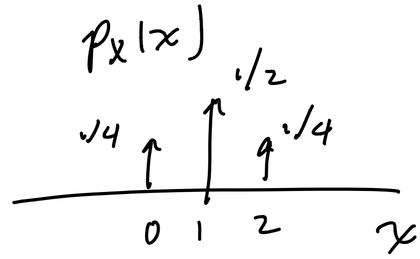
Cumulative Distribution Function (Chapter 4.1)
(Topic 2.2a)

Example: 2 fair coin tosses,

observe X = # heads



pmf: $$p_X(x) = \begin{cases} 1/4 & x=0 \\ 1/2 & x=1 \\ 1/4 & x=2 \\ 0 & \text{else} \end{cases}$$

what is $P(X \leq 1)$?

$$= \sum_{k=0}^{1} p_X(k) = \frac{1}{4} + \frac{1}{2} = \frac{3}{4}$$

In general, how do you compute $P(X \leq x)$

$$P(X \leq x) = \sum_{k=-\infty}^{x} p_X(k)$$

Note: this is a function of x because its value depends on where we set x.

This function $P(X \leq x)$ is so important we give it a name and its own notation.

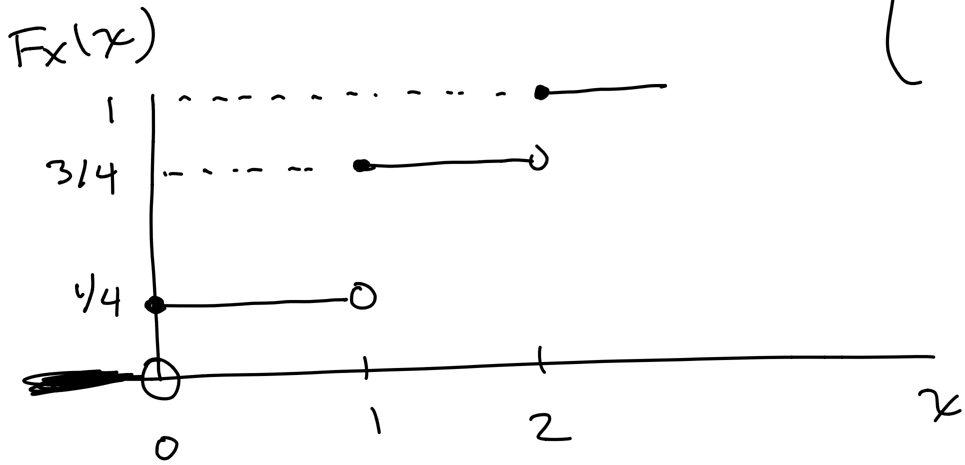
$$\boxed{F_X(x) = P(X \leq x) \quad \text{cumulative distribution function (cdf)}}$$

$\equiv$ the probability the RV X takes on a value in the interval $(-\infty, x)$

Note that in general, $P(X < x) \neq P(X \leq x)$

For the 2 fair coin tosses

cdf $F_X(x) = P(X \le x) = \begin{cases} 0 & x < 0 \\ 1/4 & 0 \le x < 1 \\ 3/4 & 1 \le x < 2 \\ 1 & x \ge 2 \end{cases}$



We can also express using the 'unit step function'

Recall $u(x) = \begin{cases} 0 & x < 0 \\ 1 & x \ge 0 \end{cases}$

$$F_X(x) = \frac{1}{4}u(x) + \frac{1}{2}u(x-1) + \frac{1}{4}u(x-2)$$

$$= p_X(0)\,u(x) + p_X(1)\,u(x-1) + p_X(2)\,u(x-2)$$

it's very easy to compute this by using the pmf for discrete RVs

Note $P(X \le 1) = 3/4$ but $P(X < 1) = 1/4$

Three types of RVs

Discrete RVs have a discrete sample space.
They have a pmf $p_X(x) = P(X=x)$
which is nonzero only for specific values of x.

Continuous RVs have a sample space that
is one interval or a union of intervals

Mixed RVs are a combination of continuous
and discrete

All 3 kinds of RVs have a CDF.
The CDF completely describes any probability
of interest for that RV.

Cumulative distribution function (cdf) (chapter 4.1)

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

Probability density function (pdf) (Chapter 4.2)

$$f_X(x) = \frac{d\, F_X(x)}{dx}$$

Note: mixed random variables can seem like a
mathematical oddity, but they are
<u>extremely</u> useful when building complicated
probability models of real world events.
We cannot just pretend they don't exist.

Continuous example

Spin an arrow attached to the center of a circular board. Let Θ be the angle where the arrow stops. $0 < \Theta \leq 2\pi$.

