STAT 345/445 Lecture 6

Section 1.5 - 1.6: Cdf, pdf, pmf

- **Cumulative distribution functions** cdf $F(x)$
- **Probability density functions** pdf $f(x)$
- **Probability mass functions** pmf $f(x)$

Cumulative distribution function

dummy variable

$F(t) = P(X \le t)$

Definition

The **cumulative distribution function (cdf)** of a random variable X is defined as

$$F_X(x) = P_X(X \le x) \qquad \forall x \in \mathbb{R}$$

Just to remind us that this is the cdf for X

- Note: $F_X(x)$ is defined for all $x \in \mathbb{R}$

lower case x: the argument of the function / dummy variable / outcome.

- Note: The cdf is defined the same way for both a discrete and a continuous random variable
  - If X is discrete: $F(x)$ is a step-function
  - If X is continuous: $F(x)$ is a continuous function

Cumulative distribution function

A minor technicality

- The textbook defines random variables as discrete or continuous depending on whether the cdf is a step function or a continuous function, respectively.

- We defined random variables as discrete or continuous depending on whether the range is countable or uncountable, respectively.

- Either one works

- My opinion: thinking about the possible outcomes of X makes more intuitive sense

Example – choosing class representatives

- Have a class of 40 students, 25 of which are Stat majors.
- Let X denote the number of Stat majors in a randomly chosen task force of three students.

Note that

$$P_X(X=0) = \frac{\binom{15}{3}}{\binom{40}{3}} = 0.046 \qquad P_X(X=1) = \frac{\binom{15}{2}\binom{25}{1}}{\binom{40}{3}} = 0.266$$

$$P_X(X=2) = \frac{\binom{15}{1}\binom{25}{2}}{\binom{40}{3}} = 0.455 \qquad P_X(X=3) = \frac{\binom{25}{3}}{\binom{40}{3}} = 0.233$$

Find and sketch the cdf of X.

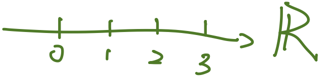
$F(x) = P(X \leq x)$

Example - continued

Only 4 possible outcomes : $0, 1, 2, 3$

If $x<0$, then $F(x) = P(X \le x) = 0$

For $x=0$: $F(0) = P(X \le 0) = P(X=0)$



$= 0.046$

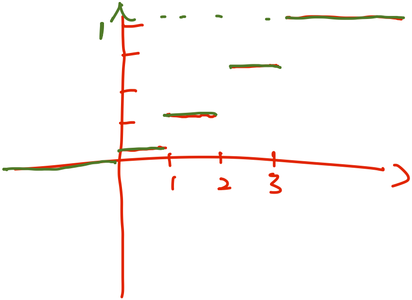
If $x \in (0,1)$,

$F(x) = P(X=0) = 0.046$.

For $x=1$

$F(1) = P(X \le 1) = P(X=0) + P(X=1) = 0.046 + 0.266 = 0.312$ etc...

$$F(x) = \begin{cases} 0 & x < 0 \\ 0.046 & 0 \le x < 1 \\ 0.312 & 1 \le x < 2 \\ 0.767 & 2 \le x < 3 \\ 1 & x \geq 3. \end{cases}$$



Properties of a cdf

Theorem 1

A function $F(x)$ is a cdf if and only if the following 3 conditions hold:

(i) $\lim\limits_{x \to -\infty} F(x) = 0$ and $\lim\limits_{x \to \infty} F(x) = 1$

(ii) $F(x)$ is a non-decreasing function of x

(iii) $F(x)$ is right continuous. That is, $\forall x_0 \in \mathbb{R}$ we have

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

Some notes on Theorem 1

- $F(x)$ is a cdf **if and only if** (i) - (iii) are true
  - Goes both ways
- $F(x)$ is a cdf $\quad\Rightarrow\quad$ (i) - (iii) are true
  - If $F(x)$ is a cdf then (i) - (iii) are true
  - (i) - (iii) are a *necessary* condition for $F(x)$ being a cdf
  - Not very hard to prove
- (i) - (iii) are true $\quad\Rightarrow\quad$ $F(x)$ is a cdf
  - If (i) - (iii) are true then $F(x)$ is a cdf
  - (i) - (iii) are a *sufficient* condition for $F(x)$ being a cdf
  - Hard to prove - *but very useful!*

