Math 180A: Intro to Probability
(for Data Science)

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

Today: § 2.4-2.5, 4.4
Next: § 3.3

Lab 2.2 due TONIGHT (maybe?)
HW3 due Friday.
Midterm next Wed evening

* Regrade requests for Lab1: window
  Thursday 10/17
  8am - 11pm

  → separate request for each problem
  → detailed, polite responses please.

* The material slotted for today’s lecture is the cutoff
  for Midterm 1.
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

$\{X_1 \in B_1\}, \{X_2 \in B_2\}, \ldots, \{X_n \in B_n\}$ 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 $H \sim 1$, $T \sim 0$. 
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 $S_n = \# \text{successful trials}$

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 \( \{0, 1, 2, 3, \ldots\} = N \).
Rare Events

If $S_n = S_{n,p} \sim \text{Bin}(n,p)$, $S_n$ is the number of successes in $n$ independent trials each with success probability $p$.

What if $p$ is very small, but $n$ is very large?

One way to handle this mathematically is a

$L \rightarrow$ For each $n$, take $p \propto \frac{1}{n}$. $p = \frac{\lambda}{n}$ for some $\lambda > 0$.

\[ P(S_{n,p} = k) = \]

What happens as $n \to \infty$?
\[ P(S_{np} = k) = \binom{n}{k} \left( \frac{\lambda}{n} \right)^k \left( 1 - \frac{\lambda}{n} \right)^{n-k} \]
Poisson Distribution

A random variable $X$ has the Poisson($\lambda$) distribution if

$$P(X=k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k=0,1,2,\ldots$$

Eg. A 100 year storm is a storm magnitude expected to occur in any given year with probability $1/100$. Over the course of a century, how likely is it to see at least 4 100 year storms?
Summary

Sampling independent trials, the most important (discrete) probability distributions are:

- **Ber(p)**: \( P(X = 1) = p, \ P(X = 0) = 1 - p \quad 0 \leq p \leq 1 \) (single trial with success probability \( p \))

- **Bin(n,p)**: \( P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad 0 \leq k \leq n \) (number of successes in \( n \) independent trials with rate \( p \))

- **Geom(p)**: \( P(N = k) = (1-p)^{k-1} p \quad k = 0, 1, 2, ... \) (first successful trial in repeated independent trials with rate \( p \))

- **Poisson(\( \lambda \))**: \( P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad k = 0, 1, 2, ... \quad \lambda > 0 \) (Approximates \( \text{Bin}(n, \frac{\lambda}{n}) \); number of rare events in many trials)