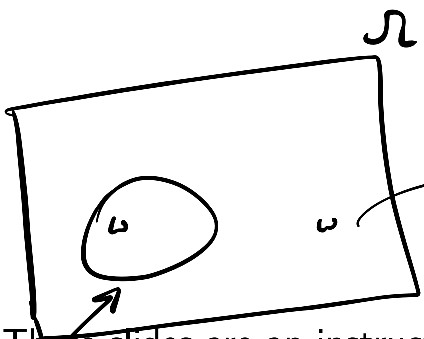
STAT 712 fa 2022 Lec 4 slides

Random variables

Karl B. Gregory

University of South Carolina



$X(\omega)$

$\mathbb{R} = (-\infty, \infty)$

$\in \mathcal{B}(\mathbb{R})$

$A = (a,b) \subset \mathbb{R}$

These slides are an instructional aid; their sole purpose is to display, during the lecture, definitions, plots, results, etc. which take too much time to write by hand on the blackboard. They are not intended to explain or expound on any material.

X

$X^{-1}\big((a,b)\big) = \underbrace{\{\omega \in \Omega : X(\omega) \in (a,b)\}} \in \mathcal{A}$

A random variable is a numeric encoding of the outcome of an experiment.

Random variable

such that

Given a p. space $(\Omega, \mathcal{B}, P)$, a *random variable* X is a function $\boxed{X}: \Omega \to \mathbb{R}$ s.t.

$$X^{-1}(A) := \{\omega \in \Omega : X(\omega) \in A\} \in \mathcal{B} \quad \text{for all } A \in \mathcal{B}(\mathbb{R}),$$

where $\boxed{\mathcal{B}(\mathbb{R}) \text{ is the Borel } \sigma\text{-algebra.}}$ (a, b)

The inverse image under X of any Borel set must belong to $\mathcal{B}$.

$\mathcal{X}$

Let $\boxed{\mathcal{X}}$ be the set of values X may take. This is called the *support* of X.

Exercise: Make $(\Omega, \mathcal{B}, P)$ for each experiment and check if X is an rv:

1. Flip a coin and let $X = 1$ if heads, $X = 0$ if tails.
2. Flip a coin three times and let $X =$ number of heads.
3. Let $X =$ time until you drop your new phone.

② $\Omega = \{H, T\}$

$\mathcal{B} = \{\emptyset, \{H,T\}, \{H\}, \{T\}\}$ ←

$$P(\omega) = \begin{cases} \frac{1}{2} & \omega = H \\ \frac{1}{2} & \omega = T \end{cases}$$

$$X(\omega) = \begin{cases} 1 & \omega = H \\ 0 & \omega = T \end{cases}$$

$\mathcal{X} = \{0, 1\}$

Choose a set in $\mathcal{B}(\mathbb{R})$ [borel σ-alg. on $\mathbb{R}$], e.g. $\underline{(1/2, 2)}$.

$$X^{-1}((\tfrac{1}{2}, 2)) = \{\omega \in \Omega : X(\omega) \in (\tfrac{1}{2}, 2)\}$$

$$= \{H\} \quad ✓$$

$$\in \mathcal{B}$$

② $\Omega = \{$ HHH   THH, HTH, HHT   TTH, THT, HTT   TTT $\}$

$\mathcal{B} = \{$all subsets of Ω, including Ω itself$\}$ ←

$P(\omega) = \frac{1}{8}$ for all $\omega \in \Omega$

$$X(\omega) = \begin{cases} 0 & \omega = TTT \\ 1 & \omega \in \{TTH,\ THT,\ HTT\} \\ 2 & \omega \in \{HHT,\ HTH,\ THH\} \\ 3 & \omega = HHH \end{cases}$$

$$\mathcal{X} = \{0, 1, 2, 3\}$$

look at $(-1, 2)$.

$$X^{-1}\big((-1,2)\big) = \{\omega \in \Omega : X(\omega) = (-1,2)\}$$

$$= \{TTT,\ TTH,\ THT,\ HTT\}$$

$$\in \mathcal{B}$$

$P_X(X \in (a,b)) = P(\{\omega \in \Omega : X(\omega) \in (a,b)\})$

A probability function for X

For a random variable X on $(\Omega, \mathcal{B}, P)$, the function

$$P_X(X \in A) = P(\underbrace{\boxed{\{\omega \in \Omega : X(\omega) \in A\}}}_{X^{-1}(A)}) \quad \text{for all } A \in \mathcal{B}(\mathbb{R})$$

is a probability function on $\mathcal{B}(\mathbb{R})$.

