

Lesson 9: Cumulative distribution functions (CDFs)

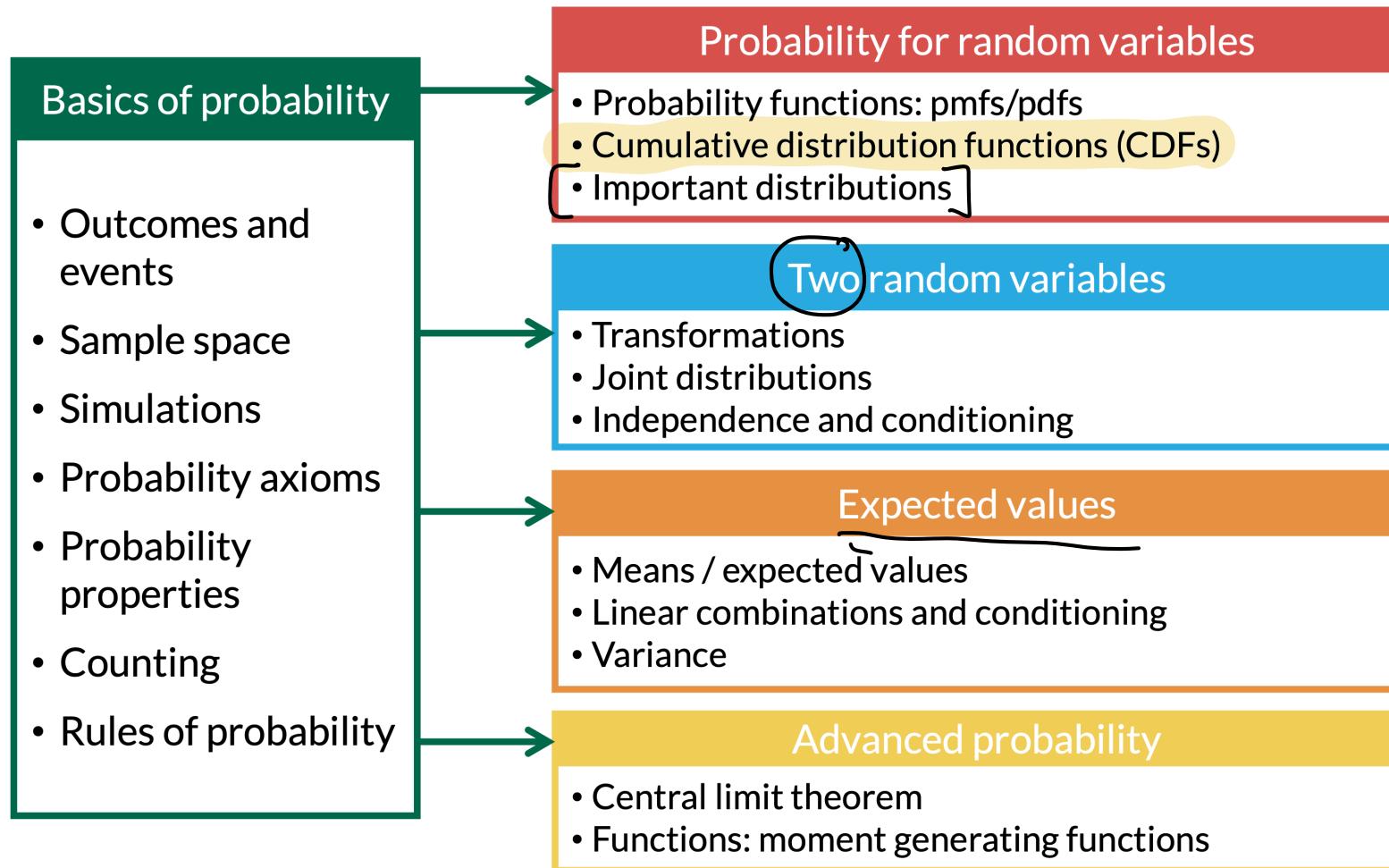
Nicky Wakim

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?



