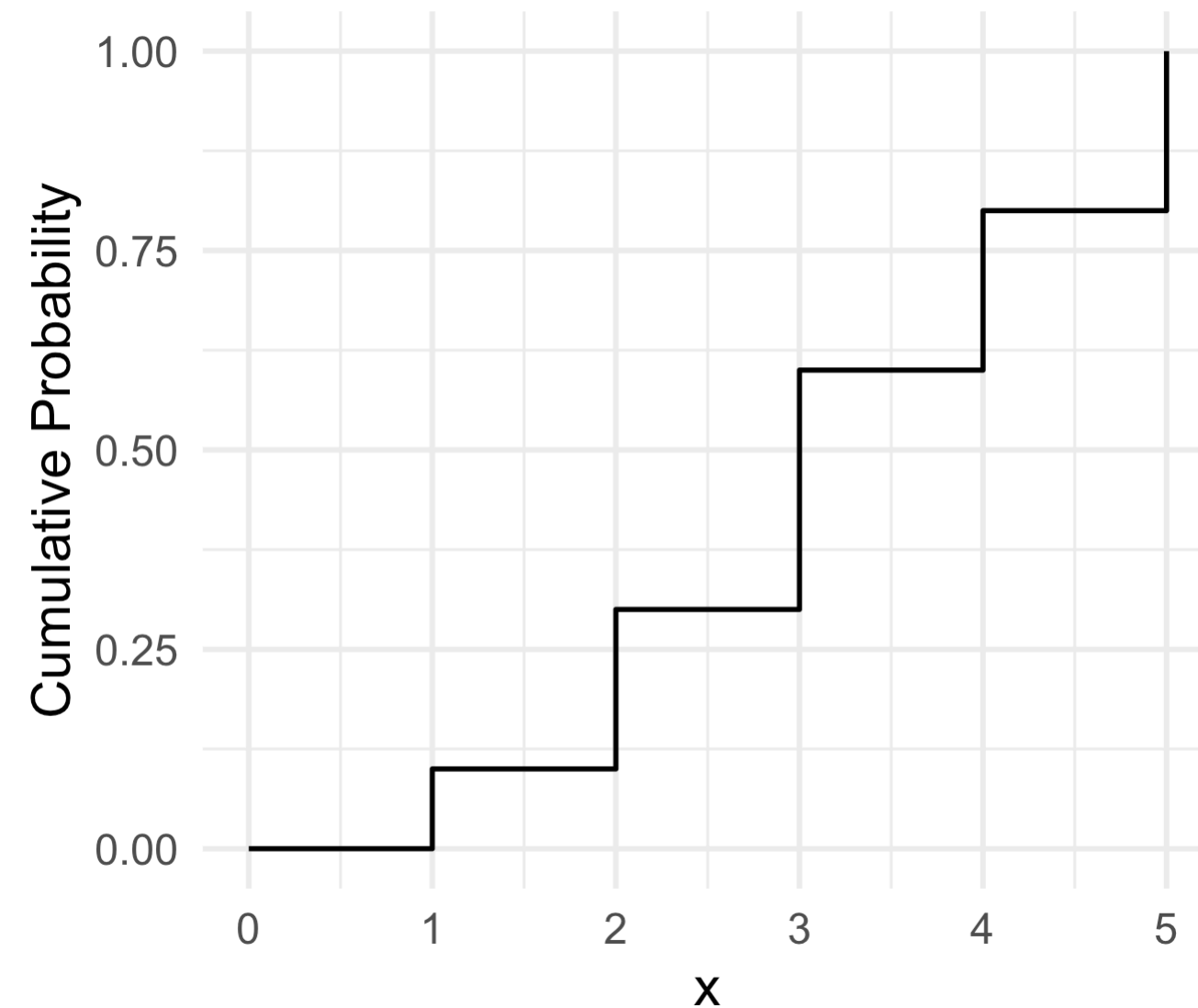| Cumulative distribution function (CDF) for discrete random variable	Cumulative distribution function (CDF) for continuous random variable
The **cumulative distribution function (cdf)** of a **discrete RV** X with pmf $p_X(x)$, is defined for every value x by $$F_X(x) = \mathbb{P}(X \le x) = \sum_{\{all\ y:\ y \le x\}} p_X(y)$$	The **cumulative distribution function (cdf)** of a **continuous RV** X, is the function $F_X(x)$, such that for all real values of x, $$F_X(x) = \mathbb{P}(X \le x) = \int_{-\infty}^{x} f_X(s)ds$$

