MATH 180A: INTRO TO PROBABILITY (FOR DATA SCIENCE)

www.math.ucsd.edu/~tkemp/180A

Today: §3.2, 2.4

Next: §2.5, 4.4

Lab 2.2 due TONIGHT!

HW3 due Friday.

Midterm next Wed evening

* Regrade requests for HW1:
  Regrade requests for Lab1:
  } window Tuesday 10/15 – Thursday 10/17, 8am – 11pm
  - ↳ separate request for each problem
  - ↳ detailed, polite responses please.

* HW3 Problem 8: changed 2.25 → 2.21

Cumulative Distribution Function (CDF)

For any random variable X, $F_X(r) = \mathbb{P}(X \le r)$. $r \in \mathbb{R}$

(1) Monotone increasing: $s \le t \Rightarrow F_X(s) \le F_X(t)$

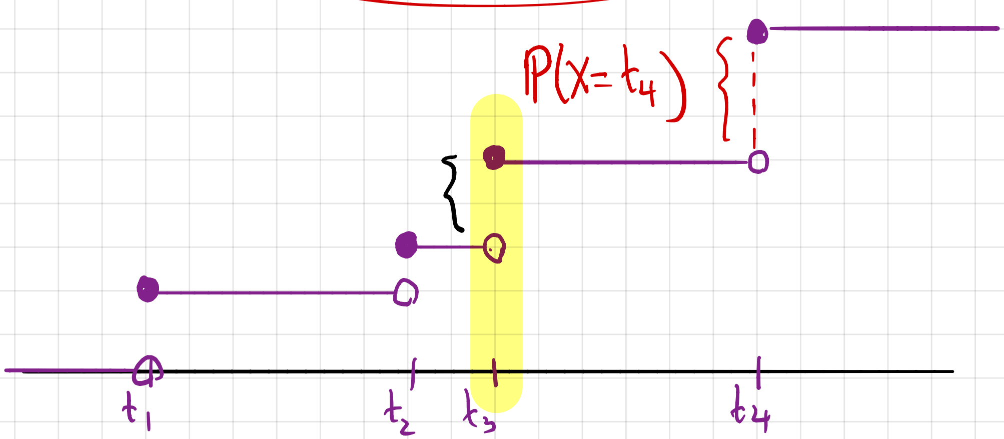
(2) $\lim\limits_{r \to -\infty} F_X(r) = 0$, $\lim\limits_{r \to +\infty} F_X(r) = 1$.

(3) The function F_X is right-continuous: $\lim\limits_{t \to r+} F_X(t) = F_X(r)$.

Discrete random variable:

finite or countable set of values $t_1, t_2, t_3, \ldots$ with $\mathbb{P}(X = t_j) > 0$ and

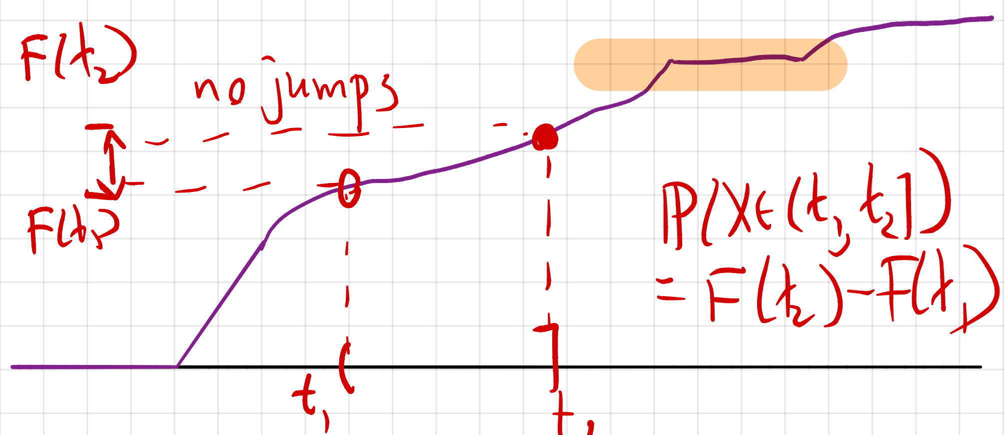
$$\sum_j \mathbb{P}(X = t_j) = 1.$$



Continuous random variable

for each real number t, $\mathbb{P}(X = t) = 0$.