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.

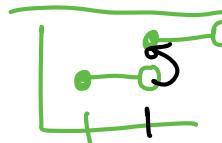
What is a cumulative distribution function?

Cumulative distribution function (CDF) for discrete random variable

The cumulative distribution function (cdf) of a discrete RV X with pmf $p_X(x)$, is defined for every value x by

$$F_X(x) = \mathbb{P}(X \leq x) = \sum_{\{all \ y: y \leq x\}} p_X(y)$$

- $F(x)$ is increasing or flat (never decreasing)
- $\min_x F(x) = 0 \leftarrow f_X(x) \text{ starts @ } 0 \rightarrow$
- $\max_x F(x) = 1 \leftarrow @ \max x, \text{ all probs summed/integrated, so equals } 1 \rightarrow$
- CDF is a step function



Cumulative distribution function (CDF) for continuous random variable

The cumulative distribution function (cdf) of a continuous RV X , is the function $F_X(x)$, such that for all real values of x ,

$$F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^x f_X(s) ds$$

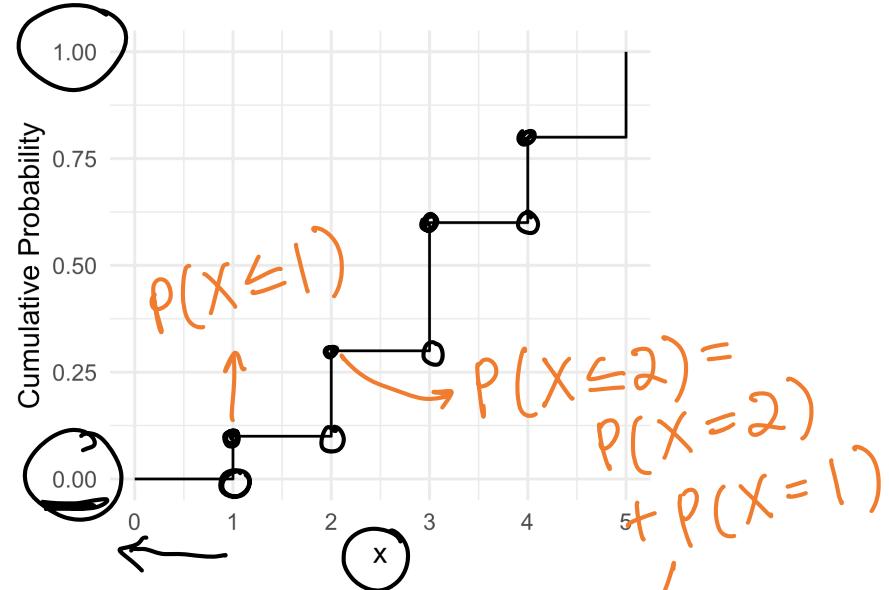
$s = -\infty$
to $s = x$

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

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

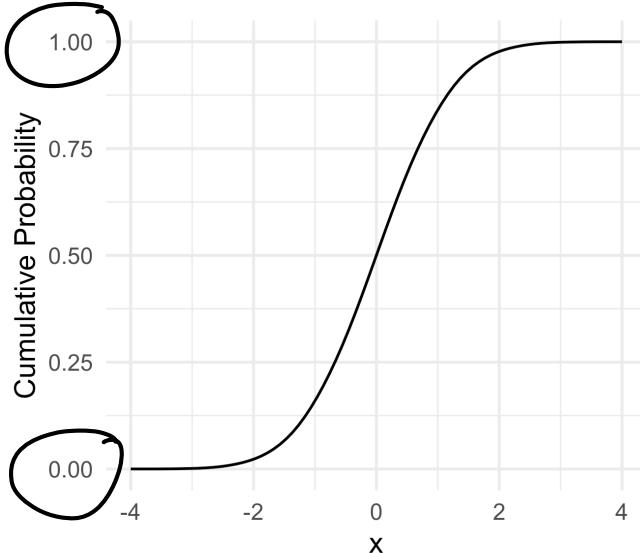
How to define CDFs for discrete and continuous RVs?