- $F(x)$ is increasing or flat (never decreasing)
- $\min_{x} F(x) = 0$
- $\max_{x} F(x) = 1$
- CDF is a step function

Remarks: In general, $F_X(x)$ is increasing and

- $\lim_{x \to -\infty} F_X(x) = 0$
- $\lim_{x \to \infty} F_X(x) = 1$
- $P(X > a) = 1 - P(X \le a) = 1 - F_X(a)$
- $P(a \le X \le b) = F_X(b) - F_X(a)$

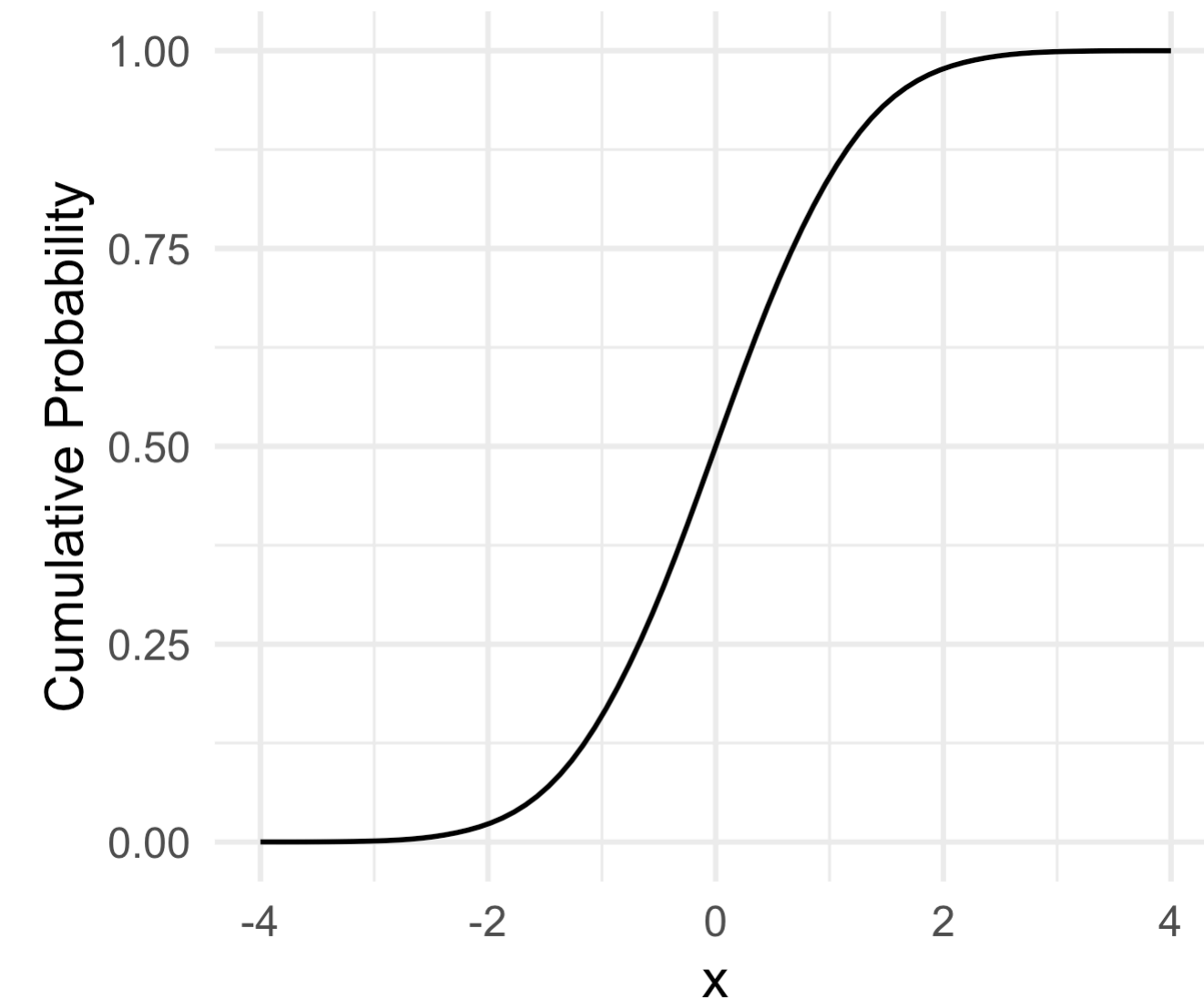# How to define CDFs for discrete and continuous RVs?