So X makes a new p. space $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_X)$.

We often refer to P_X as the *probability distribution* of X.

Exercise: For rolling two dice we have $(\Omega, \mathcal{B}, P)$ given by

$$\begin{aligned} \Omega &= \{(1,1),(1,2),(2,1)\dots\} \\ \mathcal{B} &= \{\text{all subsets}\} \\ P(A) &= \#\{\text{points in } A\}/36, \quad \text{for all } A \in \mathcal{B}. \end{aligned}$$

Let X = sum of rolls and describe P_X by making a table of $P_X(X = x)$ for $x \in \mathcal{X}$

$\mathcal{X} = \{2, 3, 4, \ldots, 12\}$

x	2	3	4	5	6	7	8	9	10	11	12
$P_X(X=x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

$P_X(X=2) = P(\{\omega \in \Omega : X(\omega) = 2\}) = P(\{(1,1)\}) = 1/36$

$P_X(X=3) = P(\{\omega \in \Omega : X(\omega) = 3\})$

$= P(\{(1,2), (2,1)\}) = 2/36$

$P_X(X=4) = P(\{\omega \in \Omega : X(\omega) = 4\})$

$= P(\{(1,3), (2,2), (3,1)\}) = 3/36$

rv: random variable

Cumulative distribution function

The *cumulative distribution function (cdf)* F_X of an rv X is the function given by

$$F_X(x) = \boxed{P_X(X \leq x)} \text{ for all } x \in \mathbb{R}.$$

X: rv; x: specific value

$P_X(X \in (-\infty, x])$

Exercise: Let X = # times you get ⚅ in 2 rolls of a die.

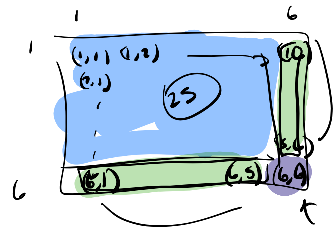
1. Give the cdf $F_X(x)$ for all $x \in \mathbb{R}$.
2. Draw a detailed picture of F_X.
3. Discuss interpretation of $F_X(1/2)$, say.
4. Discuss interpretation of jump sizes.
5. Discuss $F_X(x)$ for $x < 0$ and for $x \geq 2$

$\Omega = \{(1,1), (1,2), (2,1) \ldots\}$, $\mathcal{B} = \{\text{all subsets}\}$

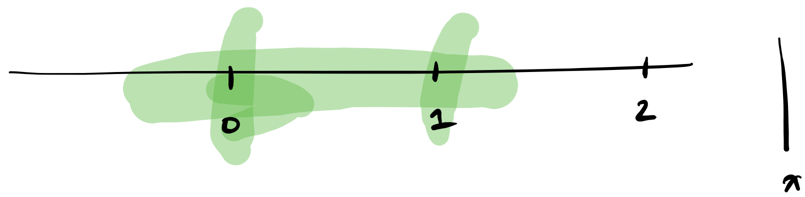
$P(A) = \frac{|A|}{36} \quad \forall A \in \mathcal{B}$

$\mathcal{X} = \{0, 1, 2\}$

x	0	1	2
$P_X(X=x)$	$\frac{25}{36}$	$\frac{10}{36}$	$\frac{1}{36}$



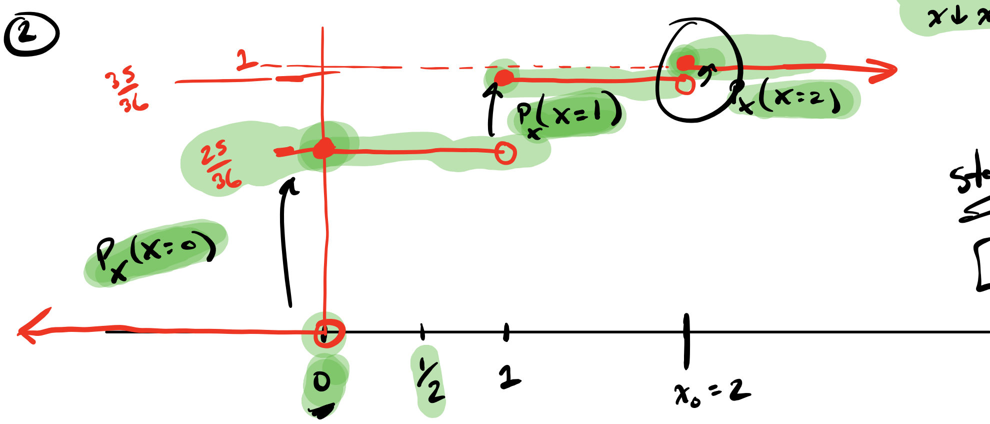
$$P_X(X=0) = P(\{\omega \in \Omega : X(\omega) = 0\}) = \frac{25}{36}$$