Define the probability Θ falls in any subinterval to be proportional to the length of the subinterval and let $X = \Theta / 2\pi$. Note $0 < X \leq 1$

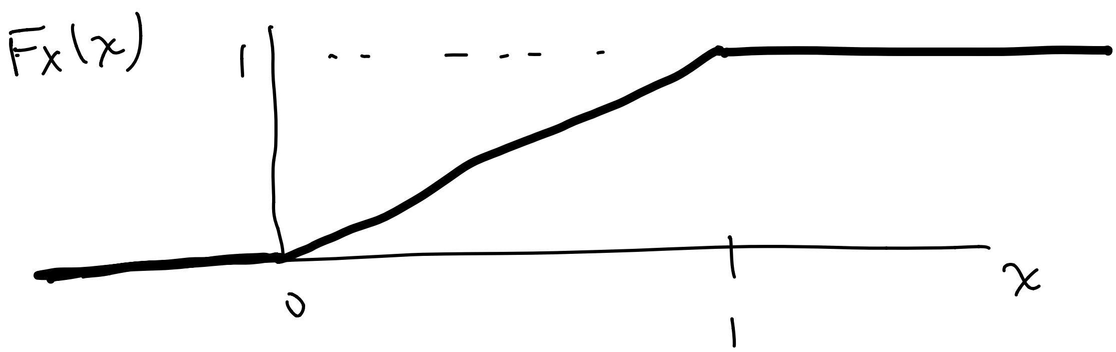
What is $F_X(x)$, the cdf of X Recall - find $F_X(x)$ for all x!

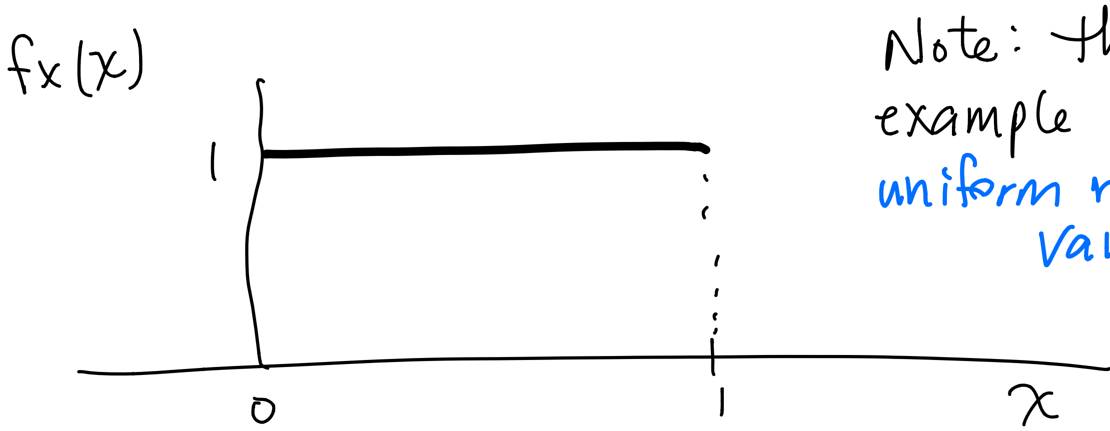
Case 1: $x < 0$: $F_X(x) = P(X \leq x) = P(\phi) = 0$

Case 2: $x > 1$: $F_X(x) = P(X \leq x) = P(S) = 1$

Case 3: $0 < x \leq 1$: $F_X(x) = P(X \leq x)$

$$= P(\Theta \leq 2\pi x) = \frac{2\pi x}{2\pi} = x$$





Note: this is an example of a uniform random variable

A mixed RV example X = wait time for a taxi

ⓐ if a taxi is available when you get to the taxi stand, the wait time $X = 0$
This happens with probability p.

ⓑ if no taxi is there, then wait time is uniformly distributed between $[0, 1]$ hours

What is $F_X(x)$, the cdf of X

$$F_X(x) = P(X \le x) = P(X \le x \mid \text{taxi there})\, P(\text{taxi there}) + P(X \le x \mid \text{no taxi})\, P(\text{no taxi})$$

by the theorem of ~~total~~ probability.