Discrete RV X:

Continuous RV X:

- pmf: $p_X(x) = P(X = x)$
- CDF: $F_X(x) = P(X \leq x) = \sum\limits_{\{all\ y:\ y \leq x\}} p_X(y)$

- density: $f_X(x)$
- probability: $P(a \leq X \leq b) = \int_a^b f_X(x)dx$
- CDF: $F_X(x) = P(X \leq x) = \int_{-\infty}^x f_X(s)ds$

Learning Objective

1. Understand the definition of cumulative distribution functions (CDFs) for discrete and continuous random variables.

2. Compute CDFs for given probability mass functions (pmfs).

3. Compute CDFs for given probability density functions (pdfs) and pdfs from CDFs.

Falls in Older Adults Revisited (1/5)

Recall our pmf:

$$P(X = x) = \binom{8}{x} 0.25^x 0.75^{8-x}, x = 0, 1, 2, \ldots, 8$$

Falls in Older Adults Revisited (2/5)

1. Write the CDF of X.

Recall our pmf:

$$P(X = x) = \binom{8}{x} 0.25^x 0.75^{8-x}, x = 0, 1, 2, \ldots, 8$$

$$F_X(x) = P(X \leq x) = \sum_{k=0}^{x} \binom{8}{k} 0.25^y 0.75^{8-k}$$

Falls in Older Adults Revisited (3/5)

2. Use R to calculate the cumulative probability for each possible value of X.

```
1  n = 8
2  p = 0.25
3
4  falls_cdf <- tibble(
5    x = 0:n,
6    c_prob = pbinom(x, size = n, prob = p)
7  )
```
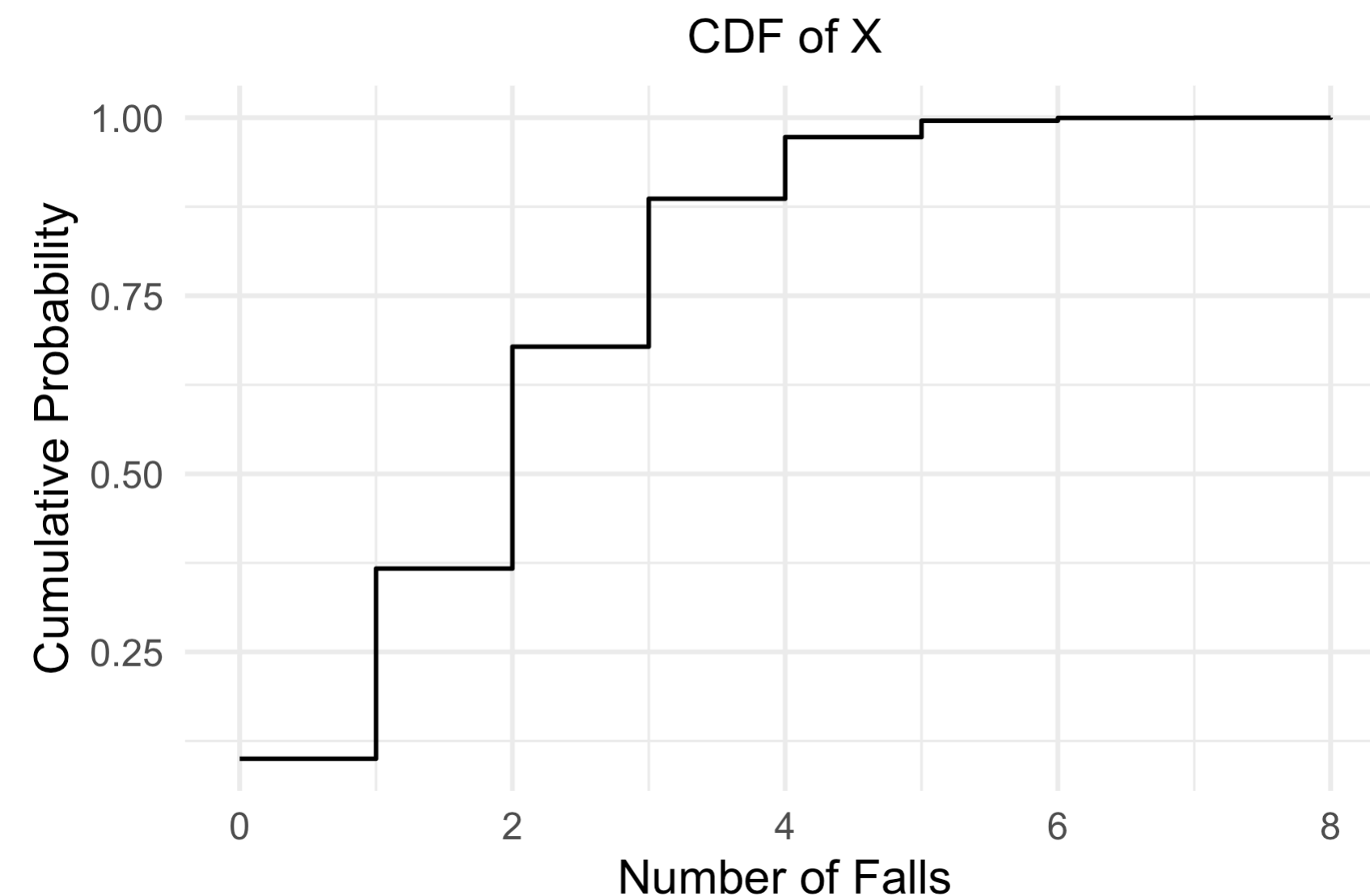
```
1  falls_cdf
```

```
# A tibble: 9 × 2
      x c_prob
  <int>  <dbl>
1     0  0.100
2     1  0.367
3     2  0.679
4     3  0.886
5     4  0.973
6     5  0.996
7     6  1.00
8     7  1.00
9     8  1
```

