CSE 312
Foundations of Computing II
Lecture 28: Victory Lap, What’s Next, & Review
Announcements / Logistics

• Do the class evaluation!

• Check the exam instructions post
  – https://edstem.org/us/courses/29595/discussion/2225663
  – Q&A session on Sunday, 2:00pm, over zoom!

• Longer office hours today for me (2:30pm – 3:20pm)
What you’ve learned …

The essentials of probability and some statistics,

hands-on applications,

• Naïve Bayes SPAM filtering
• Bloom Filters
• MinHash for Distinct Elements
• Markov Chains and PageRank

some societal connections

• Differential privacy
• Detecting statistical lies

and some Python…

a great headstart for CSE 446 (ML)
What’s next?

• Some places to apply and extend your knowledge
  – CSE 421 Algorithms - counting and more basic probability
  – CSE 422 Toolkit for Modern Algorithms – probability everywhere
  – CSE 426 Cryptography – randomness, reasoning about probability essential
  – CSE 427 Computational Biology
  – CSE 446 Machine Learning – this course + linear algebra essential
  – CSE 447 Natural Language Processing
  – CSE 473 Artificial Intelligence – Bayes nets, probability, etc.
  – CSE 490Q Quantum Computing – the quantum world is inherently random
Agenda

• What you’ve learned
• What’s next
• Review
Counting: Sum & Product Rules

• **Sum rule:**
  If you can choose from
  – EITHER one of \( n \) options,
  – OR one of \( m \) options with NO overlap with the previous \( n \),
  then the number of possible outcomes of the experiment is \( n + m \)

• **Product rule:**
  In a sequential process, if there are
  – \( n_1 \) choices for the 1\(^{st} \) step,
  – \( n_2 \) choices for the 2\(^{nd} \) step (given the first choice), ..., and
  – \( n_k \) choices for the \( k^{th} \) step (given the previous choices),
  then the total number of outcomes is \( n_1 \times n_2 \times n_3 \times \cdots \times n_k \)
**Counting: Permutations & Combinations**

**Permutations.** The number of orderings of \( n \) distinct objects

\[
n! = n \times (n-1) \times \cdots \times 2 \times 1
\]

*Example: How many sequences in \( \{1,2,3\}^3 \) with no repeating elements?*

**k-Permutations.** The number of orderings of **only** \( k \) out of \( n \) distinct objects

\[
P(n,k) = n \times (n-1) \times \cdots \times (n-k+1) = \frac{n!}{(n-k)!}
\]

*Example: How many sequences of 5 distinct alphabet letters from \( \{A,B,\ldots,Z\} \)?

**Combinations / Binomial Coefficient.** The number of ways to select \( k \) out of \( n \) objects, where ordering of the selected \( k \) does not matter:

\[
\binom{n}{k} = \frac{P(n,k)}{k!} = \frac{n!}{k! (n-k)!}
\]

*Example: How many size-5 subsets of \( \{A,B,\ldots,Z\} \)?

*Example: How many shortest paths from Gates to Starbucks?*

*Example: How many solutions \((x_1, \ldots, x_k)\) such that \(x_1, \ldots, x_k \geq 0 \) and \( \sum_{i=1}^{k} x_i = n\)?*
Counting: When order only partly matters

We often want to count # of partly ordered lists:

Let $M = \#$ of ways to produce fully ordered lists

$P = \#$ of partly ordered lists

$N = \#$ of ways to produce corresponding fully ordered list given a partly ordered list

Then $M = P \cdot N$ by the product rule. Often $M$ and $N$ are easy to compute:

$$P = \frac{M}{N}$$

Dividing by $N$ “removes” part of the order.
Multinomial Coefficients

If we have $k$ types of objects ($n$ total), with $n_1$ of the first type, $n_2$ of the second, ..., and $n_k$ of the $k^{th}$, then the number of orderings possible is

$$
\binom{n}{n_1, n_2, \ldots, n_k} = \frac{n!}{n_1! n_2! \cdots n_k!}
$$
Counting using binary encoding/star and bars

The number of ways to distribute \( n \) indistinguishable balls into \( k \) distinguishable bins is

\[
{n + k - 1 \choose k - 1} = {n + k - 1 \choose n}
\]

E.g., \# of ways to add \( k \) non-negative integers up to \( n \)

Encode using one symbol (1 or *) for items, the other (0 or |) for dividers
Counting: Binomial Theorem

**Theorem.** Let $x, y \in \mathbb{R}$ and $n \in \mathbb{N}$ a positive integer. Then,

$$(x + y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$$
Counting: Inclusion-Exclusion

Let $A, B$ be sets. Then

$$|A \cup B| = |A| + |B| - |A \cap B|$$

In general, if $A_1, A_2, ..., A_n$ are sets, then

$$|A_1 \cup A_2 \cup \cdots \cup A_n| = \text{singles} - \text{doubles} + \text{triples} - \text{quads} + \cdots$$

$$= (|A_1| + \cdots + |A_n|) - (|A_1 \cap A_2| + \cdots + |A_{n-1} \cap A_n|) + \cdots$$
Counting: Pigeonhole Principle

If there are \( n \) pigeons in \( k < n \) holes, then one hole must contain at least \( \left\lceil \frac{n}{k} \right\rceil \) pigeons!

**Reason.** Can’t have fractional number of pigeons

Syntax reminder:

- Ceiling: \( [x] \) is \( x \) rounded up to the nearest integer (e.g., \( [2.731] = 3 \))
- Floor: \( [x] \) is \( x \) rounded down to the nearest integer (e.g., \( [2.731] = 2 \))
**Probability**

**Definition.** A **sample space** \( \Omega \) is the set of all possible outcomes of an experiment.

**Examples:**
- Single coin flip: \( \Omega = \{H,T\} \)
- Two coin flips: \( \Omega = \{HH,HT,TH,TT\} \)
- Roll of a die: \( \Omega = \{1,2,3,4,5,6\} \)

**Definition.** An **event** \( E \subseteq \Omega \) is a subset of possible outcomes.

**Examples:**
- Getting at least one head in two coin flips: \( E = \{HH,HT,TH\} \)
- Rolling an even number on a die: \( E = \{2,4,6\} \)
**Definition.** A (discrete) probability space is a pair $(\Omega, P)$ where:

- $\Omega$ is a set called the **sample space**.
- $P$ is the **probability measure**, a function $P: \Omega \to \mathbb{R}$ such that:
  - $P(x) \geq 0$ for all $x \in \Omega$
  - $\sum_{x \in \Omega} P(x) = 1$

For $A \subseteq \Omega$:

$$P(A) = \sum_{x \in A} P(x)$$
**Definition.** A **random variable (RV)** for a probability space \((\Omega, P)\) is a function \(X: \Omega \rightarrow \mathbb{R}\).