$$F_X(x) = P_X(X \le x) = \begin{cases} 0 & x < 0 \\ \frac{25}{36} & 0 \le x < 1 \\ \frac{35}{36} & 1 \le x < 2 \\ \frac{36}{36} = 1 & 2 \le x \end{cases}$$

$$\lim_{x \downarrow x_0} F_X(x) = F_X(x_0)$$

②



step function

[X is discrete]

③ $F_X(\frac{1}{2}) = P_X(X \le \frac{1}{2}) = \frac{25}{36}$

Theorem (Properties of a cdf)

The function F_X is a cdf if and only if

1. $F_X(x) \le F_X(y)$ *for all* $x \le y$. Nondecreasing
2. $\lim_{x\to-\infty} F_X(x) = 0$ *and* $\lim_{x\to\infty} F_X(x) = 1$.
3. $\lim_{x\downarrow x_0} F_X(x) = F_X(x_0)$ *for all* $x_0 \in \mathbb{R}$. "Right-continuous"

Exercise: Show that F_X is a cdf $\Longrightarrow$ F_X has the above properties.

1. Let $x \le y$. Then

$$F_X(x) = P_X(X \le x)$$
$$= P(\{\omega \in \Omega : X(\omega) \le x\})$$

$\{\omega \in \Omega : X(\omega) \le x\}$
$\subset \{\omega \in \Omega : X(\omega) \le y\}$

$A \subset B \Rightarrow P(A) \le P(B)$

$$= P(\{\omega \in \Omega : X(\omega) \le y\})$$

$$= P_X(X \le y)$$

$$= F_X(y) \implies \text{Nondecreasing}$$

② $\lim_{x \to \infty} F_X(x) = 1$:

Choose any increasing sequence $\{x_n\}_{n \ge 1}$ such that $\lim_{n \to \infty} x_n = \infty$.

Then define sets $\boxed{A_n = (-\infty, x_n]}$ for $n \ge 1$.

Then $A_n \subset A_{n+1} \subset A_{n+2}, \ldots$ so $\{A_n\}_{n \ge 1}$ is an increasing seq. of sets.

$$\left[\bigcup_{n=1}^{\infty} A_n = (-\infty, \infty) \right]$$

We have

$$\lim_{n \to \infty} F_X(x_n) = \lim_{n \to \infty} P_X(X \le x_n)$$

$$= \lim_{n \to \infty} P_X(X \in (-\infty, x_n])$$

$$= \lim_{n \to \infty} P_X(X \in A_n)$$

$$= \lim_{n \to \infty} P(\{\omega \in \Omega : X(\omega) \in A_n\})$$

Continuity of $P(\cdot)$

$$= P(\lim_{n \to \infty} \underbrace{\{\omega \in \Omega : X(\omega) \in A_n\}}_{\text{inc. seq. of sets because } A_n \subset A_{n+1} \subset A_{n+2} \subset \cdots})$$

lim of inc. seq. of sets is union

$$= P\left(\bigcup_{n=1}^{\infty} \{\omega \in \Omega : X(\omega) \in A_n\}\right)$$

$$= P(\{\omega \in \Omega : X(\omega) \in \underset{(-\infty,\infty)}{\underset{\uparrow}{\mathbb{R}}}\})$$

$$= P(\Omega)$$

$$= 1.$$

$\lim_{x \to -\infty} F_X(x) = 0$ [Show similarly]

③ $\lim_{x \downarrow x_0} F_X(x) = F_X(x_0)$.

Choose any decreasing seq $\{\varepsilon_n\}_{n \geq 1}$ such that $\lim_{n\to\infty} \varepsilon_n = 0$.

Set $A_n = (-\infty, x_0 + \varepsilon_n]$, $n \geq 1$ so $A_n \supset A_{n+1} \supset A_{n+2}$, dec seq.