Because of (1) & (3) above, this implies that F_X is continuous.



$\mathbb{P}(X \in (t_1, t_2]) = F(t_2) - F(t_1)$

Densities

Some continuous random variables have probability densities. This is an infinitesimal version of a probability mass function.

X discrete, $\in \{t_1, t_2, t_3, \ldots\}$

$p_X(t) = \mathbb{P}(X=t)$ probability mass function

$$\mathbb{P}(X \in A) = \sum_{t \in A} \mathbb{P}(X=t)$$

$$= \sum_{t \in A} p_X(t)$$

$p_X(t) \geq 0$, $\sum_t p_X(t) = 1$.

X continuous

$\mathbb{P}(X=t) = 0$ for all $t \in \mathbb{R}$.

BUT

Maybe there is an "infinitesimal" prob. mass function f_X.

$$\mathbb{P}(X \in A) = \int_A f_X(t)\,dt$$

i.e., $A = (-\infty, r]$

$$\mathbb{P}(X \leq r) = \int_{-\infty}^{r} f_X(t)\,dt$$

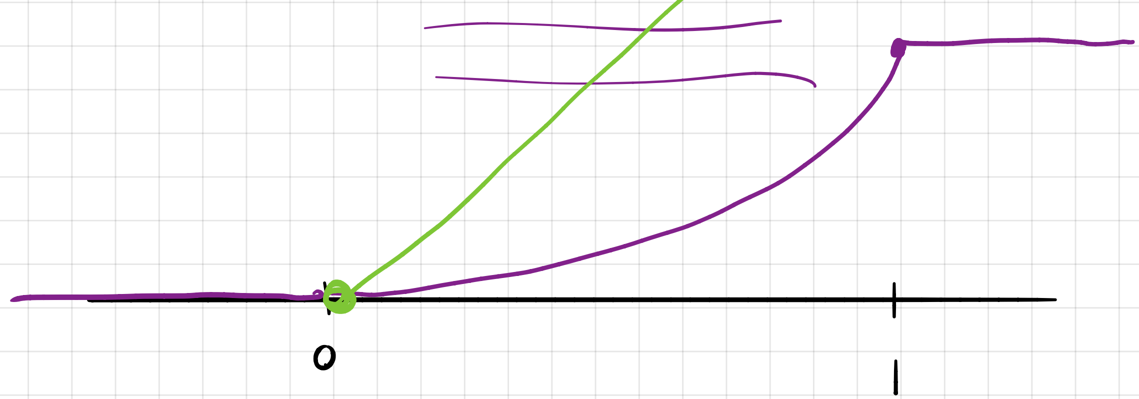
$$\mathbb{P}(X \in [a,b]) = \int_a^b f_X(t)\,dt$$

$\int_{-\infty}^{\infty} f_X(t)\,dt = 1$, $\quad f_X(t) \geq 0$.

Eg. Shoot an arrow at a circular target of radius 1.

Y = distance from center.

$$\int_{-\infty}^{r} f(t)\,dt \overset{?}{=} \mathbb{P}(Y \in (-\infty, r]) = F_Y(r) = \begin{cases} 0, & r < 0 \\ r^2, & 0 \le r \le 1 \\ 1, & r \ge 1 \end{cases}$$



"Solve for f"

$$\frac{d}{dr}\int_0^r f(t)\,dt = \frac{d}{dr}\begin{cases} r^2 & 0 \le r \le 1 \\ 1 & r > 1 \end{cases}$$