The set of values that \(X\) can take on is its **range/support**: \(X(\Omega)\) or \(\Omega_X\)

\[
\{X = x_i\} = \{\omega \in \Omega \mid X(\omega) = x_i\}
\]

Random variables **partition** the sample space.

\[
\sum_{x \in X(\Omega)} P(X = x) = 1
\]
Probability Mass Function (PMF) and CDF (Discrete Case)

Definitions:

For a RV $X: \Omega \to \mathbb{R}$, the probability mass function (pmf) of $X$ specifies, for any real number $x$, the probability that $X = x$

$$p_X(x) = P(X = x) = P(\{\omega \in \Omega \mid X(\omega) = x\})$$

$$\sum_{x \in \Omega_X} p_X(x) = 1$$

For a RV $X: \Omega \to \mathbb{R}$, the cumulative distribution function (cdf) of $X$ specifies, for any real number $x$, the probability that $X \leq x$

$$F_X(x) = P(X \leq x)$$
**Probability Density Function**

- **Non-negativity:** $f_X(x) \geq 0$ for all $x \in \mathbb{R}$

- **Normalization:** $\int_{-\infty}^{+\infty} f_X(x) \, dx = 1$

$$P(a \leq X \leq b) = \int_{a}^{b} f_X(x) \, dx$$

$$\frac{P(X \approx y)}{P(X \approx z)} \approx \frac{\epsilon f_X(y)}{\epsilon f_X(z)} = \frac{f_X(y)}{f_X(z)}$$

What $f_X(x)$ measures: The local rate at which probability accumulates
Definition. The cumulative distribution function (cdf) of \( X \) is

\[
F_X(a) = P(X \leq a) = \int_{-\infty}^{a} f_X(x) \, dx
\]

By the fundamental theorem of Calculus \( f_X(x) = \frac{d}{dx} F_X(x) \)

Therefore: \( P(X \in [a, b]) = F_X(b) - F_X(a) \)

\( F_X \) is monotone increasing, since \( f_X(x) \geq 0 \). That is \( F_X(c) \leq F_X(d) \) for \( c \leq d \)
Continuous Random Variables

Probability Density Function (PDF).
\[ f: \mathbb{R} \to \mathbb{R} \quad \text{s.t.} \]
- \[ f(x) \geq 0 \text{ for all } x \in \mathbb{R} \]
- \[ \int_{-\infty}^{+\infty} f(x) \, dx = 1 \]

Cumulative Distribution Function (CDF).
\[ F(y) = \int_{-\infty}^{y} f(x) \, dx \]

Theorem. \[ f(x) = \frac{dF(x)}{dx} \]

Density ≠ Probability!

\[ P(X \in [a, b]) = \int_{a}^{b} f_X(x) \, dx = F_X(b) - F_X(a) \]

\[ F_X(y) = P(X \leq y) \]
Probability: Inclusion-Exclusion

Let $A, B$ be events. Then

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

In general, if $A_1, A_2, \ldots, A_n$ are events, then

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = \text{singles} - \text{doubles} + \text{triples} - \text{quads} + \ldots$$

$$= (P(A_1) + \ldots + P(A_n))$$
$$- (P(A_1 \cap A_2) + \ldots + P(A_{n-2} \cap A_n) + P(A_{n-1} \cap A_n))$$
$$+ \ldots$$
Conditional Probability

• **Conditional Probability**
  
  \[
  P(B|A) = \frac{P(A \cap B)}{P(A)} \quad \text{for } P(A) \neq 0
  \]

• **Bayes Theorem**
  
  \[
  P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad \text{if } P(A) \neq 0, P(B) \neq 0
  \]

• **Law of Total Probability**
  
  \[E_1, \ldots, E_n \text{ partition } \Omega\]

  \[
  P(F) = \sum_{i=1}^{n} P(F \cap E_i) = \sum_{i=1}^{n} P(F|E_i) P(E_i)
  \]
Bayes Theorem with Law of Total Probability

Bayes Theorem with LTP: Let $E_1, E_2, ..., E_n$ be a partition of the sample space, and $F$ and event. Then,

$$P(E_1|F) = \frac{P(F|E_1)P(E_1)}{P(F)} = \frac{\sum_{i=1}^{n} P(F|E_i)P(E_i)}{P(F)}$$

Simple Partition: In particular, if $E$ is an event with non-zero probability, then

$$P(E|F) = \frac{P(F|E)P(E)}{P(F|E)P(E) + P(F|E^C)P(E^C)}$$
Chain rule & Independence

**Theorem. (Chain Rule)** For events $A_1, A_2, \ldots, A_n$,

$$P(A_1 \cap \cdots \cap A_n) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap A_2 \cap \cdots \cap A_{n-1})$$

**Definition.** Two events $A$ and $A$ are (statistically) **independent** if

$$P(A \cap B) = P(A) \cdot P(B).$$

“Equivalently,” $P(A|B) = P(A)$.