Now write

$$\lim_{n\to\infty} F_X(x_0 + \varepsilon_n) = \lim_{n\to\infty} P_X(X \leq x_0 + \varepsilon_n)$$

$$= \lim_{n\to\infty} P_X(X \in A_n)$$

$$= \lim_{n\to\infty} P(\{\omega \in \Omega : X(\omega) \in A_n\})$$

Cont. of $P(\cdot)$ ↓

$$= P(\lim_{n\to\infty} \underbrace{\{\omega \in \Omega : X(\omega) \in A_n\}}_{\text{decreasing seq. of sets}})$$

$$= P(\bigcap_{n=1}^{\infty} \{\omega \in \Omega : X(\omega) \in A_n\})$$

$$= P(\{\omega \in \Omega : X(\omega) \le x_0\})$$

$$= P_X(X \le x_0)$$

$$= F_X(x_0).$$

Exercise: Let $X = \#$ free throw attempts required for you make one. Assume the attempts are independent with success probability $p \in (0, 1)$.

1. Give the support $\mathcal{X}$ of X.
2. Begin tabulating the values $P_X(X = x)$ for each $x \in \mathcal{X}$.
3. Find an expression which gives $P_X(X = x)$ for any $x \in \mathcal{X}$.
4. Find an expression for $F_X(x) = P_X(X \leq x)$ for any $x \in \mathcal{X}$.
5. Draw a picture of F_X when $p = 1/2$.
6. Verify that F_X has the three properties of a cdf.

Continuous and discrete random variables

A random variable X with cdf F_X is called a

1. *continuous rv* if $F_X(x)$ is a continuous function of x
2. *discrete rv* if $F_X(x)$ is a step function of x.
3. *mixture rv* if F_X has jumps and increasing continuous parts.

For a continuous rv X, for any $a, b \in \mathbb{R}$, $a < b$, we have

$$P_X(a < X < b) = P_X(a \leq X \leq b) = P_X(a < X \leq b) = P_X(a \leq X < b) = F_X(b) - F_X(a).$$

Exercise: Show that if X is continuous, $\boxed{P_X(X = x) = 0}$ for all x.

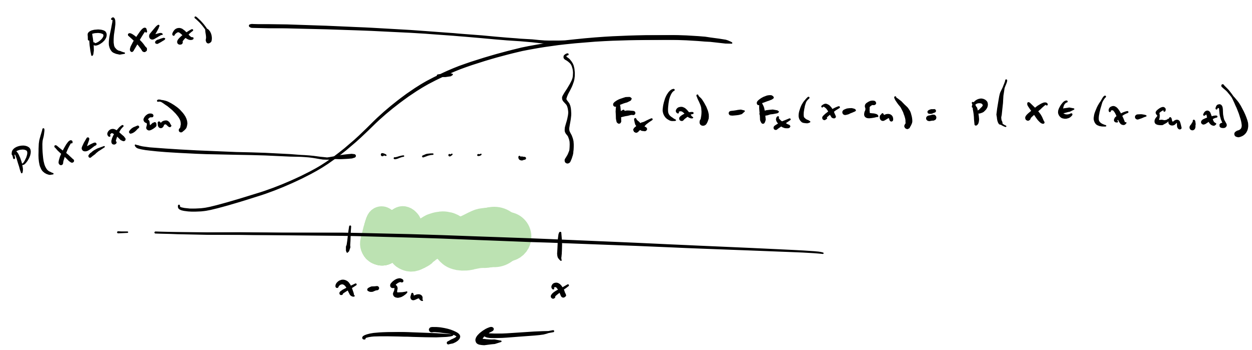
Choose any seq $\{\varepsilon_n\}_{n \geq 1} \downarrow 0.$

Then $\{x\} \subset (x-\varepsilon_n, x]$ for all $n \geq 1$.

$$P_X(X=x) \leq P_X(X \in (x-\varepsilon_n, x])$$
$$\vdots$$
$$= F_X(x) - F_X(x-\varepsilon_n) \quad \forall\, n \geq 1$$

$A \subset B$

$\Rightarrow P(A) \leq P(B)$



Then write

$$0 \leq P_X(X=x) \leq \lim_{n\to\infty} \left[F_X(x) - F_X(x-\varepsilon_n) \right]$$

probabilities always nonneg.