Falls in Older Adults Revisited (4/5)

3. Plot the CDF of X.


```r
1  ggplot(
2    falls_cdf,
3    aes(x = x, y = c_prob)
4        ) +
5    geom_step(
6      size = 1,
7      color = "black"
8      ) +
9    labs(
10     x = "Number of Falls",
11     y = "Cumulative Probability",
12     title = "CDF of X"
13     )
```
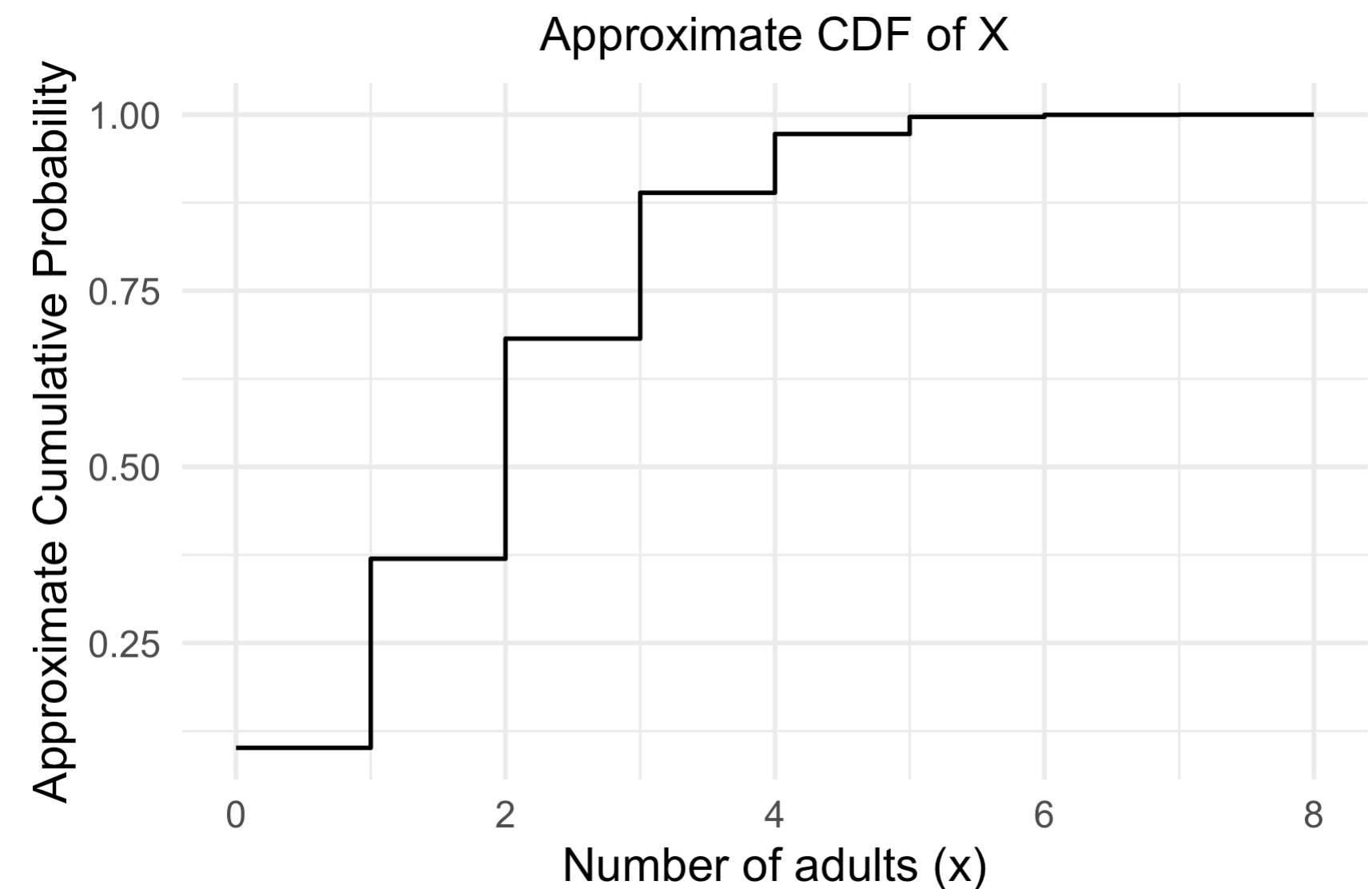
Falls in Older Adults Revisited (4/5)