**Definition.** Two events $A$ and $B$ are **independent conditioned on** $C$ if

$P(C) \neq 0$ and $P(A \cap B | C) = P(A | C) \cdot P(B | C)$. 
Multiple Events – Mutual Independence

**Definition.** Events $A_1, \ldots, A_n$ are **mutually independent** if for every non-empty subset $I \subseteq \{1, \ldots, n\}$, we have

$$P\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} P(A_i).$$
**Expected Value of a Random Variable (Discrete Case)**

**Definition.** Given a discrete RV $X: \Omega \rightarrow \mathbb{R}$, the **expectation** or **expected value** or **mean** of $X$ is

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega) \cdot P(\omega)$$

or equivalently

$$\mathbb{E}[X] = \sum_{x \in X(\Omega)} x \cdot P(X = x) = \sum_{x \in \Omega_X} x \cdot p_X(x)$$

Intuition: “Weighted average” of the possible outcomes (weighted by probability)
Theorem. For any random variables $X_1, \ldots, X_n$, and real numbers $a_1, \ldots, a_n \in \mathbb{R}$,

$$
\mathbb{E}[a_1X_1 + \cdots + a_nX_n] = a_1\mathbb{E}[X_1] + \cdots + a_n\mathbb{E}[X_n].
$$

Very important: In general, we do not have $\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$. 
Linearity of Expectation with Indicator Variables.

We flip $n$ coins, each one heads with probability $p$. $Z$ is the number of heads, what is $\mathbb{E}[Z]$?

- $X_i = \begin{cases} 1, & i\text{th coin flip is heads} \\ 0, & i\text{th coin flip is tails.} \end{cases}$

Fact. $Z = X_1 + \cdots + X_n$

Linearity of Expectation:

$$\mathbb{E}[Z] = \mathbb{E}[X_1 + \cdots + X_n] = \mathbb{E}[X_1] + \cdots + \mathbb{E}[X_n] = n \cdot p$$

$$P(X_i = 1) = p$$
$$P(X_i = 0) = 1 - p$$

$$\mathbb{E}[X_i] = p \cdot 1 + (1 - p) \cdot 0 = p$$
No independence required for Linearity of Expectation

Each coin shows up heads half the time.

Two fair coins

\[
\begin{align*}
P(HT) &= P(TH) = 0.25 \\
P(HH) &= P(TT) = 0.25 \\
\mathbb{E}(X) &= 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} = 1
\end{align*}
\]

Glued coins

\[
\begin{align*}
P(HT) &= P(TH) = 0.5 \\
P(HH) &= P(TT) = 0 \\
\mathbb{E}(X) &= 1 \cdot 1 = 1
\end{align*}
\]

Attached coins

\[
\begin{align*}
P(HH) &= P(TT) = 0.4 \\
P(HT) &= P(TH) = 0.1 \\
\mathbb{E}(X) &= 1 \cdot 0.2 + 2 \cdot 0.4 = 1
\end{align*}
\]
LOTUS: Expected Value of $g(X)$ (Discrete Case)

**Definition.** Given a discrete RV $X: \Omega \rightarrow \mathbb{R}$, the expectation or expected value or mean of $g(X)$ is

$$
\mathbb{E}[g(X)] = \sum_{\omega \in \Omega} g(X(\omega)) \cdot P(\omega)
$$

or equivalently

$$
\mathbb{E}[g(X)] = \sum_{x \in X(\Omega)} g(x) \cdot P(X = x) = \sum_{x \in \Omega_x} g(x) \cdot p_X(x)
$$

Also known as **LOTUS**: “Law of the unconscious statistician”
Linearity is special!

In general $\mathbb{E}[g(X)] \neq g(\mathbb{E}[X])$

E.g., $X = \begin{cases} +1 \text{ with prob } 1/2 \\ -1 \text{ with prob } 1/2 \end{cases}$

Then: $\mathbb{E}[X^2] \neq \mathbb{E}[X]^2$
Variance (Discrete Case)

**Definition.** The variance of a (discrete) RV $X$ is

$$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \sum_x p_x(x) \cdot (x - \mathbb{E}[X])^2$$

**Theorem.** For any $a, b \in \mathbb{R}$, $\text{Var}(a \cdot X + b) = a^2 \cdot \text{Var}(X)$

**Theorem.** $\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$

**Definition.** The standard deviation of a (discrete) RV $X$ is $\sigma_X = \sqrt{\text{Var}(X)}$

**Note.** For any $a \geq 0, b \in \mathbb{R}$, $\sigma_{a \cdot X + b} = a \cdot \sigma_X$
**Definition.** The expected value of a continuous RV $X$ is defined as

$$E[X] = \int_{-\infty}^{+\infty} f_X(x) \cdot x \, dx$$

**Fact.** $E[aX + bY + c] = aE[X] + bE[Y] + c$

**Definition.** The variance of a continuous RV $X$ is defined as

$$\text{Var}(X) = \int_{-\infty}^{+\infty} f_X(x) \cdot (x - E[X])^2 \, dx = E[X^2] - E[X]^2$$

$$\text{Var}(aX + b) = a^2 \text{Var}(X)$$
Definition. Given a continuous RV $X: \mathbb{R} \to \mathbb{R}$, the expectation or expected value or mean of $g(X)$ is

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) \, dx$$
## Review: From Discrete to Continuous

<table>
<thead>
<tr>
<th></th>
<th>Discrete</th>
<th>Continuous</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PMF/PDF</strong></td>
<td>( p_X(x) = P(X = x) )</td>
<td>( f_X(x) \neq P(X = x) = 0 )</td>
</tr>
<tr>
<td><strong>CDF</strong></td>
<td>( F_X(x) = \sum_{t \leq x} p_X(t) )</td>
<td>( F_X(x) = \int_{-\infty}^{x} f_X(t) , dt )</td>
</tr>
<tr>
<td><strong>Normalization</strong></td>
<td>( \sum_x p_X(x) = 1 )</td>
<td>( \int_{-\infty}^{\infty} f_X(x) , dx = 1 )</td>
</tr>
<tr>
<td><strong>Expectation</strong></td>
<td>( \mathbb{E}[g(X)] = \sum_x g(x) , p_X(x) )</td>
<td>( \mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) , f_X(x) , dx )</td>
</tr>
</tbody>
</table>
Properties of Independent Random Variables

**Theorem.** If $X, Y$ independent, $\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$

**Theorem.** If $X, Y$ independent, $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$

**Corollary.** If $X_1, X_2, \ldots, X_n$ mutually independent,

$$\text{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i} \text{Var}(X_i)$$
Joint PMFs and Joint Range