FTC $\quad f(r) = \begin{cases} 2r & 0 \le r \le 1 \\ 0 & r > 1 \end{cases}$

$$f_Y(r) = \begin{cases} 0 & r < 0 \\ 2r & 0 \le r \le 1 \\ 0 & r > 1 \end{cases}$$

$$\mathbb{P}(Y \in [0.1, 0.2] \cup [0.9, 1]) = \int_{0.1}^{0.2} 2r\,dr + \int_{0.9}^{1} 2r\,dr$$

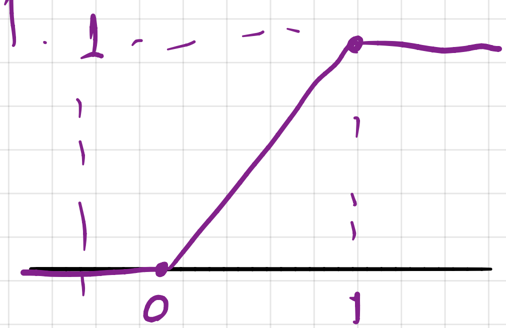
$(0.2)^2 - (0.1)^2$
$+ 1^2 - (0.9)^2$

Theorem: If F_X is continuous and piecewise differentiable, then X has a density $f_X = F_X'$.

Proof: FTC. $\boxed{}$

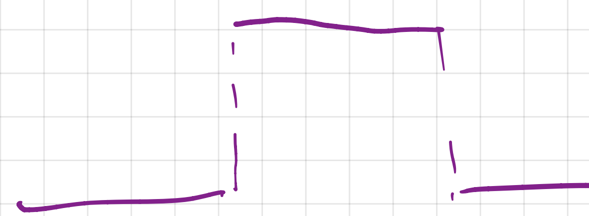
Eg. Let X = a uniformly random number in $[0,1]$.
As we discussed in lecture 2, this means

$$\longrightarrow \mathbb{P}(X \in [s,t]) = t-s \quad \text{if} \quad 0 \le s < t \le 1.$$



$$F_X(r) = \mathbb{P}(X \le r) = \begin{cases} 0 & r \le 0 \\ r-0 & 0 \le r \le 1 \\ 1 & r \ge 1 \end{cases}$$

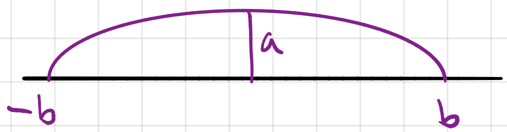
$$\therefore f_X(r) = F_X'(r) = \begin{cases} 0 & r < 0 \\ 1 & 0 \le r \le 1 \\ 0 & r > 1 \end{cases}$$



$X \sim \mathrm{Unif}([0,1])$

$Z \sim \mathrm{Unif}([a,b]) \longrightarrow f_Z(t) = \begin{cases} 0 & t < a \\ \frac{1}{b-a} & a \le t \le b \\ 0 & t > b \end{cases}$

Eg. Let $f(t) = c\sqrt{b^2-t^2}$ for $|t| \le b$, 0 otherwise (for some positive constants $b, c > 0$).



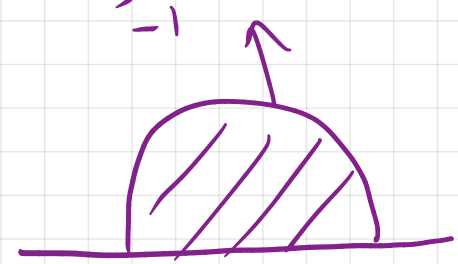
Is f a probability density?

- $f \ge 0$ ✓
- $\displaystyle\int_{-\infty}^{\infty} f(t)\,dt = \int_{-b}^{b} c\sqrt{b^2-t^2}\,dt = c\int_{-b}^{b}\sqrt{b^2-t^2}\,dt$

Subs: $t = bs$. $\displaystyle c\int_{-1}^{1}\sqrt{b^2-(bs)^2}\,b\,ds$

$$= cb\int_{-1}^{1}\sqrt{b^2(1-s^2)}\,ds$$

$$= cb^2\int_{-1}^{1}\sqrt{1-s^2}\,ds$$



$$\|$$

$$cb^2\frac{\pi}{2} = 1$$

Must have $cb^2 = \frac{2}{\pi}$.

