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

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

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

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 $P(X = t_j) > 0$ and $\sum_j P(X = t_j) = 1$.

**Continuous random variable**:

- For each real number $t$, $P(X = t) = 0$.

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

- No jumps
Densities

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

\[
X \text{ discrete, } \epsilon \{ t_1, t_2, t_3, \ldots \} \quad \text{X continuous} \\
\Pr(X = t) \quad \text{probability mass function} \quad \Pr(X = t) = 0 \text{ for all } t \in \mathbb{R}.
\]

\[
\Pr(X \in A) = \sum_{t \in A} \Pr(X = t) \\
= \sum_{t \in A} p_x(t) \\
p_x(t) \geq 0, \quad \sum_{t} p_x(t) = 1.
\]
Eg. Shoot an arrow at a circular target of radius $1$. 

$Y = \text{distance from center.}$

$$\int_{-\infty}^{r} f(t) \, dt \quad \overset{?}{=} \quad \mathbb{P}(Y \in (-\infty, r]) = F_Y(r) = \begin{cases} 0, & r \leq 0 \\ r^2, & 0 \leq r \leq 1 \\ 1, & r \geq 1 \end{cases}$$
**Theorem:** If $F_x$ is continuous and piecewise differentiable, then $X$ has a density $f_x = F'_x$.

**Proof:**

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

$$P(X \in [s,t]) = t - s \text{ if } 0 \leq s < t \leq 1.$$
Eg. Let \( f(t) = c \sqrt{b^2 - t^2} \) for \( |t| \leq b \), 0 otherwise.

(for some positive constants \( b > 0 \)).

Is \( f \) a probability density?
E.g. Your car is in a minor accident; the damage repair cost is a random number between $100 and $1500. Your insurance 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

A collection $X_1, X_2, X_3, \ldots, X_n$ of random variables defined on the same sample space are independent if for any $B_1, B_2, \ldots, B_n \subseteq \mathbb{R}$, the events

$$\{X_i \in B_i\}, \{X_2 \in B_2\}, \ldots, \{X_n \in B_n\}$$

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

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 $1-p$.

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

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 $p$.

$X_1, X_2, \ldots, X_n \sim \text{Ber}(p)$.

Let $S_n$ = number of 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

\[ N = \text{first success in repeated independent trials (success rate } p). \]

Model trials with (unlimited number) of independent \( \text{Ber}(p) \)'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 \( \text{Geom}(p) \) on \( \{0, 1, 2, 3, \ldots\} = N \).