4. Simulate X for 10000 groups and plot the approximated CDF.

```
1  ggplot(falls2, aes(x = x, y = c_prob)) +
2    geom_step(size = 1, color = "black") +
3    labs(
4      title = "Approximate CDF of X",
5      x = "Number of adults (x)",
6      y = "Approximate Cumulative Probability
7    )
```

```
1  set.seed(4764)
2  reps = 10000
3
4  sims = rbinom(n = reps,
5                size = n,
6                prob = p)
7
8  sims %>% head(., 14)
```

```
[1] 2 1 2 1 3 3 2 2 4 2 3 0 2 0
```

```
1  falls2 <- tibble(x = 0:n) %>%
2    rowwise() %>%
3    mutate(c_prob = sum(sims <= x) / reps)
```


Approximate CDF of X

Learning Objective

1. Understand the definition of cumulative distribution functions (CDFs) for discrete and continuous random variables.

2. Compute CDFs for given probability mass functions (pmfs).

3. Compute CDFs for given probability density functions (pdfs) and pdfs from CDFs.

Let's demonstrate the CDF with an example

Example 2

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

Derivatives of the CDF

Theorem 1

If X is a continuous random variable with pdf $f_X(x)$ and cdf $F_X(x)$, then for all real values of x at which $F'_X(x)$ exists,

$$\frac{d}{dx}F_X(x) = F'_X(x) = f_X(x)$$

Finding the PDF from a CDF

Example 3

Let X be a RV with cdf

$$F_X(x) = \begin{cases} 0 & x < 2.5 \\ 2x - 5 & 2.5 \leq x \leq 3 \\ 1 & x > 3 \end{cases}$$

Find the pdf $f_X(x)$.

Let's go through another example (1/7)

Example 4

Let X be a RV with pdf $f_X(x) = 2e^{-2x}$, for $x > 0$.

1. Show $f_X(x)$ is a pdf.

2. Find $\mathbb{P}(1 \leq X \leq 3)$.

3. Find $F_X(x)$.

4. Given $F_X(x)$, find $f_X(x)$.

5. Find $\mathbb{P}(X \geq 1 | X \leq 3)$.

6. Find the median of the distribution of X.

Let's go through another example (2/7)

Let X be a RV with pdf
$f_X(x) = 2e^{-2x}$, for $x > 0$.

1. Show $f_X(x)$ is a pdf.

Let's go through another example (3/7)

Do this problem at home for extra practice.

Example 4.2

Let X be a RV with pdf
$f_X(x) = 2e^{-2x}$, for $x > 0$.

2. Find $\mathbb{P}(1 \leq X \leq 3)$.

Let's go through another example (4/7)

Example 4.3

Let X be a RV with pdf
$f_X(x) = 2e^{-2x}$, for $x > 0$.

3. Find $F_X(x)$.

Let's go through another example (5/7)

Do this problem at home for extra practice.

Let X be a RV with pdf
$f_X(x) = 2e^{-2x}$, for $x > 0$.

4. Given $F_X(x)$, find $f_X(x)$.

Let's go through another example (6/7)

Let X be a RV with pdf
$f_X(x) = 2e^{-2x}$, for $x > 0$.

5. Find $\mathbb{P}(X \geq 1 | X \leq 3)$.

Let's go through another example (7/7)

Example 4.6

Let X be a RV with pdf
$f_X(x) = 2e^{-2x}$, for $x > 0$.

6. Find the median of the
   distribution of X.