Random variables

Cumulative distribution function (cdf)

Definition: The **cumulative distribution function (cdf)** of a random variable X, denoted by $F(\cdot)$, is defined by

$$F(x) = P(X \leq x), \quad \text{any } x \in \mathbb{R}$$

Sometimes written with subscript $F_X(x)$

A function $F(x)$, $x \in \mathbb{R}$, is a cdf for some random variable **if and only if** the following hold:

1. $F(x)$ is a nondecreasing function of x

2. $\lim_{x \to -\infty} F(x) = 0 \qquad \lim_{x \to \infty} F(x) = 1$

3. $F(x)$ is right continuous, i.e., $\lim_{x \downarrow x_0} F(x) = F(x_0)$ for any $x_0 \in \mathbb{R}$

   $\lim_{x \to x_0^+} F(x)$

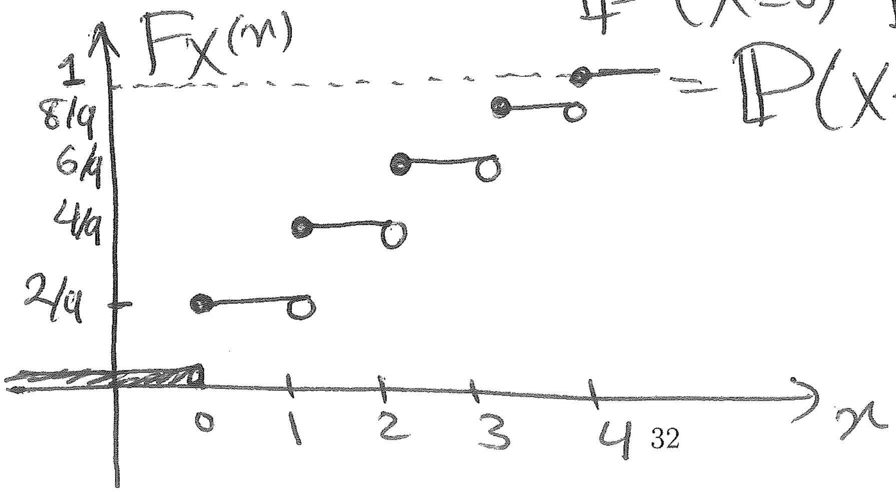
e.g., to show part 1., take $y \geq x$ and $F(y) - F(x) = P(X \leq y) - P(X \leq x)$.

$y \geq x \Rightarrow \{X \leq x\} \subset \{X \leq y\}$

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

Example of a cdf: 10 people seated randomly around a table; $X = \#$ of seats between $A \& B$

$P(X=0) = P(X=1) = P(X=2) = P(X=3) = 2/9$, $P(X=4) = 1/9$



Random variables

Discrete random variables & probability mass functions

- *Definition:* If a cdf F is a step function (with jumps at a countable collection of points $x_i \in \mathbb{R}$), then we say the distribution described by F is **discrete** (with support or range $x_i \in \mathbb{R}$).

  If a random variable X has a cdf $F = F_X$ which is a step function, then we say X is a **discrete random variable.**

  X is discrete (i.e. cdf F is step-function) If $\mathbb{P}(X \in \{x_i\}) = 1$ for some set $\{x_i\}$ of countable values in $\mathbb{R}$

- Besides the cdf, there are other (equivalent) ways to state the probability distribution for a discrete distribution/discrete r.v. X.