Discrete RV X :



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

Continuous RV X :



- 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)

Example 1: Falls in Older Adults

A major public health concern is falls among older adults (age 65+). National data suggests that 25% of older adults will experience at least one fall within a given year. A community health program is tracking a random group of $n = 8$ older adults for one year. Assume the likelihood of falling is independent from person to person.

Let X be the random variable representing the number of individuals in this group who experience at least one fall.

1. Write the CDF of X ~~and make a table of values.~~
2. Use R to calculate the cumulative probability for each possible value of X .
3. Plot the CDF of X .
4. Simulate X for 10000 groups and plot the approximated CDF.

Recall our pmf:

Binomial
RV

$$\underbrace{P(X = x)}_{\text{Binomial RV}} = \binom{8}{x} 0.25^x 0.75^{8-x}, x = \underbrace{0}_{\text{Binomial RV}} 1, 2, \dots, \underbrace{8}_{\text{Binomial RV}}$$

Falls in Older Adults Revisited (2/5)

Example 1: Falls in Older Adults

1. Write the CDF of X .

Recall our pmf:

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

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

CDF @ 2

$$F_X(2) = P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$$

$$\begin{aligned} &= \binom{8}{0} 0.25^0 0.75^8 + \binom{8}{1} 0.25^1 0.75^7 + \binom{8}{2} 0.25^2 0.75^6 \\ &= 0.679 \end{aligned}$$

Falls in Older Adults Revisited (3/5)

Example 1: Falls in Older Adults

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 )
```

create data frame

possible x -values

$\rightarrow P(X \leq x)$

```
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

x

c_prob

sum across ALL
 x 's \Rightarrow total prob = 1

$$P(X \leq 8) = P(X=0) + P(X=1)$$

$$+ P(X=2) + P(X=3) +$$

$$P(X=4) + P(X=5) +$$

$$P(X=6) + P(X=7) + P(X=8)$$

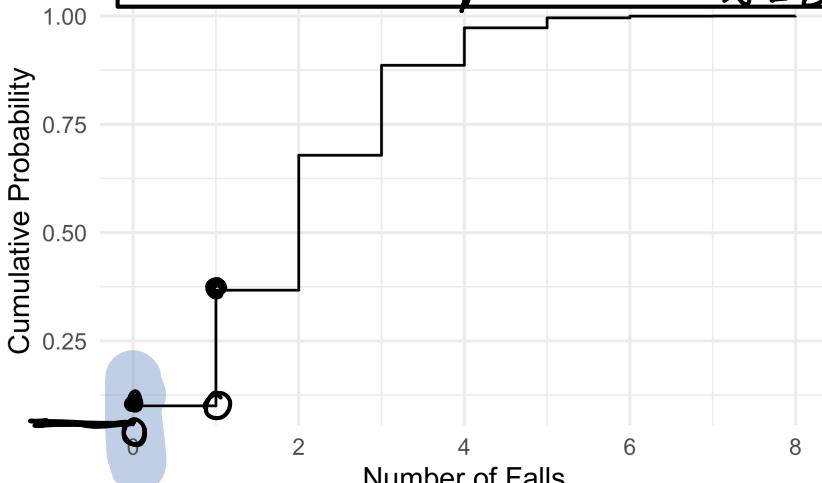
Falls in Older Adults Revisited (4/5)

Example 1: Falls in Older Adults

3. Plot the CDF of X .

```
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  )
```

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 0.1 & 0 \leq x < 1 \\ 0.367 & 1 \leq x < 2 \\ 0.679 & 2 \leq x < 3 \\ 0.886 & 3 \leq x < 4 \\ 0.973 & 4 \leq x < 5 \\ 0.996 & 5 \leq x < 6 \\ 0.9999... & 6 \leq x < 7 \\ 0.9999... & 7 \leq x < 8 \\ 1 & x \geq 8 \end{cases}$$



