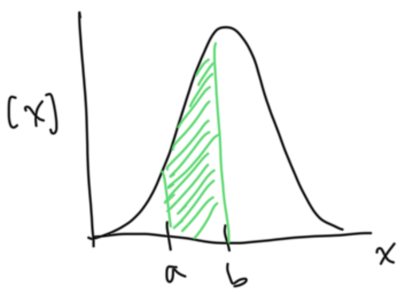
PROBABILITY BOARDWORK (part 2)

- we learned about discrete r.v.'s. Now we turn to second class:

Continuous random variables

- unlike discrete case, continuous r.v.'s can take on infinitely many values
  - e.g. mass, proportion, count / m^2
- the distribution of continuous X is characterized by its probability density function (PDF) $f_X(x) = [x]$



such that the following holds for all $a < b$:

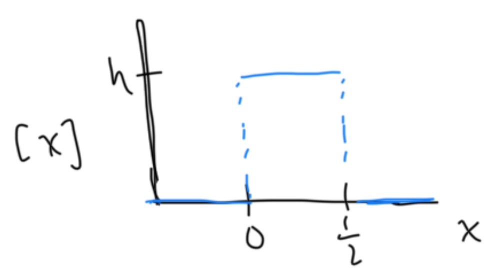
$$\Pr(a < X < b) = \int_a^b f_X(x)\,dx = \int_a^b [x]\,dx$$

- the support of X, S_X, is set of values x such that $[x] > 0$
- Like PMFs, PDF has two properties:
  1) $[x] \geq 0$
  2) $\int_{-\infty}^{\infty} [x]\,dx = \int_{S_X} [x]\,dx = 1$

this was a sum instead of integral for discrete X

- interpretation: area under density equals 1

- e.g. consider PDF $[x]$ that looks like this:



Q: what is S_X?

A: $S_X = [0, \frac{1}{2}]$

Q: what is value of h?

A: By property 2, need area under the "curve" to equal 1.

using geometry: $1 = (\frac{1}{2} - 0)h = \frac{1}{2}h \Rightarrow h = 2$

So, the PDF of X is $[x] = \begin{cases} 2 & \text{if } x \in [0, \frac{1}{2}] \\ 0 & \text{otherwise} \end{cases}$

* This motivates the following cautionary reminder: the PDF evaluated at a value is NOT a probability!!

  - i.e. if X is continuous, then $[a] \neq \Pr(X = a)$
  - we clearly see this from the example, since $[x] = 2$ for $x \in S_X$, and 2 is not a legal probability

  ↳ why not?

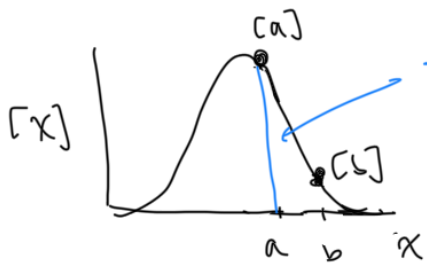
$$\Pr(X = a) = \Pr(a \le X \le a) = \int_a^a [x]\,dx = 0$$

i.e. prob. of X taking on any single value in its infinite support is 0!

- for this reason, $\Pr(a < X < b) = \Pr(a \le X \le b)$

** However, this doesn't mean the event $X = a$ is impossible!

Rationale: the probability in distribution of X is spread so thin, we can only "see" it on sets like non-degenerate intervals



the line has 0 area in 2D, but that doesn't mean the line isn't there!!

even though $[a]$ & $[b]$ are not probabilities, they still help tell us which values of X are more likely!

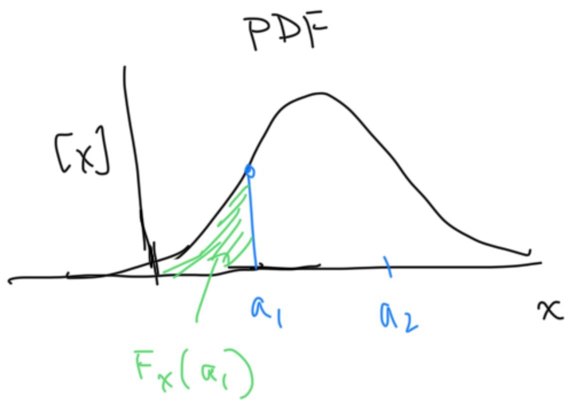
- in picture above, $[a] > [b]$ $\Rightarrow$ a is more likely than b

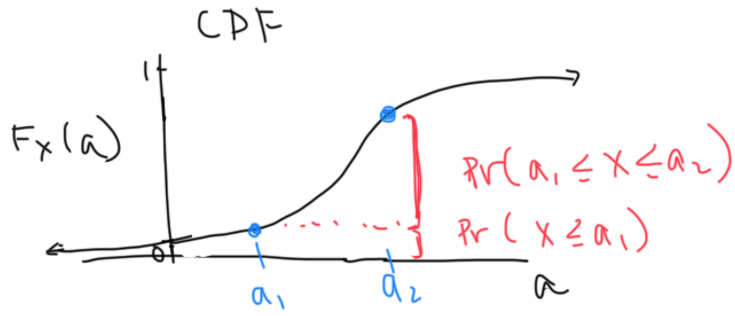
- okay... if I cannot talk about probabilities of single values for continuous r.v's, how are they useful?

- obtain probabilities by integrating PDF over some set of values in S_X. The CDF for continuous X evaluated at a is:

has same properties as discrete CDF!

$$F_X(a) = \Pr(X \le a) = \Pr(-\infty < X \le a) = \int_{-\infty}^{a} [x]\,dx$$





as we slide blue line to right, we just add area

- can also obtain summary values / moments:

$$\mathbb{E}[X] = \mu_X = \int_{-\infty}^{\infty} x[x]\,dx = \int_{S_X} x[x]\,dx \quad \leftarrow \text{another weighted average}$$

$$\mathrm{var}(X) = \sigma^2_X = \int_{-\infty}^{\infty} (x-\mu_X)^2[x]\,dx = \int_{S_X} (x-\mu_X)^2[x]\,dx$$

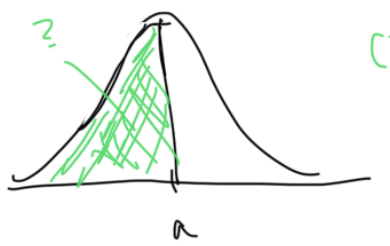
- continuous CDFs have a nicely behaved inverse function, which is called the inverse-CDF or quantile function of X

  - The CDF $F_X(a) = Pr(X \leq a) = p$ returns a probability

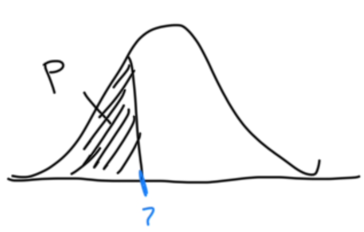
  - The quantile function $F_X^{-1}(p)$ returns value $a \in S_X$ such that

$$F_X(a) = p \qquad \text{(i.e. the } p\text{-th quantile)}$$

visually:



CDF returns area to left of a



inverse CDF returns value in S_X which has prob. p to left of it

- as an example: the median of a cont. distribution is the value m with 50% prob. to left & right.
  - i.e. median is 0.5th quantile

To find value of m, we use inverse CDF!

$$m = F_X^{-1}(0.5)$$