$$= F_X(x) - \lim_{n\to\infty} F_X(x-\varepsilon_n)$$

$$= F_X(x) - F_X\left(\lim_{n\to\infty} (x-\varepsilon_n)\right)$$

$$= F_X(x) - F_X(x)$$

$$= 0$$

$0 \leq P_X(X=x) \leq 0 \Rightarrow P_X(X=x) = 0.$

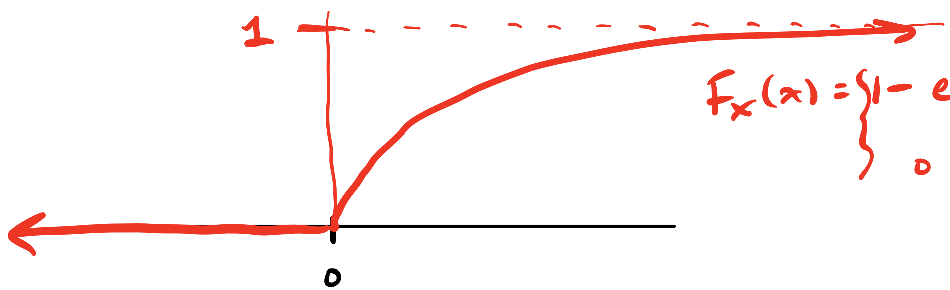
Exercise: Let $X =$ time (months) until you drop your new phone and suppose X has the cdf given by

$$F_X(x) = \begin{cases} 1 - e^{-x/10}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

1. Verify that F_X has the three properties of a cdf.
2. Use F_X to obtain $P_X(X \leq 1)$
3. Use F_X to obtain $P_X(2 < X)$
4. Use F_X to obtain $P_X(1 < X \leq 3)$

continuous.

1



$$F_X(x) = \begin{cases} 1 - e^{-x/10} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

(i) Nondecreasing? Yes

(ii) $\lim_{x \to -\infty} F_X(x) = 0$, $\lim_{x \to \infty} F_X(x) = 1$ Yes

(iii) $\lim_{x \downarrow x_0} F_X(x) = F_X(x_0)$ Yes

[Continuous $\Rightarrow$ Right continuous]

Identically distributed-ness of two random variables

Two rvs X and Y on the same probability space are called *identically distributed* if

$$P_X(X \in A) = P_Y(Y \in A)$$

for every $A \in \mathcal{B}(\mathbb{R})$. We write $X \overset{d}{=} Y$.

Example: The following random variables are identically distributed:

$$X = \#\text{ times you get ⚀ in 2 rolls of a die}$$
$$Y = \#\text{ times you get ⚁ in 2 rolls of a die}$$

Theorem (Identically distributed-ness result)

The following two statements are equivalent:

- $X \stackrel{d}{=} Y$
- $F_X(x) = F_Y(x)$ *for every* $x \in \mathbb{R}$.

So, two random variables are identically distributed if they have the same cdf.

Probability mass function of a discrete random variable

The *probability mass function (pmf)* p_X of a discrete rv X with probability distribution P_X is defined as

$$\boxed{p_X(x)} = P_X(X = x) \text{ for all } x \in \mathbb{R}.$$

Exercise: Find the pmfs of the following rvs based on independent Bernoulli trials with success probability p:

1. $X = 1$ if first trial a success, $X = 0$ if a failure.
2. ~~X = number of successes in 3 trials.~~
3. X = number of successes in n trials.
4. X = number of trial on which the first success occurs.

$\Omega = \{\text{success}, \text{failure}\}$

P

$P(\{\text{success}\}) = p$

(1)

$$X(\omega) = \begin{cases} 1 & \text{if } \omega = \text{success} \\ 0 & \text{if } \omega = \text{failure} \end{cases} \qquad \mathcal{X} = \{0,1\}$$

$$p_X(x) = P_X(X = x) \qquad \text{for all } x \in \mathbb{R}$$

$$= \begin{cases} p & x = 1 \\ 1-p & x = 0 \\ 0 & \text{otherwise} \end{cases}$$

$$p_X(x) = p^x (1-p)^{1-x} \, \mathbb{1}(x \in \{0,1\})$$

$\mathbb{1}(\cdot)$ the "indicator function".

$$\mathbb{1}(\text{statement}) = \begin{cases} 1 & \text{if statement true} \\ 0 & \text{if false} \end{cases}$$

(2) $X = \#$ successes in n trials

$\mathcal{X} = \{0, 1, \ldots, n\}$ $\qquad \Omega = \{ \boxed{\underbrace{S\,S\,F\,S\,F \ldots S}_{n}} \}$

x successes

$n - x$ failures

$$p_X(x) = \boxed{P_X(X = x)} = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{for } x = 0, 1, 2, \ldots, n \\ 0 & \text{otherwise.} \end{cases}$$

$$p_X(x) = \binom{n}{x} p^x (1-p)^{n-x} \, \mathbb{1}(x \in \{0,1,2,\ldots,n\})$$

For continuous rv $P_X(X=x)=0$ for all $x \in \mathbb{R}$.