**Definition.** Let $X$ and $Y$ be discrete random variables. The **Joint PMF** of $X$ and $Y$ is

$$p_{X,Y}(a, b) = P(X = a, Y = b)$$

**Definition.** The **joint range** of $p_{X,Y}$ is

$$\Omega_{X,Y} = \{(c, d) : p_{X,Y}(c, d) > 0\} \subseteq \Omega_X \times \Omega_Y$$

Note that

$$\sum_{(s, t) \in \Omega_{X,Y}} p_{X,Y}(s, t) = 1$$
Marginal PMF

**Definition.** Let $X$ and $Y$ be discrete random variables and $p_{X,Y}(a, b)$ their joint PMF. The **marginal PMF** of $X$

$$p_X(a) = \sum_{b \in \Omega_Y} p_{X,Y}(a, b)$$

Similarly, $p_Y(b) = \sum_{a \in \Omega_X} p_{X,Y}(a, b)$
Continuous distributions on $\mathbb{R} \times \mathbb{R}$

**Definition.** The joint probability density function (PDF) of continuous random variables $X$ and $Y$ is a function $f_{X,Y}$ defined on $\mathbb{R} \times \mathbb{R}$ such that

- $f_{X,Y}(x, y) \geq 0$ for all $x, y \in \mathbb{R}$
- $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y) \, dx \, dy = 1$

for $A \subseteq \mathbb{R} \times \mathbb{R}$ the probability that $(X, Y) \in A$ is $\iint_{A} f_{X,Y}(x, y) \, dx \, dy$

The (marginal) PDFs $f_X$ and $f_Y$ are given by

- $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) \, dy$
- $f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) \, dx$
Independence and joint distributions

Discrete random variables $X$ and $Y$ are independent iff

- $p_{X,Y}(x, y) = p_X(x) \cdot p_Y(y)$ for all $x \in \Omega_X, y \in \Omega_Y$

Continuous random variables $X$ and $Y$ are independent iff

- $f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y)$ for all $x, y \in \mathbb{R}$
Conditional Expectation

**Definition.** Let $X$ be a discrete random variable then the conditional expectation of $X$ given event $A$ is

$$
\mathbb{E}[X \mid A] = \sum_{x \in \Omega_x} x \cdot P(X = x \mid A)
$$

**Notes:**

- Can be phrased as a “random variable version”

  $$
  \mathbb{E}[X \mid Y = y]
  $$

- Linearity of expectation still applies here

  $$
  \mathbb{E}[aX + bY + c \mid A] = a \mathbb{E}[X \mid A] + b \mathbb{E}[Y \mid A] + c
  $$
Law of Total Expectation

Law of Total Expectation (event version). Let $X$ be a random variable and let events $A_1, \ldots, A_n$ partition the sample space. Then,

$$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X | A_i] \cdot P(A_i)$$

Law of Total Expectation (random variable version). Let $X$ be a random variable and $Y$ be a discrete random variable. Then,

$$\mathbb{E}[X] = \sum_{y \in \Omega_Y} \mathbb{E}[X | Y = y] \cdot P(Y = y)$$
<table>
<thead>
<tr>
<th></th>
<th>Discrete</th>
<th>Continuous</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Joint PMF/PDF</strong></td>
<td>$p_{x,y}(x, y) = P(X = x, Y = y)$</td>
<td>$f_{x,y}(x, y) \neq P(X = x, Y = y)$</td>
</tr>
<tr>
<td><strong>Joint CDF</strong></td>
<td>$F_{x,y}(x, y) = \sum_{t \leq x} \sum_{s \leq y} p_{x,y}(t, s)$</td>
<td>$F_{x,y}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{x,y}(t, s),ds,dt$</td>
</tr>
<tr>
<td><strong>Normalization</strong></td>
<td>$\sum_{x} \sum_{y} p_{x,y}(x, y) = 1$</td>
<td>$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{x,y}(x, y),dx,dy = 1$</td>
</tr>
<tr>
<td><strong>Marginal PMF/PDF</strong></td>
<td>$p_{x}(x) = \sum_{y} p_{x,y}(x, y)$</td>
<td>$f_{x}(x) = \int_{-\infty}^{\infty} f_{x,y}(x, y),dy$</td>
</tr>
<tr>
<td><strong>Expectation</strong></td>
<td>$E[g(X, Y)] = \sum_{x} \sum_{y} g(x, y)p_{x,y}(x, y)$</td>
<td>$E[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y)f_{x,y}(x, y),dx,dy$</td>
</tr>
<tr>
<td><strong>Independence</strong></td>
<td>$\forall x, y, p_{x,y}(x, y) = p_{x}(x)p_{y}(y)$</td>
<td>$\forall x, y, f_{x,y}(x, y) = f_{x}(x)f_{y}(y)$</td>
</tr>
</tbody>
</table>
Markov’s and Chebyshev’s Inequalities

Theorem (Markov’s Inequality). Let $X$ be a random variable taking only non-negative values. Then, for any $t > 0$,

$$P(X \geq t) \leq \frac{\mathbb{E}[X]}{t}.$$ 

Theorem (Chebyshev’s Inequality). Let $X$ be a random variable. Then, for any $t > 0$,

$$P(|X - \mathbb{E}[X]| \geq t) \leq \frac{\text{Var}(X)}{t^2}.$$
Theorem. Let $X = X_1 + \cdots + X_n$ be a sum of independent RVs, each taking values in $[0,1]$, such that $\mathbb{E}[X] = \mu$. Then, for every $\delta > 0$,

$$P(|X - \mu| \geq \delta \cdot \mu) \leq e^{-\frac{\delta^2 \mu}{4}}.$$  

Herman Chernoff, Herman Rubin, Wassily Hoeffding

Example: If $X \sim \text{Bin}(n, p)$, then $X = X_1 + \cdots + X_n$ is a sum of independent $\{0,1\}$-Bernoulli variables, and $\mu = np$

Note: More accurate versions are possible, but with more cumbersome right-hand side (e.g., see textbook)
Theorem (Union Bound). Let $A_1, \ldots, A_n$ be arbitrary events. Then,

$$P\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} P(A_i)$$

Intuition (3 evts.):
Bernoulli Random Variables

A random variable $X$ that takes value 1 ("Success") with probability $p$, and 0 ("Failure") otherwise. $X$ is called a Bernoulli random variable.