To show that some function $F(x)$ is a cdf for some random variable all we need to do is to verify (i) - (iii)

Proof of "$\Rightarrow$" part of Theorem 1

Part (ii)

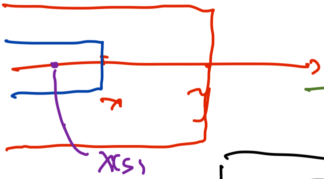
Suppose $F(x)$ is a cdf, want to prove that (i) - (iii) hold.

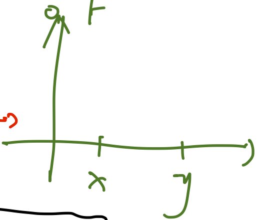
Start with (ii): Non-decreasing: ie show that if $x \le y$ then $F(x) \le F(y)$

$P(X \le y)$

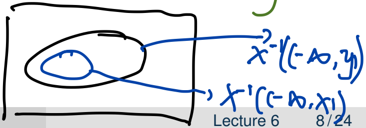
$F(x) = P(X \le x)$

$= P(\{s \in S : X(s) \le x\}) \cdots ①$





$\{s \in S : X(s) \le x\} \subset \{s \in S : X(s) \le y\}$



Therefore ① $\leq P(\{s \in S : X(s) \leq y\}) = F(y)$

Proof of "⇒" part of Theorem 1

For (i) and (iii) we need two results:

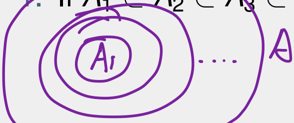
Continuity property of $P(\cdot)$

Let $A_1, A_2, A_3, \ldots \in \mathcal{B}$

set limit.

1. If $A_1 \subset A_2 \subset A_3 \subset \cdots$ and $A = \bigcup_{i=1}^{\infty} A_i$ then

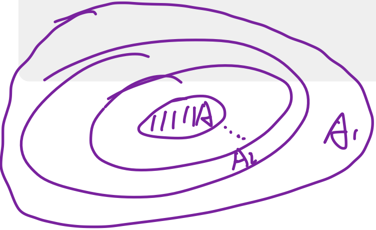
limit of numbers



$$\lim_{i \to \infty} P(A_i) = P(A)$$

2. If $A_1 \supset A_2 \supset A_3 \supset \cdots$ and $A = \bigcap_{i=1}^{\infty} A_i$ then

$$\lim_{i \to \infty} P(A_i) = P(A)$$



Proof of "⇒" part of Theorem 1

Parts (i) and (iii)

Similar argument works for * with a decreasing seq.

(iii): $\lim_{x \to -\infty} F(x) = 0$ * and $\lim_{x \to \infty} F(x) = 1$

Let x_n, $n = 1, 2, 3 \ldots$ be a monotone sequence of numbers in $\mathbb{R}$ such that

$$\lim_{n \to \infty} x_n = \infty$$

Show that

$\lim_{n \to \infty} F(x_n) = 1 \longrightarrow$ Since $\{x_n\}$ is an artifary sequence we get that $\lim_{x \to \infty} F(x) = 1$

$$\lim_{n\to\infty} F(x_n) = \lim_{n\to\infty} P(X\le x_n) = \lim_{n\to\infty} P(\underbrace{\{s: X(s)\le x_n\}}) \cdots\cdots (*)$$