Eg. For any b, $c = \frac{2}{\pi b^2}$. ✓

E.g. Your car is in a minor accident; the damage repair cost is a random number between \$100 and \$1500. Your insurace deductible is \$500. Z = your out of pocket expenses.

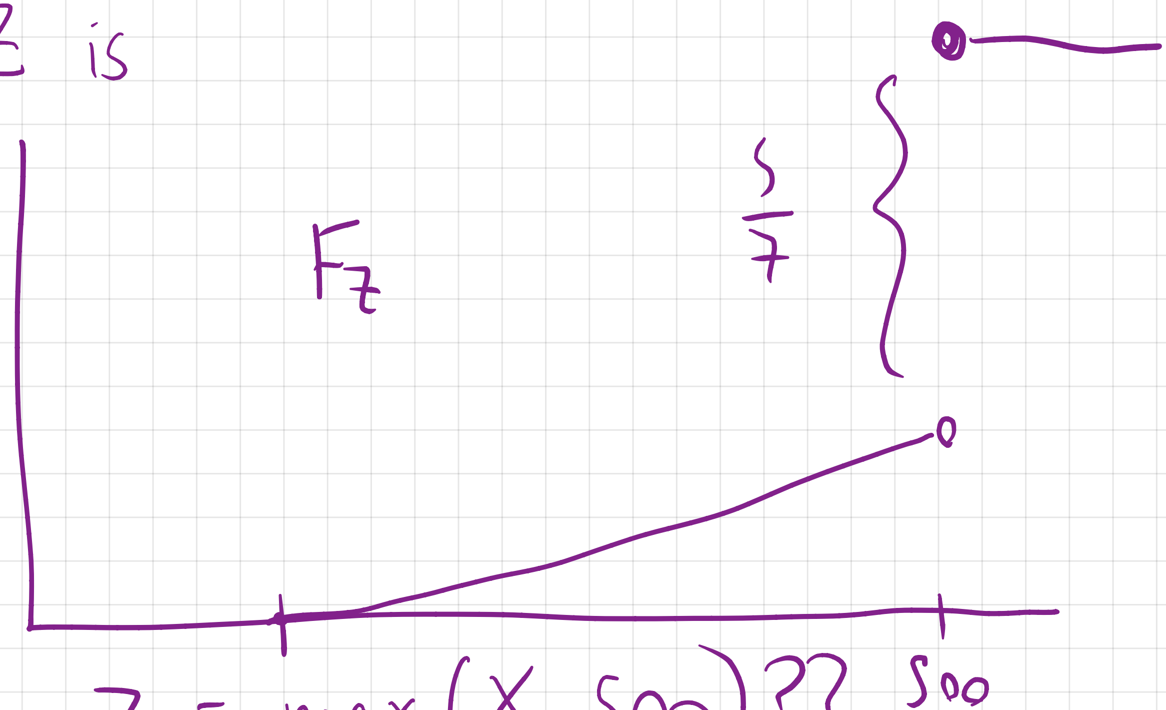
The random variable Z is

(a) continuous

(b) discrete

(c) neither

~~(d) both~~



$X \sim \text{Unif}([100, 1500])$ $\quad Z = \cancel{\max}\ \min(X, 500)$??

correction from lecture

$$f_X(t) = \begin{cases} \frac{1}{1500-100} & 100 \le t \le 1500 \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbb{P}(Z=500) = \mathbb{P}(Z \ge 500)$$
$$= \mathbb{P}(X \ge 500)$$
$$= \int_{500}^{1500} \frac{1}{1400}\,dt = \frac{5}{7} > 0$$

But if $r < 500$ $\quad \mathbb{P}(Z=r) = 0.$