  1. The **probability mass function (pmf)** of a discrete random variable X is given by

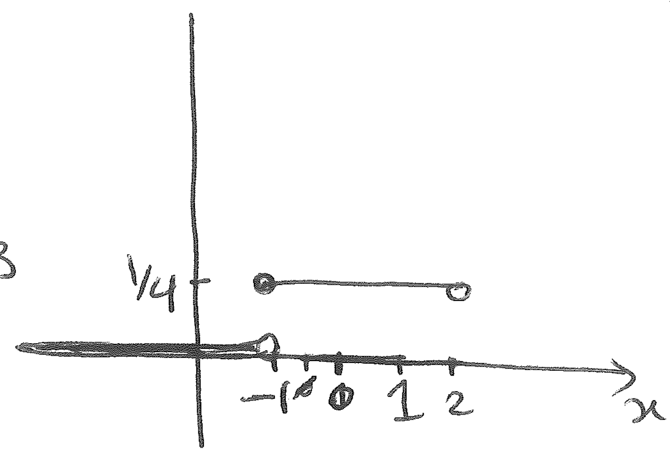
  $$f(x) = P(X = x) \geq 0, \quad \text{for any } x \in \mathbb{R}$$

  2. Equivalently, the pmf of a discrete r.v. X is

  $$f(x) = \underbrace{P(X \leq x) - P(X < x)}_{\mathbb{P}(X=x)} = F(x) - \lim_{y \to x^-} F(y)$$

e.g. random seating
X = # of seats between A & B

$$f(x) = \mathbb{P}(X=x) = \begin{cases} 2/9 & x = 0,1,2,3 \\ 1/9 & x = 4 \\ 0 & \text{o.w.} \end{cases}$$



Note that $F(x) = P(X \leq x) = \sum_{z \leq x} f(z) = \sum_{z \leq x, f(z) > 0} f(z)$

Random variables

Continuous random variables & probability density functions

- If a cdf F is such that there is a nonnegative function f such that

$$F(x) = \int_{-\infty}^{x} f(t)dt, \quad \text{for any } x \in \mathbb{R},$$

  then the distribution described by F is said to be (absolutely) **continuous** with **probability density function (pdf)** f.

  A random variable X with an (absolutely) continuous cdf F, or a pdf f, is said to be a **continuous random variable**.

- If F is (absolutely) continuous then its derivative at $x \in \mathbb{R}$ is its pdf $f(x)$

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

- If X is a continuous random variable then $P(X = x) = 0$ for any $x \in \mathbb{R}$.

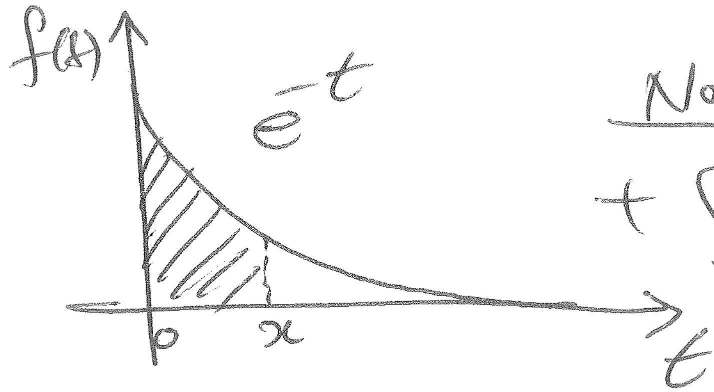
$$\mathbb{P}(a<X<b) = \mathbb{P}(a\le X\le b)$$
$$= \mathbb{P}(a\le X\le b)$$
$$= \mathbb{P}(a<X<b) = F(b)-F(a)$$
$$= \int_a^b f(t)dt$$

- Technical points: An absolutely continuous cdf has a derivative and so is more than just continuous. We *ignore* the distinction between a continuous & absolutely continuous cdf. Also, the derivative of F might not technically exist for *every* $x \in \mathbb{R}$ but this never matters (i.e., doesn't affect integrals); we just say $F'(x) = f(x)$ in any case.