$x_1 \quad x_2 \quad x_3 \cdots\cdots \longrightarrow \infty$

monotone increasing seq of sets

$$\bigcup_{n=1}^{\infty} \{s: X(s)\le x_n\} = S$$

$$(*) = P(X\in\mathbb{R}) = P(S) = 1$$

(iii) Cont. from the right

Let x_n be a monotone decreasing sequence such that

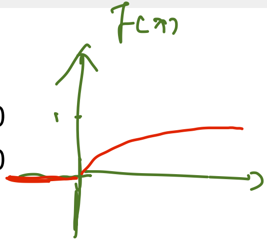
$\lim_{n\to\infty} x_n = x_0$ $\quad \{s: X(s) \le x_1\} \supset \{s: X(s) \le x_2\} \supset \cdots\cdots$

and $\bigcap_{n=1}^{\infty} \{s: X(s) \le x_n\} = \{s: X(s) \le x_0\}$

$$\Rightarrow \underbrace{\lim_{n\to\infty} P(\{s: X(s) \le x_n\})}_{\lim_{n\to\infty} F(x_n)} = \underbrace{P(\{s: X(s) \le x_0\})}_{= F(x_0)}$$

$$\Rightarrow \lim_{x \downarrow x_0} F(x) = F(x_0)$$

Example of a cdf



Show that this function is a cdf

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

(i) $$\lim_{x \to -\infty} F(x) = \lim_{x \to -\infty} 0 = 0$$

$$\lim_{x \to \infty} F(x) = \lim_{x \to \infty} 1 - e^{-x} = 1 - 0 = 1$$

(ii) If $a \leq b$ we have

$$e^{-a} \geq e^{-b} \;\Rightarrow\; 1 - e^{-a} \leq 1 - e^{-b} \;\Rightarrow\; F(a) \leq F(b)$$

(iii) $F(x)$ is continuous ($\Rightarrow$ also right continuous). Even at $x = 0$ we have

$$\lim_{x \downarrow 0} F(x) = \lim_{x \downarrow 0} = 1 - e^{-x} = 1 - 1 = 0 = F(0)$$

Identically distributed

Definition

Random variables X and Y are **identically distributed (id)** if for every $A \in \mathcal{B}^1$ we have

$$P(X \in A) = P(Y \in A)$$

e.g. $P(X \le a) = P(Y \le a)$

$\mathcal{B}^1$ = smallest σ-algebra generated by half-open intervals of $\mathbb{R}$

just a technicality

Theorem

The following are equivalent

1. X and Y are identically distributed
2. $F_X(t) = F_Y(t)$ for all $t \in \mathbb{R}$

Notation $X \stackrel{D}{=} Y$

the cdf uniquely determines the distribution.

Identically distributed does not necessarily mean equal

- Identically distributed is often denoted as $X \overset{D}{=} Y$
  - Equal in distribution

Don't write this

- $X \overset{D}{=} Y$ does NOT necessarily mean that $X = Y$

$X(s) \neq Y(s)$

for any s!

- Example: We throw one coin and let

$$X = \begin{cases} 1 & \text{if we get heads} \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad Y = \begin{cases} 1 & \text{if we get tails} \\ 0 & \text{otherwise} \end{cases}$$

  - In both cases the cdf is

$$F(t) = \begin{cases} 0 & \text{for } t < 0 \\ 0.5 & \text{for } 0 \leq t < 1 \\ 1 & \text{for } t \geq 1 \end{cases}$$

Probability mass function function

Definition

The **probability mass function (pmf)** of a *discrete* random variable is defined as

$$f(x) = P_X(X = x) \qquad \forall x \in \mathbb{R}$$

- If X is a discrete random variable with cdf $F(x)$ and pmf $f(x)$ then

$$\sum_{t \le x} P(X = t) = F(x) = \sum_{t \le x} f(t) \qquad \forall x \in \mathbb{R}$$

- Note that both $F(x)$ and pmf $f(x)$ are defined for all $x \in \mathbb{R}$

Probability density function

Definition

The **probability density function (pdf)** of a continuous random variable X is a function $f_X(x)$ that satisfies

$$P(X \leq x) = F_X(x) = \int_{-\infty}^{x} f_X(u)du \qquad \forall x \in \mathbb{R}$$

Index X is just to remind us what r.v. we are refering to

- Comparing to discrete case: Replaced the sum with an integral
- Generally use $F(x)$ for cdf and $f(x)$ for pmf and pdf

Properties of a pdf

Theorem

A function $f_X(x)$ is a pdf or pmf of a random variable X if and only if

(a) $f_X(x) \geq 0 \quad \forall x \in \mathbb{R}$

(b) $\sum_x f(x) = 1$ if X is discrete.