Falls in Older Adults Revisited (4/5)

Example 1: Falls in Older Adults

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

```

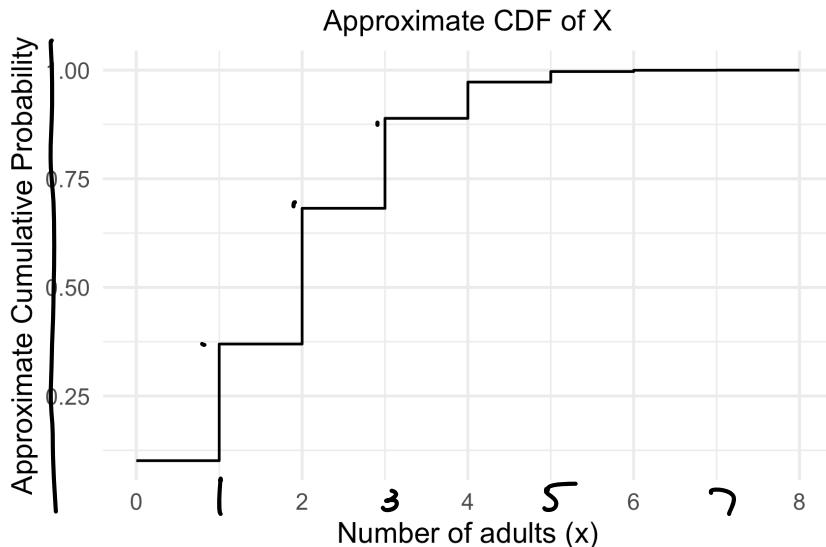
1 set.seed(4764)
2 reps = 10000
3
4 sims = rbinom(n = reps,
5   size = n,
6   prob = p)
7
8 sims %>% head(., 14)
[1] 2 1 2 1 3 3 2 2 4 2 3 0 2 0
1 falls2 <- tibble(x = 0:n) %>%
2   rowwise() %>%
3   mutate(c_prob = sum(sims <= x) / reps)

```

```

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   )

```



* at home: try 100 reps & see how affects CDF (approx)

Learning Objective

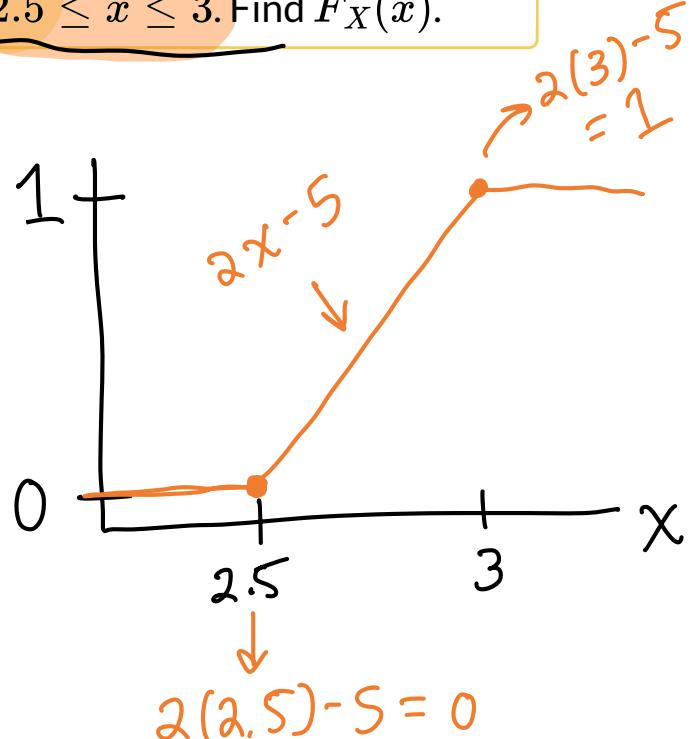
1. Understand the definition of cumulative distribution functions (CDFs) for discrete and continuous random variables.
2. Compute CDFs for given probability mass functions (pmfs).
3. Compute CDFs for given probability density functions (pdfs) and pdfs from CDFs.