Probability density function of a continuous random variable

The *probability density function (pdf)* f_X of a continuous rv X with cdf F_X is the function that satisfies

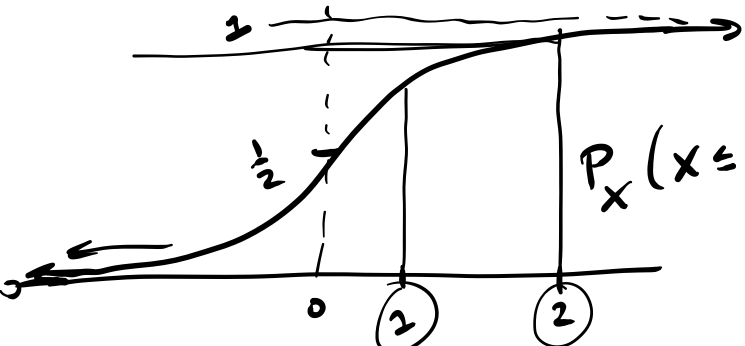
$$\underbrace{F_X(x)}_{P_X(X \le x)} = \int_{-\infty}^{x} \underbrace{f_X(t)}_{\text{pdf}} dt \text{ for all } x \in \mathbb{R}.$$

If f_X which satisfies the above is continuous, then $f_X(x) = \frac{d}{dx} F_X(x)$.

Note that the cdf of a discrete rv X with pmf p_X and support $\mathcal{X}$ can be written as

$$\boxed{F_X(x) = \sum_{\{t \in \mathcal{X} : t \le x\}} p_X(t).}$$

$$P_X(X \le x) = \sum_{\{t \in \mathcal{X} : t \le x\}} \boxed{P_X(X=t)}$$



$$P_X(X \le 0) = \frac{1}{1+\underbrace{e^{-0}}_{1}} = \frac{1}{2}$$

$$P(1 < X < 2) = P(X<2) - P(X<1)$$
$$= \frac{1}{1+e^{-2}} - \frac{1}{1+e^{-1}}.$$

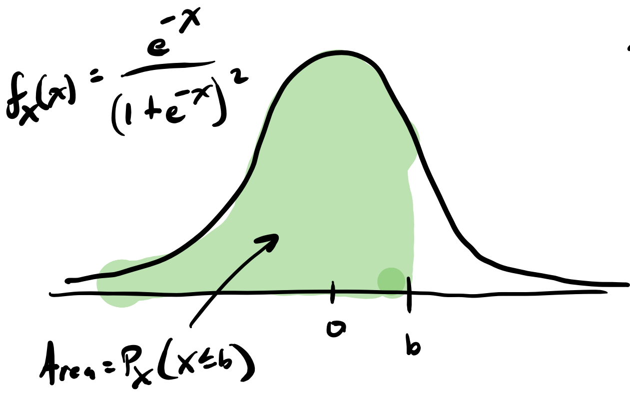
Exercise: Let X have cdf given by

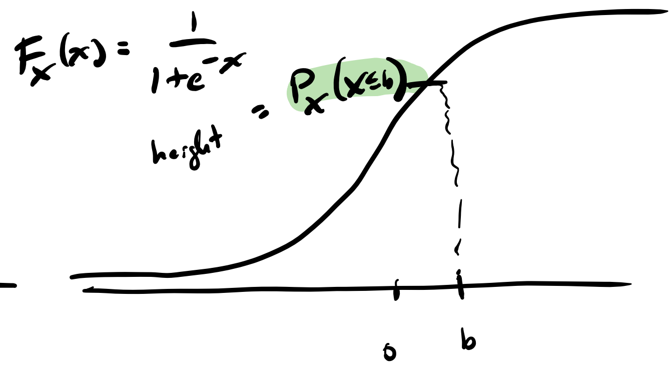
$$F_X(x) = \boxed{\frac{1}{1+e^{-x}}} \quad \text{for all } x \in \mathbb{R}.$$

Find the pdf of X.