Notation: $X \sim \text{Ber}(p)$

PMF: $P(X = 1) = p$, $P(X = 0) = 1 - p$

Expectation: $\mathbb{E}[X] = p$  
Note: $\mathbb{E}[X^2] = p$

Variance: $\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = p - p^2 = p(1 - p)$

Examples:
- Coin flip
- Randomly guessing on a MC test question
- A server in a cluster fails
- Any indicator RV
Binomial Random Variables

A discrete random variable $X$ that is the number of successes in $n$ independent random variables $Y_i \sim \text{Ber}(p)$. $X$ is a **Binomial random variable** where $X = \sum_{i=1}^{n} Y_i$

**Notation:** $X \sim \text{Bin}(n, p)$

**PMF:** $P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$

**Expectation:** $\mathbb{E}[X] = np$

**Variance:** $\text{Var}(X) = np(1 - p)$
Geometric Random Variables

A discrete random variable $X$ that models the number of independent trials $Y_i \sim \text{Ber}(p)$ before seeing the first success.

$X$ is called a geometric random variable with parameter $p$.

**Notation:** $X \sim \text{Geo}(p)$

**PMF:** $P(X = k) = (1 - p)^{k-1}p$

**Expectation:** $E[X] = \frac{1}{p}$

**Variance:** $\text{Var}(X) = \frac{1-p}{p^2}$

**Examples:**
- # of coin flips until first head
- # of random guesses on MC questions until you get one right
- # of random guesses at a password until you hit it
Uniform Distribution (Discrete)

A discrete random variable $X$ equally likely to take any (integer) value between integers $a$ and $b$ (inclusive), is uniform.

**Notation:** $X \sim \text{Unif}[a, b]$

**PMF:** $P(X = i) = \frac{1}{b-a+1}$

**Expectation:** $\mathbb{E}[X] = \frac{a+b}{2}$

**Variance:** $\text{Var}(X) = \frac{(b-a)(b-a+2)}{12}$

**Example:** value shown on one roll of a fair die is $\text{Unif}[1,6]$:
- $P(X = i) = 1/6$
- $\mathbb{E}[X] = 7/2$
- $\text{Var}(X) = 35/12$
Uniform Distribution (Continuous)

\[ X \sim \text{Unif}(a, b) \]

pdf:

\[ f_X(x) = \begin{cases} 
\frac{1}{b - a} & \text{if } x \in [a, b] \\
0 & \text{else}
\end{cases} \]

cdf:

\[ F_X(y) = \begin{cases} 
0 & \text{if } x < a \\
\frac{x - a}{b - a} & \text{if } x \in [a, b] \\
1 & \text{if } x > b
\end{cases} \]

Mean:

\[ \mathbb{E}[X] = \frac{a + b}{2} \]

Variance:

\[ \text{Var}(X) = \frac{(b - a)^2}{12} \]
Poisson Distribution

- $X$ is a **Poisson r.v. with parameter** $\lambda$ (denoted $X \sim \text{Poi}(\lambda)$) with this distribution (PMF): For all non-negative integers $k = 0, 1, 2, ...$

  $$P(Z = k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!}$$

- $\mathbb{E}[X] = \lambda$ and $\text{Var}(X) = \lambda$

Distribution of the # of events that happen, independently, at an **average** rate of $\lambda$ per unit time: car arrivals, customers, radioactive decay

**Theorem.** Let $X_1 \sim \text{Poi}(\lambda_1), \ldots, X_n \sim \text{Poi}(\lambda_n)$ be independent. Set $Z = \sum_i X_i$. Then $Z \sim \text{Poi}(\lambda)$ for $\lambda = \sum_i \lambda_i$.
Exponential Distribution

An **exponential random variable** $X$ with parameter $\lambda \geq 0$

$(X \sim \text{Exp}(\lambda))$ follows the exponential density $f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$

**CDF:** For $y \geq 0$, $F_X(y) = 1 - e^{-\lambda y}$

$\mathbb{E}[X] = \frac{1}{\lambda}$ \quad $\text{Var}(X) = \frac{1}{\lambda^2}$

**Theorem.** $X \sim \text{Exp}(\lambda)$ is **memoryless:** i.e. for all $s, t > 0$,

$P(X > s + t \mid X > s) = P(X > t)$. 

$P(X > t) = e^{-t\lambda}$
The Normal Distribution

A Gaussian (or normal) random variable $X \sim \mathcal{N}(\mu, \sigma^2)$ with parameters $\mu \in \mathbb{R}$ and $\sigma \geq 0$ has density

$$f_X(x) = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

**Fact.** If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $E[X] = \mu$, and $\text{Var}(X) = \sigma^2$

**Fact.** If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $Y = aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$

**Cor: $\frac{X-\mu}{\sigma} \sim \mathcal{N}(0, 1)$**

**Fact: Sum of independent normals is normal**
Independent and Identically Distributed (i.i.d.) RVs

Let $X_1, \ldots, X_n$ random variables, each chosen independently with the same (identical) distribution having expectation $\mu$ and variance $\sigma^2$

$$E[X_1 + \cdots + X_n] = E[X_1] + \cdots + E[X_n] = n\mu$$

$$\text{Var}(X_1 + \cdots + X_n) = \text{Var}(X_1) + \cdots + \text{Var}(X_n) = n\sigma^2$$

**Empirical observation:** $X_1 + \cdots + X_n$ looks like a normal RV as $n$ grows.
Central Limit Theorem

Theorem. (Central Limit Theorem) The CDF of $Y_n$ converges to the CDF of the standard normal $\mathcal{N}(0,1)$, i.e.,

$$
\lim_{n \to \infty} P(Y_n \leq y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-x^2/2} \, dx
$$

Also stated as:

- $\lim_{n \to \infty} Y_n \to \mathcal{N}(0,1)$
- $\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} X_i \to \mathcal{N} \left( \mu, \frac{\sigma^2}{n} \right)$ for $\mu = \mathbb{E}[X_i]$ and $\sigma^2 = \text{Var}(X_i)$
Normal approximation

• Let \( \overline{X} \) be the average of i.i.d. random variables \( X_1, \ldots, X_n \) with mean \( \mu \) and variance \( \sigma^2 \).