break until 11:15 am

Let's demonstrate the CDF with an example

Example 2

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



$$F_X(x) = P(X \leq x) = \int_{2.5}^x f_X(s) ds$$

$$\begin{aligned} &= \int_{2.5}^x 2 ds = 2s \Big|_{s=2.5}^{s=x} \\ &= 2x - 2(2.5) \\ &= 2x - 5 \end{aligned}$$

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

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,

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

integral of pdf to get CDF
derivative of CDF to get pdf

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 $f_X(x)$.

$$\frac{d}{dx}(0) = 0$$

$$\begin{aligned} \frac{d}{dx}(2x - 5) &= \frac{d}{dx}(2x) - \frac{d}{dx}(5) \\ &= 2 - 0 = 2 \end{aligned}$$

$$\frac{d}{dx}(1) = 0$$

$$f(x) = F'(x) = \begin{cases} 0 & x < 2.5 \\ 2 & 2.5 \leq x \leq 3 \\ 0 & x > 3 \end{cases}$$

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 $\mathbb{P}(X \geq 1 | X \leq 3)$. → given
- ★ 6. Find the median of the distribution of X .

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

Example 4.1

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

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

Show:

$$\textcircled{1} \quad f_X(x) > 0 \text{ for all } x$$

$$\textcircled{2} \quad \int_{-\infty}^{\infty} f_X(x) dx = 1$$

$\Rightarrow f_X(x)$ is a pdf!

$$\textcircled{1} \quad f_X(x) = 2e^{-2x} > 0 \quad ? \quad \checkmark$$

$$x=1 \quad x=0.5 \\ 2e^{-2} = \frac{2}{e^2} \quad 2e^{-1} = \frac{2}{e} \\ \text{as } x \rightarrow 0 \quad f(x) \rightarrow \infty$$

$$\text{as } x \rightarrow \infty \quad x=1000 \quad \frac{2}{e^{1000}} \rightarrow 0$$

$$\textcircled{2} \quad \int_0^{\infty} 2e^{-2x} dx$$

$u = -2x$
 $du = -2dx$
 $\hookrightarrow 2dx = -du$

$$\begin{aligned} &= \int_0^{-\infty} e^u (-du) \\ &= + \int_{-\infty}^0 e^u (+du) = \int_{-\infty}^0 e^u du = e^u \Big|_{u=-\infty}^{u=0} \end{aligned}$$

bounds: $u = -2x$
 $u = -2(\infty) = -\infty$
 $u = -2(0) = 0$

$$= e^0 - e^{-\infty} = 1 - 0 = 1 \quad \checkmark$$

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

$$\begin{aligned}\mathbb{P}(1 \leq X \leq 3) &= \int_1^3 2e^{-2x} dx \quad \text{u-sub in 4.1} \\ &= -e^{-2x} \Big|_1^3 \\ &= -e^{-2(3)} - (-e^{-2(1)}) \\ &= -e^{-6} + e^{-2}\end{aligned}$$

$$\mathbb{P}(1 \leq X \leq 3) = e^{-2} - e^{-6}$$

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 $F_X(x)$.

$$\begin{aligned} F_X(x) &= P(X \leq x) = \int_0^x f_X(s) ds \\ &= \int_0^x 2e^{-2s} ds \quad \text{anti deriv} \\ &= -e^{-2s} \Big|_{s=0}^{s=x} = -e^{-2x} + (+e^{-2(0)}) \\ &= -e^{-2x} + 1 = 1 - e^{-2x} \end{aligned}$$

$$F_X(x) = \begin{cases} 0 & x \leq 0 \\ 1 - e^{-2x} & x > 0 \end{cases}$$

where does $F_X(x) = 1$?

$$\lim_{x \rightarrow \infty} 1 - e^{-2x} = 1 \quad \xrightarrow{\rightarrow 0}$$

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

Do this problem at home for extra practice.