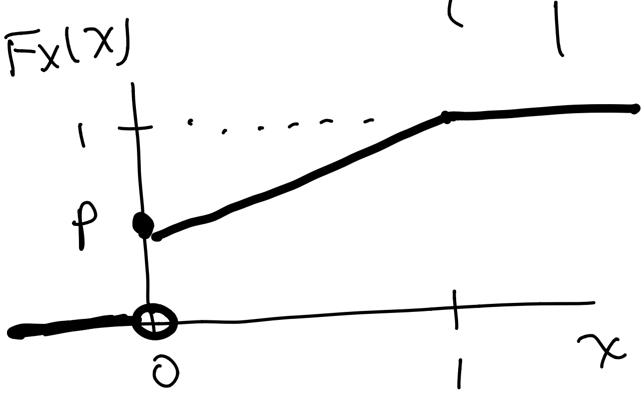
The problem statement defines the conditional probs.

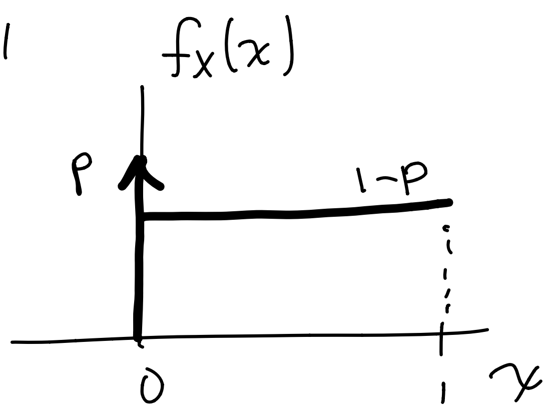
$$P(X \le x \mid \text{taxi there}) = \begin{cases} 1 & x \ge 0 \\ 0 & \text{else} \end{cases}$$

$$P(X \le x \mid \text{no taxi}) = \begin{cases} 0 & x < 0 \\ x & 0 \le x < 1 \\ 1 & x \ge 1 \end{cases}$$

(a uniform RV)

So
$$F_X(x) = \begin{cases} 0 & x < 0 \\ p + (1-p)x & 0 \le x < 1 \\ 1 & x \ge 1 \end{cases}$$





Summary: 3 types of random variables

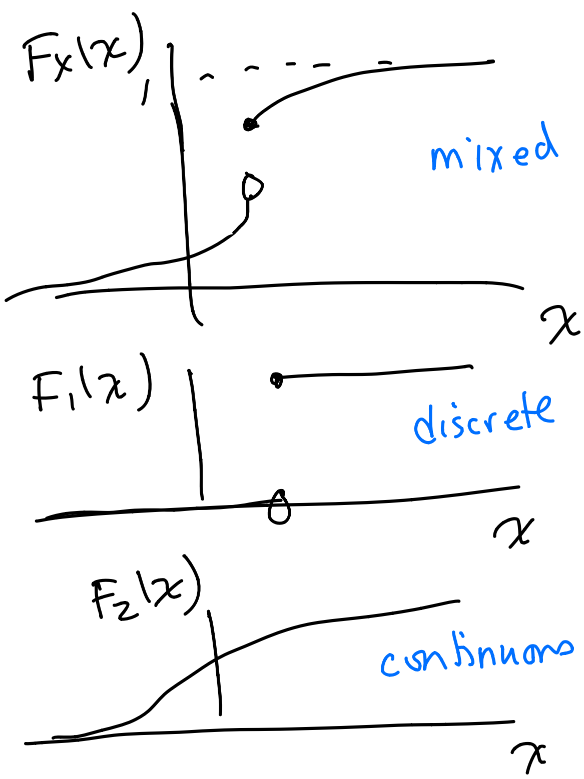
Discrete: $$F_X(x) = \sum_{x_k \le x} p_X(x_k)\, u(x - x_k)$$

a stair case that increases to the right

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

Every possible outcome has a probability zero, but probabilities of intervals are well-defined. Calculate probabilities as integrals of the probability density of intervals.

Mixed-type CDF jumps at a countable # points but also increases continuously in at least one interval



$$F_X(x) = p\, F_1(x) + (1-p)\, F_2(x)$$

for $0 < p < 1$

where $F_1(x)$ is the cdf of a discrete RV and $F_2(x)$ is a cdf of a continuous RV

Properties of CDFs $F_X(x) = P(X \le x)$

These all arise from the axioms of probability

1) $0 \le F_X(x) \le 1$ (it's a probability)

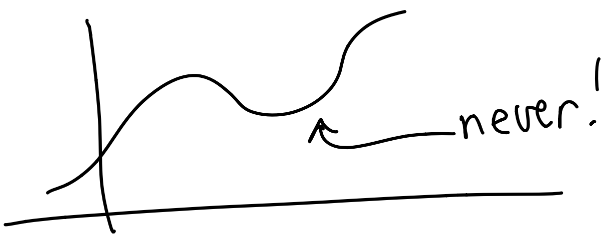
2) $\lim_{x \to \infty} F_X(x) = 1$ (as $x \to \infty$, $\{X \le x\} = S$)