$\int_{-\infty}^{\infty} f(x)\,dx = 1$ if X is continuous.

If $f(x)$ is a pdf/pmf then (a) and (b) hold

If (a), (b) hold for some function $f(x)$ then $f(x)$ is a pdf/pmf.

To determine the distribution of a random variable

- To determine the probability distribution of a random variable it is sufficient to know either F or f
pdf or pmf cdf
- Say we know $f(x)$. Then we can determine $F(x)$
    - If X is discrete:
$$F(x) = \sum_{u \leq x} f(u)$$
    - If X is continuous:
$$F(x) = \int_{-\infty}^{x} f(u)du$$

To determine the distribution of a random variable

cdf

pdf or pmf

- Say we know $F(x)$. Then we can determine $f(x)$
  - If X is discrete:

$$f(x) = F(x) - \lim_{u \uparrow x} F(u)$$

equal to jumps in the cdf

  - If X is continuous and the derivative of F exists:

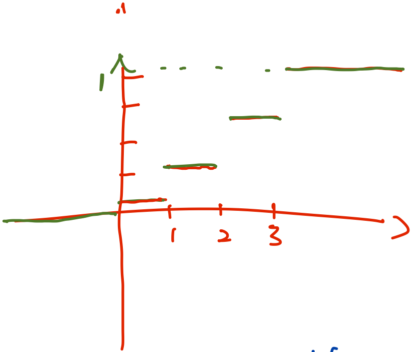
$$f(x) = \frac{d}{dx} F(x)$$

cdf for a discrete

- Notation: $X \sim F(X)$ or $X \sim f(x)$

cdf for a discrete random variable.

$$F(x) = \begin{cases} 0 & x < 0 \\ 0.046 & 0 \le x < 1 \\ 0.312 & 1 \le x < 2 \\ 0.767 & 2 \le x < 3 \\ 1 & x \ge 3. \end{cases}$$



$$f(1) = F(1) - \lim_{u \uparrow 1} F(u) = 0.312 - 0.046 = 0.266$$

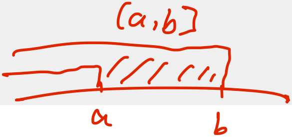
approaches from below

More on cdfs

Theorem

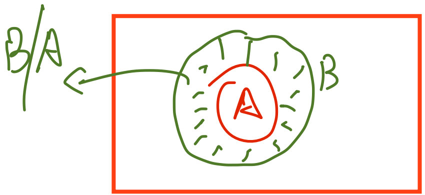
Let X be a random variable (discrete or continuous) with cdf F and let $a, b \in \mathbb{R}$ where $a < b$. Then

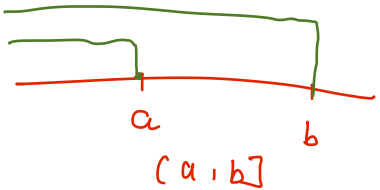
$$P(a < X \leq b) = F(b) - F(a)$$



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

$$P(a < X \leq b) = P(\{s \in S : a < X(s) \leq b\})$$





$A = \{s \in S : X(s) \le a\}$

$B = \{s \in S : X(s) \le b\}$ $\quad A \subset B$

$$P(B \setminus A) = P(B) - P(A \cap B)$$
$$= P(B) - P(A)$$

Note:

$$P(a \le X \le b)$$
$$= \underbrace{P(X \le b)}_{= F(b)} - \underbrace{P(X < a)}_{\text{may or may not be } F(a)}$$

F and f for continuous random variables

- Let X be a *continuous* random variable. Why did we not define the pdf the same way we defined the pmf, i.e. as $f(x) = P(X = x)$?