Example 4.4

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

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

$$\begin{aligned}f_X(x) &= \frac{d}{dx} F_X(x) = F'_X(x) \\&= \frac{d}{dx} (1 - e^{-2x}) = 0 - (-2)e^{-2x} \\&= 2e^{-2x}\end{aligned}$$

$$f_X(x) = 2e^{-2x} \text{ for } x > 0$$

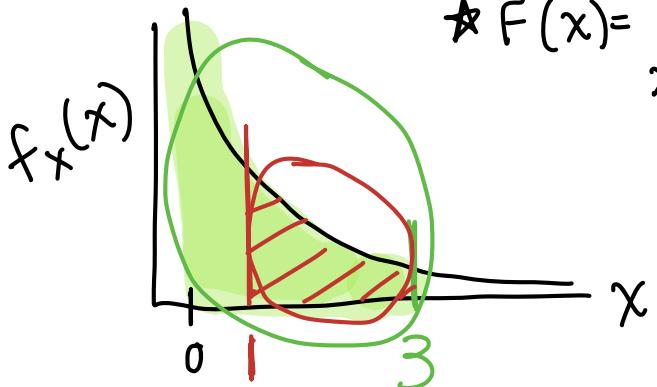
Let's go through another example (6/7)

Example 4.5

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

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

$$\star F(x) = 1 - e^{-2x} \quad x > 0$$



$$\mathbb{P}(X \geq 1 | X \leq 3) = \frac{\text{red area}}{\text{green area}}$$

$$\begin{aligned} \mathbb{P}(X \geq 1 | X \leq 3) &= \frac{P(1 \leq X \leq 3)}{P(X \leq 3)} \\ &= \frac{F_X(3) - F_X(1)}{F_X(3)} \quad \xrightarrow{F_X(3)} \\ &= \frac{[1 - e^{-2(3)}] - [1 - e^{-2(1)}]}{1 - e^{-2(3)}} \\ &= \frac{-e^{-6} + e^{-2}}{1 - e^{-6}} = \frac{e^{-2} - e^{-6}}{1 - e^{-6}} \end{aligned}$$

$$\mathbb{P}(X \geq 1 | X \leq 3) = 0.1332$$

RECALL:
 $P(A|B) = \frac{P(A \cap B)}{P(B)}$

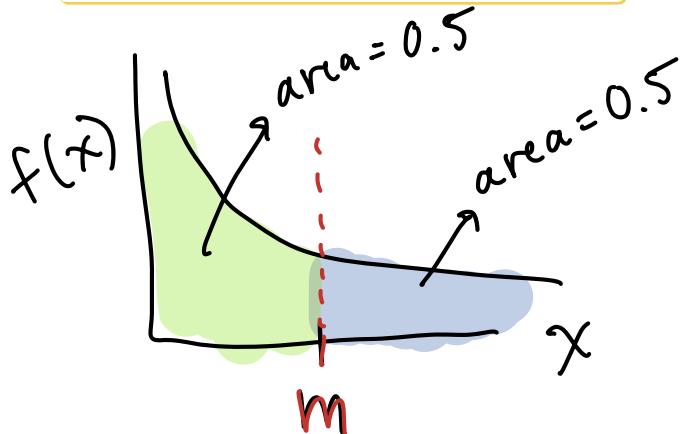
$$\begin{aligned} P(a \leq X \leq b) &= F(b) - F(a) \end{aligned}$$

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 .



median: $P(X \leq m) = 0.5$ $P(X > m) = 0.5$

$$P(X \leq m) = \int_0^m 2e^{-2x} dx = 0.5$$

↙

$$F_x(m) = 0.5$$

$$\frac{1 - e^{-2(m)}}{1 + e^{-2m}} = 0.5$$

$$\log(0.5) = -2m$$

$$\frac{\log(0.5)}{-2} = \frac{-2m}{-2}$$

$$m = \frac{\log(0.5)}{-2} = 0.34657$$

$\log = \ln$
 $\log(e^x) = x$