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, $F_X(r) = \mathbb{P}(X \leq r)$. $\quad r \in \mathbb{R}$

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

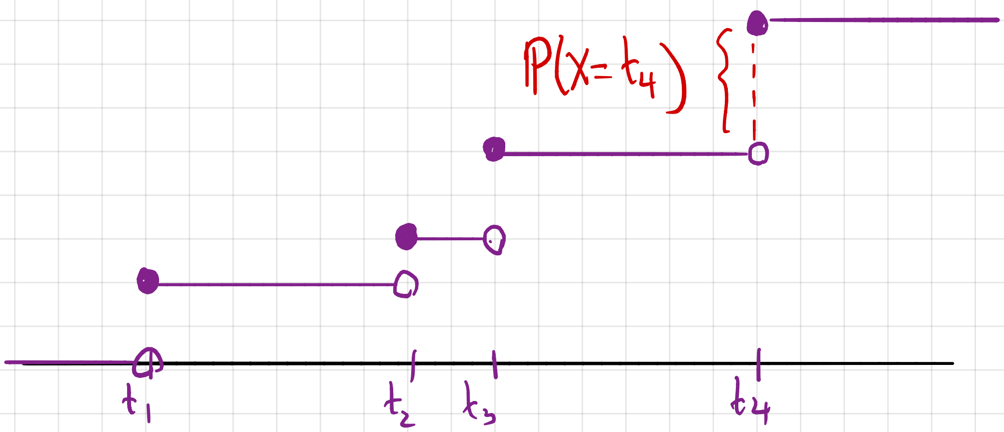
(2) $\lim_{r \to -\infty} F_X(r) = 0$, $\lim_{r \to +\infty} F_X(r) = 1$.

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

Discrete random variable:

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

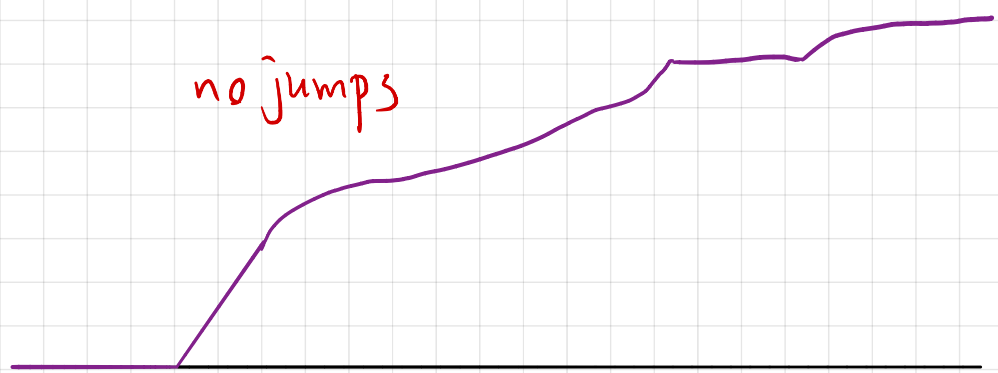
$$\sum_j \mathbb{P}(X = t_j) = 1.$$



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.



Densities

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

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

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

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

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

$$p_X(t) \geq 0, \quad \sum_{t} p_X(t) = 1.$$

X continuous

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

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

Y = distance from center.

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

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

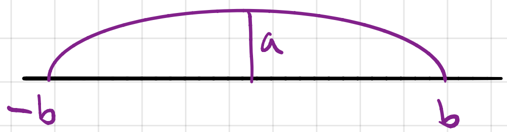
Proof:

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

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

Eg. Let $f(t) = c\sqrt{b^2 - t^2}$ for $|t| \leq b$, 0 otherwise

(for some positive constants $b, c > 0$).



Is f a probability density?

Eg. 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

Independent Random Variables

2.3

A collection $X_1, X_2, X_3, \ldots, X_n$ of random variables defined on the same sample space are <u>independent</u> if

for any $B_1, B_2, \ldots, B_n \subseteq \mathbb{R}$, the events

$$\{X_1 \in B_1\}, \{X_2 \in B_2\}, \ldots, \{X_n \in B_n\} \text{ are independent.}$$

Special Case: if the X_j are discrete random variables, it suffices to check the simpler condition

for any real numbers $t_1, t_2, \ldots, t_n$

Eg. Let $X_1, X_2, \ldots, X_n$ be fair coin tosses. Denote $H \sim 1$, $T \sim 0$.

Independent Trials

2.4

Experiments can have numerical observables, but sometimes you only observe whether there is success or failure

We model this with a random variable X taking value 1 with some probability p, and value 0 with probability $1-p$.

$$X \sim \mathrm{Ber}(p) \quad (\text{Bernoulli})$$

In practice, we usually repeat the experiment many times, making sure to use the same setup each trial. The previous trials do not influence the future ones.

How many successful trials?

Run n independent trials, each with success probability p.

$$X_1, X_2, \dots, X_n \sim \mathrm{Ber}(p).$$

Let S_n = # successful trials

What is the distribution of S_n?

Eg. Roll a fair die 10 times. What is the probability that 6 comes up at least 3 times?

Eg. What is the probability that no 6 is rolled in the 10 rolls?

Now, keep rolling. Let N denote the first roll where a 6 appears. N is a random variable. What is its distribution?

First Success Time

N = first success in repeated independent trials (success rate p).

Model trials with (unlimited number) of independent $\mathrm{Ber}(p)$'s:

$X_1, X_2, X_3, X_4, \ldots$

$$\{N=k\} = \{X_1=0, X_2=0, X_3=0, \ldots, X_{k-1}=0, X_k=1\}$$

Geometric Distribution $\mathrm{Geom}(p)$ on $\{0,1,2,3,\ldots\} = \mathbb{N}$.