• CLT says that \( \frac{\sqrt{n} \cdot (\overline{X} - \mu)}{\sigma} \) approaches \( \mathcal{N}(0,1) \) standard unit normal

• Approximate using CDF of \( \mathcal{N}(0,1) \)

\[
\Phi(z) = P(Z \leq z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-x^2/2} \, dx \quad \text{for } Z \sim \mathcal{N}(0,1)
\]

Note: \( \Phi(z) \) has no closed form – generally given via tables

Within 1 standard deviation 68% within 2 standard deviations 95%, 3 s.d.’s 99%
### Table of $\Phi(z)$ CDF of Standard Normal Distribution

<table>
<thead>
<tr>
<th>$z$</th>
<th>0.00</th>
<th>0.01</th>
<th>0.02</th>
<th>0.03</th>
<th>0.04</th>
<th>0.05</th>
<th>0.06</th>
<th>0.07</th>
<th>0.08</th>
<th>0.09</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>0.5000</td>
<td>0.5398</td>
<td>0.5772</td>
<td>0.6128</td>
<td>0.6495</td>
<td>0.6821</td>
<td>0.7120</td>
<td>0.7389</td>
<td>0.7611</td>
<td>0.7808</td>
</tr>
<tr>
<td>0.1</td>
<td>0.5398</td>
<td>0.5739</td>
<td>0.6070</td>
<td>0.6378</td>
<td>0.6650</td>
<td>0.6900</td>
<td>0.7130</td>
<td>0.7344</td>
<td>0.7528</td>
<td>0.7681</td>
</tr>
<tr>
<td>0.2</td>
<td>0.5772</td>
<td>0.6070</td>
<td>0.6378</td>
<td>0.6650</td>
<td>0.6900</td>
<td>0.7130</td>
<td>0.7344</td>
<td>0.7528</td>
<td>0.7681</td>
<td>0.7808</td>
</tr>
<tr>
<td>0.3</td>
<td>0.6128</td>
<td>0.6378</td>
<td>0.6650</td>
<td>0.6900</td>
<td>0.7130</td>
<td>0.7344</td>
<td>0.7528</td>
<td>0.7681</td>
<td>0.7808</td>
<td>0.7930</td>
</tr>
<tr>
<td>0.4</td>
<td>0.6495</td>
<td>0.6650</td>
<td>0.6900</td>
<td>0.7130</td>
<td>0.7344</td>
<td>0.7528</td>
<td>0.7681</td>
<td>0.7808</td>
<td>0.7930</td>
<td>0.8039</td>
</tr>
<tr>
<td>0.5</td>
<td>0.6821</td>
<td>0.6900</td>
<td>0.7130</td>
<td>0.7344</td>
<td>0.7528</td>
<td>0.7681</td>
<td>0.7808</td>
<td>0.7930</td>
<td>0.8039</td>
<td>0.8129</td>
</tr>
<tr>
<td>0.6</td>
<td>0.7120</td>
<td>0.7344</td>
<td>0.7528</td>
<td>0.7681</td>
<td>0.7808</td>
<td>0.7930</td>
<td>0.8039</td>
<td>0.8129</td>
<td>0.8211</td>
<td>0.8281</td>
</tr>
<tr>
<td>0.7</td>
<td>0.7344</td>
<td>0.7528</td>
<td>0.7681</td>
<td>0.7808</td>
<td>0.7930</td>
<td>0.8039</td>
<td>0.8129</td>
<td>0.8211</td>
<td>0.8281</td>
<td>0.8344</td>
</tr>
<tr>
<td>0.8</td>
<td>0.7528</td>
<td>0.7681</td>
<td>0.7808</td>
<td>0.7930</td>
<td>0.8039</td>
<td>0.8129</td>
<td>0.8211</td>
<td>0.8281</td>
<td>0.8344</td>
<td>0.8398</td>
</tr>
<tr>
<td>0.9</td>
<td>0.7681</td>
<td>0.7808</td>
<td>0.7930</td>
<td>0.8039</td>
<td>0.8129</td>
<td>0.8211</td>
<td>0.8281</td>
<td>0.8344</td>
<td>0.8398</td>
<td>0.8446</td>
</tr>
<tr>
<td>1.0</td>
<td>0.7808</td>
<td>0.7930</td>
<td>0.8039</td>
<td>0.8129</td>
<td>0.8211</td>
<td>0.8281</td>
<td>0.8344</td>
<td>0.8398</td>
<td>0.8446</td>
<td>0.8487</td>
</tr>
<tr>
<td>1.1</td>
<td>0.7930</td>
<td>0.8039</td>
<td>0.8129</td>
<td>0.8211</td>
<td>0.8281</td>
<td>0.8344</td>
<td>0.8398</td>
<td>0.8446</td>
<td>0.8487</td>
<td>0.8524</td>
</tr>
<tr>
<td>1.2</td>
<td>0.8039</td>
<td>0.8129</td>
<td>0.8211</td>
<td>0.8281</td>
<td>0.8344</td>
<td>0.8398</td>
<td>0.8446</td>
<td>0.8487</td>
<td>0.8524</td>
<td>0.8559</td>
</tr>
<tr>
<td>1.3</td>
<td>0.8129</td>
<td>0.8211</td>
<td>0.8281</td>
<td>0.8344</td>
<td>0.8398</td>
<td>0.8446</td>
<td>0.8487</td>
<td>0.8524</td>
<td>0.8559</td>
<td>0.8594</td>
</tr>
<tr>
<td>1.4</td>
<td>0.8211</td>
<td>0.8281</td>
<td>0.8344</td>
<td>0.8398</td>
<td>0.8446</td>
<td>0.8487</td>
<td>0.8524</td>
<td>0.8559</td>
<td>0.8594</td>
<td>0.8624</td>
</tr>
<tr>
<td>1.5</td>
<td>0.8281</td>
<td>0.8344</td>
<td>0.8398</td>
<td>0.8446</td>
<td>0.8487</td>
<td>0.8524</td>
<td>0.8559</td>
<td>0.8594</td>
<td>0.8624</td>