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt$$

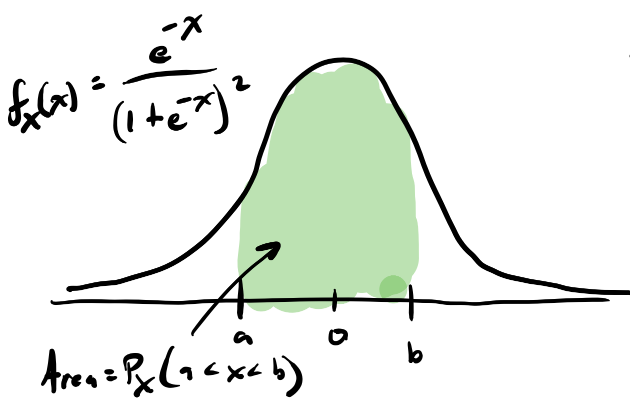
$$f_X(x) = \frac{d}{dx} F_X(x)$$
$$= -\frac{1}{(1+e^{-x})^2} e^{-x}(-1)$$
$$= \frac{e^{-x}}{(1+e^{-x})^2} \quad \text{for all } x \in \mathbb{R}.$$

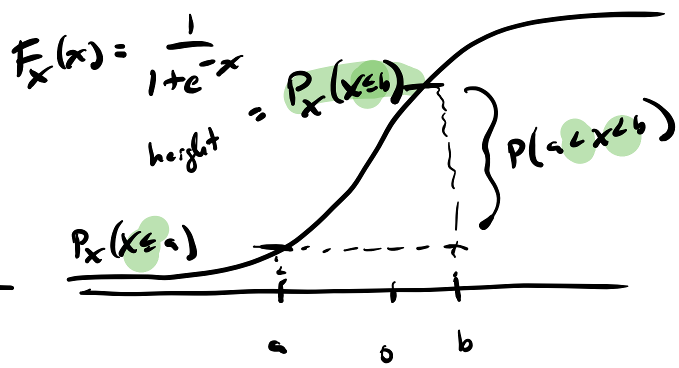




$$P_X(X \le b) = F_X(b) = \int_{-\infty}^{b} f_X(t)\,dt$$

$$P_X(a < X < b)$$





$$P_X(a < X < b) = F_X(b) - F_X(a) = \int_a^b f_X(t)\,dt$$

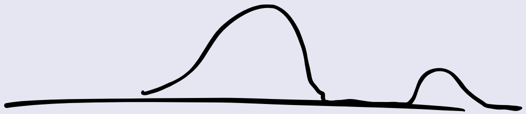
Theorem (Properties of pmfs and pdfs)

The function p_x *is a pmf if and only if*

- $p_X(x) \geq 0$ *for all* $x \in \mathbb{R}$
- $\sum_{x \in \mathcal{X}} p_X(x) = 1.$ $\quad [\, P(\Omega) = 1 \,]$

The function f_X *is a pdf if and only if*

- $f_X(x) \geq 0$ *for all* $x \in \mathbb{R}$
- $\int_{-\infty}^{\infty} f_X(x) dx = 1$



Exercise: Consider a discrete rv X with pmf given by

$$p_X(x) = \begin{cases} \dfrac{e^{-\lambda}\lambda^x}{x!}, & x = 0, 1, 2, \ldots \\ 0, & \text{otherwise} \end{cases}$$

for some $\lambda > 0$. Show that p_X is a valid pmf.

(i) $p_X(x) \geq 0$ ☑

(ii) $\sum_{x \in \mathcal{X}} p_X(x) = 1$ ☑ $\mathcal{X} = \{0, 1, 2, \ldots\}$

$$\sum_{x \in \{0,1,2,\ldots\}} \frac{e^{-\lambda} \lambda^x}{x!}$$

$$= \sum_{x=0}^{\infty} \frac{e^{-\lambda} \lambda^x}{x!}$$

$$= e^{-\lambda} \underbrace{\boxed{\sum_{x=0}^{\infty} \frac{\lambda^x}{x!}}}_{= e^{\lambda}}$$

$$= e^{-\lambda} e^{\lambda}$$

$$= 1.$$

Taylor expansion of e^x around $x = 0 = x_0$

$$\underset{\uparrow x_0}{f(x)} = \sum_{k=0}^{\infty} \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k \qquad \frac{d}{dx} e^x = e^x$$

$$e^x = \frac{e^0 (x-0)^0}{0!} + \frac{e^0}{1!} (x-0)^1 + \frac{e^0 (x-0)^2}{2!} \ldots$$

$$= \sum_{i=0}^{\infty} \frac{x^i}{i!}$$

Exercise: Consider a continuous rv X with pdf given by

$$f_X(x) = \begin{cases} c \cdot e^{-x/\lambda}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

for some $\lambda > 0$. Find the value c which makes f_X a valid pdf.