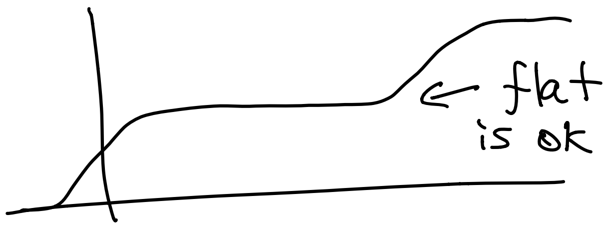
3) $\lim_{x \to -\infty} F_X(x) = 0$ (as $x \to -\infty$, $\{X \le x\} = \emptyset$)

4) $F_X(x)$ is nondecreasing

$F_X(a) \le F_X(b)$ if $a < b$

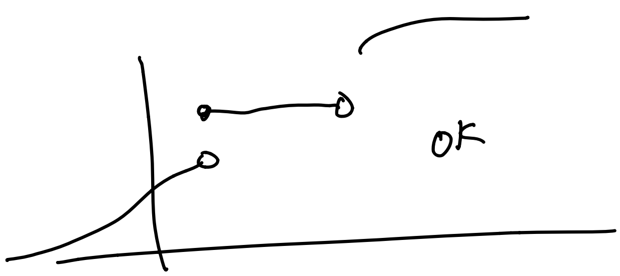
(as x increases, the set $\{X \le x\}$ cannot get smaller)

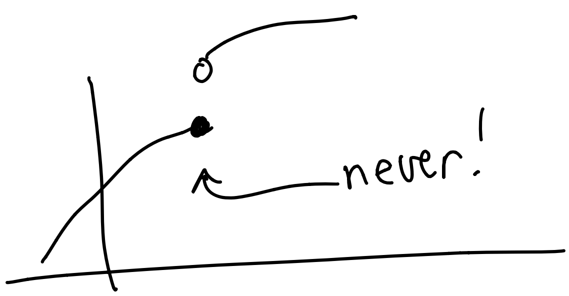




5) $F_X(x)$ is continuous from the right

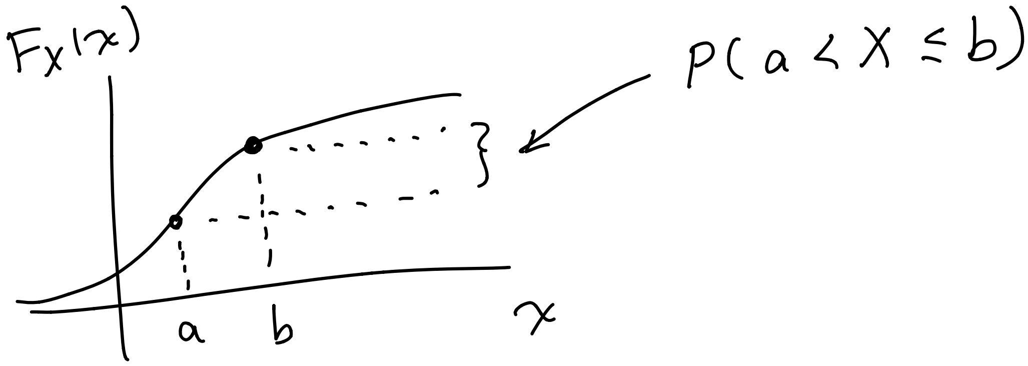
$F_X(b) = \lim_{h \to 0^+} F_X(b+h) = F_X(b^+)$ for $h > 0$





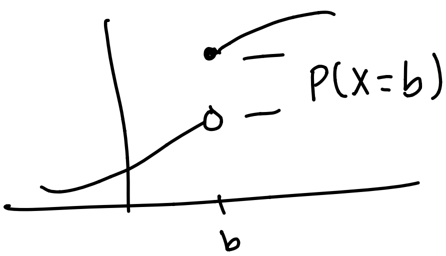
Computing probabilities

6) $P(a < X \le b) = F_X(b) - F_X(a)$



7) $P(X = b) = F_X(b) - F_X(b^-)$

$$= F_X(b) - \lim_{\substack{h \to 0 \\ h > 0}} F_X(b-h)$$



(the value at the open circle is $\lim_{h \to 0} F_X(b-h)$)

8) $P(X > x) = 1 - F_X(x)$

$$= P(\{X \le x\}^c) = 1 - P(X \le x)$$

Be careful of strict ($<$) and loose ($\le$) inequalities. IT MATTERS!