<td>0.8651</td>
</tr>
<tr>
<td>1.6</td>
<td>0.8344</td>
<td>0.8398</td>
<td>0.8446</td>
<td>0.8487</td>
<td>0.8524</td>
<td>0.8559</td>
<td>0.8594</td>
<td>0.8624</td>
<td>0.8651</td>
<td>0.8677</td>
</tr>
<tr>
<td>1.7</td>
<td>0.8398</td>
<td>0.8446</td>
<td>0.8487</td>
<td>0.8524</td>
<td>0.8559</td>
<td>0.8594</td>
<td>0.8624</td>
<td>0.8651</td>
<td>0.8677</td>
<td>0.8699</td>
</tr>
<tr>
<td>1.8</td>
<td>0.8446</td>
<td>0.8487</td>
<td>0.8524</td>
<td>0.8559</td>
<td>0.8594</td>
<td>0.8624</td>
<td>0.8651</td>
<td>0.8677</td>
<td>0.8699</td>
<td>0.8719</td>
</tr>
<tr>
<td>1.9</td>
<td>0.8487</td>
<td>0.8524</td>
<td>0.8559</td>
<td>0.8594</td>
<td>0.8624</td>
<td>0.8651</td>
<td>0.8677</td>
<td>0.8699</td>
<td>0.8719</td>
<td>0.8739</td>
</tr>
<tr>
<td>2.0</td>
<td>0.8524</td>
<td>0.8559</td>
<td>0.8594</td>
<td>0.8624</td>
<td>0.8651</td>
<td>0.8677</td>
<td>0.8699</td>
<td>0.8719</td>
<td>0.8739</td>
<td>0.8758</td>
</tr>
<tr>
<td>2.1</td>
<td>0.8559</td>
<td>0.8594</td>
<td>0.8624</td>
<td>0.8651</td>
<td>0.8677</td>
<td>0.8699</td>
<td>0.8719</td>
<td>0.8739</td>
<td>0.8758</td>
<td>0.8776</td>
</tr>
<tr>
<td>2.2</td>
<td>0.8594</td>
<td>0.8624</td>
<td>0.8651</td>
<td>0.8677</td>
<td>0.8699</td>
<td>0.8719</td>
<td>0.8739</td>
<td>0.8758</td>
<td>0.8776</td>
<td>0.8792</td>
</tr>
<tr>
<td>2.3</td>
<td>0.8624</td>
<td>0.8651</td>
<td>0.8677</td>
<td>0.8699</td>
<td>0.8719</td>
<td>0.8739</td>
<td>0.8758</td>
<td>0.8776</td>
<td>0.8792</td>
<td>0.8807</td>
</tr>
<tr>
<td>2.4</td>
<td>0.8651</td>
<td>0.8677</td>
<td>0.8699</td>
<td>0.8719</td>
<td>0.8739</td>
<td>0.8758</td>
<td>0.8776</td>
<td>0.8792</td>
<td>0.8807</td>
<td>0.8821</td>
</tr>
<tr>
<td>2.5</td>
<td>0.8677</td>
<td>0.8699</td>
<td>0.8719</td>
<td>0.8739</td>
<td>0.8758</td>
<td>0.8776</td>
<td>0.8792</td>
<td>0.8807</td>
<td>0.8821</td>
<td>0.8835</td>
</tr>
<tr>
<td>2.6</td>
<td>0.8699</td>
<td>0.8719</td>
<td>0.8739</td>
<td>0.8758</td>
<td>0.8776</td>
<td>0.8792</td>
<td>0.8807</td>
<td>0.8821</td>
<td>0.8835</td>
<td>0.8848</td>
</tr>
<tr>
<td>2.7</td>
<td>0.8719</td>
<td>0.8739</td>
<td>0.8758</td>
<td>0.8776</td>
<td>0.8792</td>
<td>0.8807</td>
<td>0.8821</td>
<td>0.8835</td>
<td>0.8848</td>
<td>0.8859</td>
</tr>
<tr>
<td>2.8</td>
<td>0.8739</td>
<td>0.8758</td>
<td>0.8776</td>
<td>0.8792</td>
<td>0.8807</td>
<td>0.8821</td>
<td>0.8835</td>
<td>0.8848</td>
<td>0.8859</td>
<td>0.8870</td>
</tr>
<tr>
<td>2.9</td>
<td>0.8758</td>
<td>0.8776</td>
<td>0.8792</td>
<td>0.8807</td>
<td>0.8821</td>
<td>0.8835</td>
<td>0.8848</td>
<td>0.8859</td>
<td>0.8870</td>
<td>0.8880</td>
</tr>
<tr>
<td>3.0</td>
<td>0.8776</td>
<td>0.8792</td>
<td>0.8807</td>
<td>0.8821</td>
<td>0.8835</td>
<td>0.8848</td>
<td>0.8859</td>
<td>0.8870</td>
<td>0.8880</td>
<td>0.8887</td>
</tr>
</tbody>
</table>
Continuity Correction

Round to next integer!

To estimate probability that discrete RV lands in set $S$ of integers include all surrounding values that round to $S$.

For interval $\{a, \ldots, b\}$, compute probability for interval $\left[a - \frac{1}{2}, b + \frac{1}{2}\right]$. 
Parameter Estimation – Workflow

Distribution \( P(x; \theta) \) → Independent samples \( X_1, \ldots, X_n \) from \( P(x; \theta) \) → Estimation Algorithm → \( \hat{\theta} \)

\( \theta = \text{unknown} \) parameter

**Example:** coin flip distribution with unknown \( \theta = \) probability of heads

Observation: \( HTTHHHTHTHTTTTTHTHTTTTTHT \)

**Goal:** Estimate \( \theta \)
Maximum Likelihood Estimation (MLE)

1. **Input** Given $n$ i.i.d. samples $x_1, \ldots, x_n$ from parametric model with parameter (or vector of parameters) $\theta$.