$c > 0$

To find c:

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

$$= \underbrace{\int_{-\infty}^{0} 0 \, dx}_{=0} + \underbrace{\int_{0}^{\infty} c\, e^{-x/\lambda} \, dx}$$

$$= c \left[-\lambda e^{-x/\lambda} \right] \Big|_0^{\infty}$$

$$= c \left[0 - \left(-\lambda e^{-0/\lambda} \right) \right]$$

$$= c\lambda$$

$$= 1$$

$$c = \frac{1}{\lambda}$$

$$f_X(x) = \begin{cases} \frac{1}{\lambda} e^{-x/\lambda} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

$$= \frac{1}{\lambda} e^{-x/\lambda} \, \mathbb{1}(x \geq 0).$$

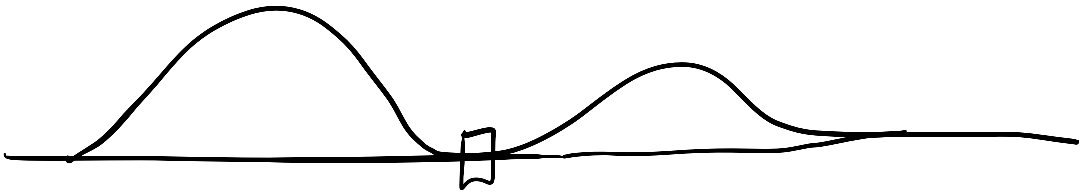
For a random variable X with probability distribution P_X, we often write

- $X \sim \boxed{F_X}$ if P_X has the cdf F_X
- $\boxed{X \sim p_X}$ if P_X has the pmf p_X
- $X \sim f_X$ if P_X has the pdf f_X

We can tell the support of an rv from its pdf or pmf.

The support $\mathcal{X}$ of a random variable X is given by

- $\{x \in \mathbb{R} : f_X(x) > 0\}$, if X is continuous with pdf f_X (wherever pdf is positive)
- $\{x \in \mathbb{R} : p_X(x) > 0\}$ if X is discrete with pmf p_X (wherever pmf is positive).



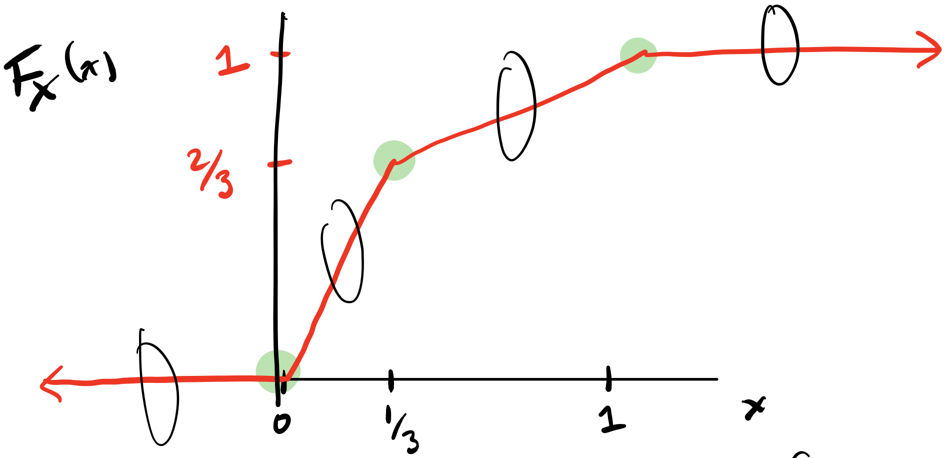
More on finding the pdf from the cdf of a continuous rv:

- If F_X has a continuous derivative F'_X, then $f_X = F'_X$.
- Otherwise set $f_X(x) = \frac{d}{dx} F_X(x)$ on intervals over which F_X is differentiable.

Exercise: Let X be a continuous rv with cdf given by

$$F_X(x) = \begin{cases} 0, & -\infty < x < 0 \\ 2x, & 0 \leq x < 1/3 \\ 2/3 + 1/2(x - 1/3), & 1/3 \leq x < 1 \\ 1, & 1 \leq x < \infty. \end{cases}$$

1. Draw a picture of F_X.
2. Find the pdf f_X of the rv X
3. Draw a picture of f_X.



$$f_X(x) = \begin{cases} 0 & -\infty < x < 0 \\ 2 & 0 \le x < 1/3 \\ 1/2 & 1/3 \le x < 1 \\ 0 & 1 \le x \end{cases}$$