Random variables

Continuous random variables & probability density functions (cont'd)

A continuous cdf example. Consider the function $f(x) = \begin{cases} e^{-x} & x > 0 \\ 0 & x \le 0 \end{cases}$

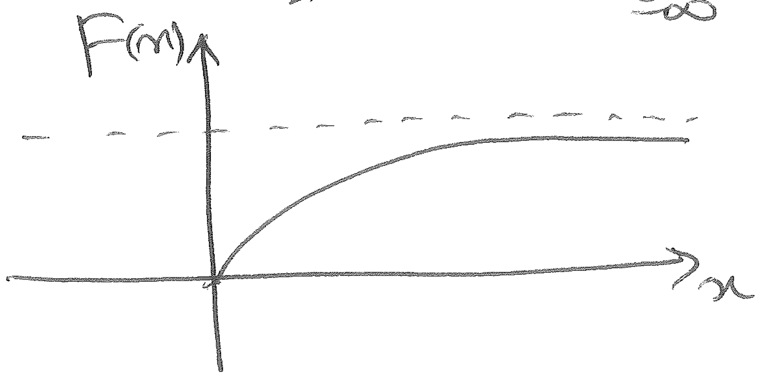


Note: $f(x) \geq 0 \ \forall x \in \mathbb{R}$

$+ \int_{-\infty}^{\infty} f(x)\,dx = \int_0^{\infty} e^{-x}\,dx = 1$

pdf $f \Longrightarrow$ cdf F

$$F(x) = \mathbb{P}(X \le x) = \int_{-\infty}^{x} f(t)\,dt = \begin{cases} \int_{-\infty}^{x} 0\,dt = 0 & \text{If } x \le 0 \\ \int_0^x e^{-t}\,dt = 1 - e^{-x} & \text{If } x > 0 \end{cases}$$



$\mathbb{P}(1 \le X \le 2)$

$= F(2) - F(1) = e^{-1} - e^{-2}$

$= \int_1^2 \underbrace{f(x)}_{e^{-x}}\,dx$

$\mathbb{P}(X = 2) = \int_2^2 f(x)\,dx = 0$

$$f(x) = \frac{dF(x)}{dx} = \begin{cases} \frac{d0}{dx} = 0 & \text{If } x < 0 \\ \frac{d(1-e^{-x})}{dx} = e^{-x} & \text{If } x > 0 \end{cases}$$

Random variables

Properties of probability density or mass functions

A function $f(x)$ is a pdf (or pmf) for some random variable if and only if

1. $f(x) \geq 0$ for any $x \in \mathbb{R}$
2. $\int_{-\infty}^{\infty} f(x)dx = 1 \quad (\text{or } \sum_x f(x) = 1)$

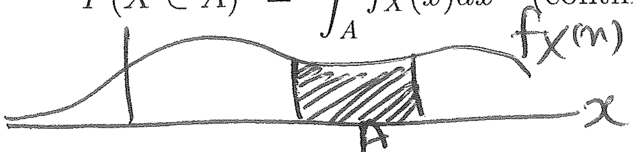
Any nonnegative function having a finite integral (or sum) can be turned into a pdf (or pmf) f by dividing by its integral (or sum)

We will write $X \sim f_X(x)$ (or $X \sim F_X(x)$) to denote that X has a distribution given by f (or F)

To find general probabilities using pmf/pdf, note that for $A \subset \mathbb{R}$,

$$P(X \in A) = \sum_{x \in A} f_X(x) = \sum_{x \in A, f_X(x) > 0} f_X(x) \quad \text{(discrete case using pmf)}$$

$$P(X \in A) = \int_A f_X(x)dx \quad \text{(continuous case using pdf)}$$



$\mathbb{P}(X \leq b)$ $\mathbb{P}(X \leq a)$

- *Discrete r.v. case:* Relating cdf to pmf

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

- *Continuous r.v. case:* Relating cdf to pmf

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

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

$$= \int_a^b f(x)\,dx$$

STAT 542: Summary to date
Where we have been
& where we are headed

- Completed
  - Intro to probability
    - * axioms and properties using set theory
    - * conditional probability and independence
  - Random variables
    - * definition
    - * discrete/continuous
    - * cdf, pdf/pmf
- Next
  - Transformations (an intro)
    "new" r.v. $Y = g(X)$ of "old" r.v. X
  - Expected values (mean, variance, moment generating function)
  - Probability-moment inequalities (Markov, Chebychev, Jensen)