Because if X is continuous then $P(X=x)=0$

Consider the interval $(x-\varepsilon, x]$ for some $x \in \mathbb{R}$ and $\varepsilon > 0$

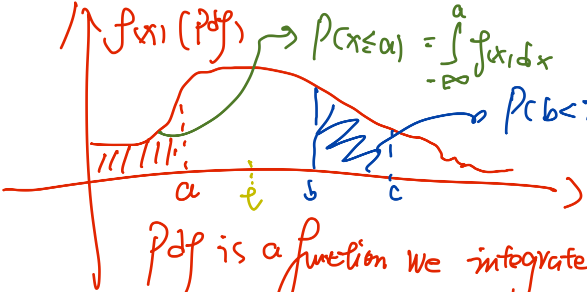
then $\lim_{\varepsilon \downarrow 0} (x-\varepsilon, x] = \{x\}$

$0 \le P(X=x) \le P(x-\varepsilon < X \le x)$ for all $\varepsilon > 0$

$\Rightarrow$ Therefore $P(X=x) \le \lim_{\varepsilon \downarrow 0} P(x-\varepsilon < X \le x) = \lim_{\varepsilon \to 0} F(x) - F(x-\varepsilon)$

$$= F(x) - \lim_{\varepsilon \downarrow 0} F(x-\varepsilon) = F(x) - F(x) = 0$$

when x is continuous,
the $F(x)$ is continous, (also from left)



$P(x \le a) = \int_{-\infty}^{a} f(x)\,dx$

$P(b < x < c) = \int_{b}^{c} f(x)\,dx$

$P(x = t) = \int_{t}^{t} f(x)\,dx = 0$

Pdf is a function we integrate over to get probabilities

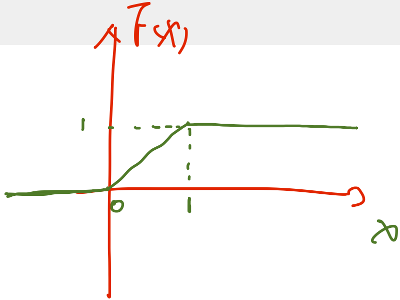
F and f for continuous random variables

- Let X be a *continuous* random variable. Then

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

Example

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } 0 \leq x \leq 1 \\ 1 & \text{if } x > 1 \end{cases}$$



1. Find the corresponding pdf
2. Find $P(0.3 < X < 0.5)$ using f and using F

$f(x) = \frac{d}{dx}F(x) = \frac{d}{dx}x = 1$ if $x \in [0,1]$

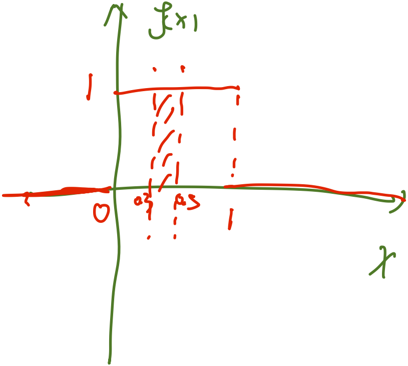
if $x<0$ then $f(x) = \frac{d}{dx}0 = 0$

if $x>1$ then $f(x) = \frac{d}{dx}1 = 0$

$\Longrightarrow f(x) = \begin{cases} 1 & x \in [0,1] \\ 0, & \text{otherwise} \end{cases}$

$\longrightarrow$ uniform$(0,1)$ distribution

② : $P(0.3 < X < 0.5)$



pdf $\int_{0.3}^{0.5} 1\,dx$

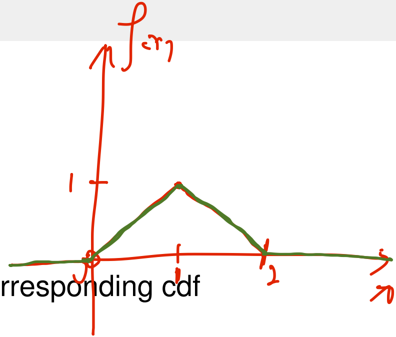
$= x \Big|_{0.3}^{0.5} = 0.5 - 0.3 = 0.2$