2. **Likelihood** Define your likelihood $\mathcal{L}(x_1, \ldots, x_n \mid \theta)$.
   - For discrete $\mathcal{L}(x_1, \ldots, x_n \mid \theta) = \prod_{i=1}^{n} P(x_i \mid \theta)$
   - For continuous $\mathcal{L}(x_1, \ldots, x_n \mid \theta) = \prod_{i=1}^{n} f(x_i \mid \theta)$

3. **Log** Compute $\ln \mathcal{L}(x_1, \ldots, x_n \mid \theta)$

4. **Differentiate** Compute $\frac{\partial}{\partial \theta_j} \ln \mathcal{L}(x_1, \ldots, x_n \mid \theta)$ for each parameter in $\theta$ (also check discontinuities)

5. **Solve for** $\hat{\theta}$ by setting derivatives to 0 and solving for max.

Generally, you need to do a second derivative test to verify it is a maximum, but we won’t ask you to do that in CSE 312.
An estimation algorithm like MLE defines $\hat{\theta}_n$ as a function of the random variables $X_1, \ldots, X_n$. $\hat{\theta}_n(X_1, \ldots, X_n)$ is a r.v. whose expectation we can evaluate using LOTUS.

**Definition.** An estimator is **unbiased** if $\mathbb{E} [\hat{\theta}_n] = \theta$ for all $n \geq 1$. 

$\theta = \text{unknown}$ parameter

Independent samples $X_1, \ldots, X_n$ from $P(x; \theta)$

Estimation Algorithm

Parameter estimate $\hat{\theta}_n$
Estimators for the Normal Distribution

Normal outcomes $X_1, \ldots, X_n$ i.i.d. according to $\mathcal{N}(\mu, \sigma^2)$  
Assume: $\sigma^2 > 0$

$$\hat{\theta}_\mu = \frac{\sum^n_i X_i}{n}$$  
Sample mean (MLE) – Unbiased!

$$\hat{\theta}_{\sigma^2} = \frac{1}{n} \sum^n_{i=1} (X_i - \hat{\theta}_\mu)^2$$  
Population variance (MLE) – Biased!

$$S_n^2 = \frac{1}{n - 1} \sum^n_{i=1} (X_i - \hat{\theta}_\mu)^2$$  
Sample variance – Unbiased!

But population variance (like every MLE) is consistent in that $\lim_{n \to \infty} \mathbb{E}[\hat{\theta}_{\sigma^2}] = \sigma^2$. 
Markov chain

At each time step $t$
- Can be in one of a set of states
  - Work, Surf, Email
- If I am in some state $s$ at time $t$
  - the **labels of out-edges** of $s$ give the **probabilities** of moving to each of the states at time $t + 1$ (as well as staying the same)
    - so **labels on out-edges sum to 1**

E.g. If in Email, there is a 50-50 chance it will be in each of Work or Email at the next time step, but it will never be in state Surf in the next step.
Transition Probability Matrix and distribution of $X^{(t)}$

$$[q_W^{(t)}, q_S^{(t)}, q_E^{(t)}] \begin{bmatrix} 0.4 & 0.6 & 0 \\ 0.1 & 0.6 & 0.3 \\ 0.5 & 0 & 0.5 \end{bmatrix} = [q_W^{(t+1)}, q_S^{(t+1)}, q_E^{(t+1)}]$$

Vector-matrix multiplication

$M$ is the Transition Probability Matrix

Probability vector for state variable $X^{(t)}$ at time $t$: $q^{(t)} = [q_W^{(t)}, q_S^{(t)}, q_E^{(t)}]$

For all $t \geq 0$, $q^{(t+1)} = q^{(t)}M$

Equivalently, $q^{(t)} = q^{(0)}M^t$ for all $t \geq 0$
Stationary Distribution of a Markov Chain

**Definition.** The *stationary distribution of a Markov Chain* with $n$ states is the $n$-dimensional row vector $\pi$ such that

$$\pi M = \pi$$

and $\pi$ is a probability distribution.

Intuition: Distribution over states at next step is the same as the distribution over states at the current step.
Computing a Stationary Distribution

Stationary Distribution satisfies

- \( \pi = \pi \mathbf{M} \), where \( \pi = (\pi_W, \pi_S, \pi_E) \)
- \( \pi_W + \pi_S + \pi_E = 1 \)

\[ \pi_W = \frac{10}{34}, \quad \pi_S = \frac{15}{34}, \quad \pi_E = \frac{9}{34} \]

\[
[\pi_W, \pi_S, \pi_E] \begin{bmatrix} 0.4 & 0.6 & 0 \\ 0.1 & 0.6 & 0.3 \\ 0.5 & 0 & 0.5 \end{bmatrix} = [\pi_W, \pi_S, \pi_E]
\]

Solve system of equations:

- \( 0.4 \cdot \pi_W + 0.1 \cdot \pi_S + 0.5 \cdot \pi_E = \pi_W \)
- \( 0.6 \cdot \pi_W + 0.6 \cdot \pi_S = \pi_S \)
- \( 0.3 \cdot \pi_S + 0.5 \cdot \pi_E = \pi_E \)

\[ \pi_W + \pi_S + \pi_E = 1 \]
Intuition: $q^{(t)}$ is the distribution of being at each state at time $t$ computed by $q^{(t)} = q^{(0)} M^t$. Often as $t$ gets large $q^{(t)} \approx q^{(t+1)}$.

**Fundamental Theorem of Markov Chains**: For a Markov Chain that is aperiodic* and irreducible*, with transition probabilities $M$ and for any starting distribution $q^{(0)}$ over the states

$$\lim_{t \to \infty} q^{(0)} M^t = \pi$$

where $\pi$ is the stationary distribution of $M$ (i.e., $\pi M = \pi$).

*These concepts are way beyond us but they turn out to cover a very large class of Markov chains of practical importance.
Thank you from the 312 team!
Announcements / Logistics

• Do the class evaluation!
• Check the exam instructions post
  – Q&A session on Sunday, 2:00pm, over zoom!
• Longer office hours today for me (2:30pm – 3:20pm)