cdf $P(0.3 < X < 0.5) = F(0.5) - F(0.3) = 0.5 - 0.3$

$= 0.2$

Example

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



1. Show that f is a pdf and find the corresponding cdf

(a) $f(x) \geq 0 \quad \forall x$

Show that $\int_{-\infty}^{\infty} f(x)\,dx = 1$

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{0} 0\,dx + \int_{0}^{1} x\,dx + \int_{1}^{2} (2-x)\,dx + \int_{2}^{\infty} 0\,dx$$

$$= \frac{1}{2}x^2\Big|_0^1 + \left(2x - \frac{x^2}{2}\right)\Big|_1^2 = \frac{1}{2} - 0 + 4 - \frac{4}{2} - 2 + \frac{1}{2} = 1$$

cdf: $F(x) = \int_{-\infty}^{x} f(u)\,du$

for $x \le 0$: $F(x) = \int_{-\infty}^{0} 0\,du = 0$

for $x \in [0,1]$: $F(x) = \int_{-\infty}^{0} 0\,du + \int_{0}^{x} u\,du = \frac{u^2}{2}\Big|_0^x = \frac{x^2}{2}$

for $x \in [1,2]$: $F(x) = \int_{-\infty}^{0} 0\,du + \int_{0}^{1} u\,du + \int_{1}^{x} (2-u)\,du$

$$= \frac{1}{2} + 2x - \frac{1}{2}x^2 - 2 + \frac{1}{2}$$

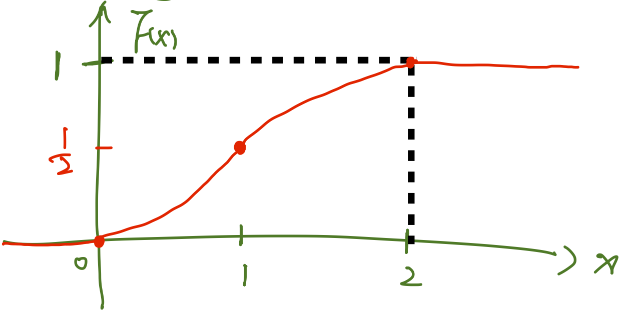
$$= -\frac{1}{2}x^2 + 2x - 1$$

for $x > 2$: $F(x) = 1$

$$F(x) = \int_{-\infty}^{0} 0\,du + \int_{0}^{1} u\,du + \int_{1}^{2} (2-u)\,du$$

$$+ \int_{2}^{x} 0\,du = 1$$

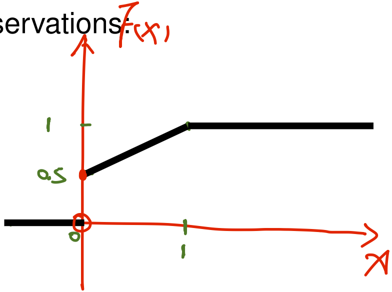
cdf:



Example: Partly discrete and partly continuous X

For example used to model truncated observations:

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{1}{2} + \frac{x}{2} & \text{if } 0 \le x \le 1 \\ 1 & \text{if } x \ge 1 \end{cases}$$



a "Point mass" at zero

Example: If negative outcomes are all recorded as 0.

This is actually a $\boxed{\text{mixture distribution.}}$

If $F_1(x)$ and $F_2(x)$ are both cdfs,

and if $0 \le \alpha \le 1$ then

$F(x) = \alpha F_1(x) + (1-\alpha) F_2(x)$ is also a cdf.

Here we have $F_1(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases}$

and $F_2(x) = \begin{cases} 0 & x < 0 \\ x & x \in [0,1] \\ 1 & x > 1 \end{cases}$

Then

$$F(x) = \frac{1}{2} F_1(x) + \frac{1}{2} F_2(x)$$

$$F(x) = \begin{cases} \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 0, & \text{if } x < 0 \\ \frac{1}{2} \cdot 1 + \frac{1}{2} x, & \text{if } x \in [0, 1] \\ \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot 1, & \text{if } x > 1 \end{cases}$$