Prep: bring ID, water, glasses, jacket, pen and the handwritten version of this set of notes

You got this!

Distributions

Bernoulli Distribution: $X \sim Bernoulli(p)$

$\mathbb{P}[X=0]=(1-p)$	$\mathbb{E}[X]=p$
$\mathbb{P}[X=1]=p$	$\mathrm{Var}[X]=p(1-p)$

Binomial Distribution: $X \sim Binom(n,p)$

$\mathbb{P}[X=i]=\binom{n}{i}p^i(1-p)^{n-i},\ i=0,1,\dots,n$	
$\mathbb{E}[X]=np$	$\mathrm{Var}[X]=np(1-p)$

Geometric Distribution: $X \sim Geometric(p)$

$\mathbb{P}[X=i]=(1-p)^{i-1}p, i=1,2,\dots$	
$\mathbb{E}[X]=\frac{1}{p}$	$\mathrm{Var}[X]=\frac{1-p}{p^2}$
$pgf(x)=\frac{px}{1-(1-p)x}$	

Poisson Distribution: $X \sim Poisson(\lambda)$

$\mathbb{P}[X=i]=\frac{\lambda^i}{i!}e^{-\lambda}, i=0,1,2,\dots$	
$\mathbb{E}[X]=\lambda$	$\mathrm{Var}[X]=\lambda$
$X+Y \sim Poisson(\lambda+\mu)$	

Exponential Distribution: $X \sim Expo(\lambda)$

$f(x)=\begin{cases}\lambda e^{-\lambda x}, & x\ge 0\\ 0, & \text{otherwise}\end{cases}$	
$\mathbb{E}[X]=\frac{1}{\lambda}$	$\mathrm{Var}[X]=\frac{1}{\lambda^2}$
$F(x)=\mathbb{P}[X\le x]=1-e^{-\lambda x}$	

Uniform Continuous: $X \sim Uniform([a,b])$

$\mathbb{E}[X]=\frac{a+b}{2}$	$\mathrm{Var}[X]=\frac{(b-a)^2}{12}$
$f(x)=\frac{1}{b-a}$	$F(x)=\frac{x-a}{b-a}$

Normal Distribution: $X \sim N(\mu,\sigma^2)$

$X\sim N(\mu,\sigma^2)\to Z=\frac{X-\mu}{\sigma}\sim N(0,1)$	
$f(x)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	$f_Z(z)=\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}$

* No closed formula for CDF.

Distribution (Hacks)

Trick:

$$\min(X_1,X_2,\dots)\sim Geometric\big(1-\Pi(1-p_i)\big)$$

λ is like expected number of success

$$X\sim Binom\left(n,\frac{\lambda}{n}\right)$$

$$\mathbb{P}[X=i]\to\left(\frac{\lambda^i}{i!}\right)e^{-\lambda}\text{ as } n\to\infty$$

Memoryless Property

$$\mathbb{P}[X\ge x+y|X\ge x]=\mathbb{P}[X\ge y]$$

Useful trick:

$$\min(X_1,X_2,\dots)\sim Expo(\Sigma\lambda_i)$$

$$\mathbb{P}[X<Y]=\frac{\lambda_X}{\lambda_X+\lambda_Y}$$

λ: success rate per unit time

Functions on Random Variables

Covariance (bilinear)

$Cov[X,Y]=E[(X-\mu_X)(Y-\mu_Y)]=E[XY]-E[X]E[Y]$
$Cov[X,X]=Var[X]$
$Var[X+Y]=Var[X]+Var[Y]+2\,Cov[X,Y]$
$Cov[aX_1+bX_2,cY_1+dY_2]=ac\,Cov[X_1,Y_1]+ad\,Cov[X_2,Y_1]+bc\,Cov[X_2,Y_1]+bd\,Cov[X_2,Y_2]$

For independent X,Y: $Cov[X,Y]=0$ (converse not true)

Correlation

$Corr[X,Y]=\frac{Cov[X,Y]}{\sigma_X\sigma_Y}$
$X'=\frac{X-\mu_X}{\sigma_X}, Y'=\frac{Y-\mu_Y}{\sigma_Y}$ $-1\le Corr[X,Y]=Cov[X',Y']\le 1$
$Corr[X,Y]=1\Rightarrow Y=AX+B, A>0\ (Y'=X')$ $Corr[X,Y]=-1\Rightarrow Y=AX+B, A<0\ (Y'=-X'$

Miscellaneous Hacks

$$\int_{-\infty}^{\infty}e^{-\frac{x^2}{2}}\mathrm{d}x=\sqrt{2\pi}$$

$$\mathbb{P}[A|B]=\frac{\mathbb{P}[B|A]\mathbb{P}[A]}{\mathbb{P}[B]}$$

Discrete Tail Sum (X nonnegative)

$$\mathbb{E}[X]=\sum_{i=0}^{\infty}\mathbb{P}[X>i]$$

Continuous Tail sum (Z nonnegative)

$$\mathbb{E}[Z]=\int_0^{\infty}\mathbb{P}[Z\ge z]\,\mathrm{d}z$$

Probability Density Hacks

$$cdf_X(x)=\mathbb{P}[X\le x]$$

$$pdf_X(x)=\frac{\mathrm{d}}{\mathrm{d}x}cdf_X\Big|_x$$

Quick hacks:

$$\mathbb{P}[X\le Y]=\int_{-\infty}^{\infty}\mathbb{P}[x\le X\le x+\mathrm{d}x]\,\mathbb{P}[x\le Y]$$

$$\mathbb{E}[g(X)]=\int_{-\infty}^{\infty}g(x)\mathbb{P}[X=x]=\int_{-\infty}^{\infty}g(x)f_X(x)\mathrm{d}x$$

$$n!\sim\sqrt{2\pi n}\left(\frac{n}{e}\right)^n$$

Properties of Conditional Expectations

$$\mathbb{E}[Y|X] = f(X)$$

1. (Linearity)
$$\mathbb{E}[a_1Y_1 + a_2Y_2|X] = a_1\mathbb{E}[Y_1|X] + a_2\mathbb{E}[Y_2|X]$$
2. (Factoring Known Values)
$$\mathbb{E}[h(X)Y|X] = h(X)\mathbb{E}[Y|X]$$
3. (Smoothing)
$$\mathbb{E}\big[\mathbb{E}[Y|X]\big] = \mathbb{E}[Y]$$
4. (Independence) If X, Y independent:
$$\mathbb{E}[Y|X] = \mathbb{E}[Y]$$

Estimation and Linear Regression

Case #1: Know the joint distribution

Want to find $L[Y|X] = g(X) = a + bX$ that minimizes cost function $C(g)$

$$C(g) = \mathbb{E}[|Y - g(X)|^2] = \mathbb{E}[|Y - a - bX|^2]$$

$$b = \frac{\mathrm{Cov}[X,Y]}{\mathrm{Var}[X]}$$

$$a = \mathbb{E}[Y] - \mathbb{E}[X] \cdot \frac{\mathrm{Cov}[X,Y]}{\mathrm{Var}[X]}$$

$$L[Y|X] = \mathbb{E}[Y] + \frac{\mathrm{Cov}[X,Y]}{\mathrm{Var}[X]}(X - \mathbb{E}[X])$$

Case #2: Linear regression

Observed K samples $(X_1, Y_1), (X_2, Y_2), \dots (X_K, Y_K)$

Choose a, b to minimize $\frac{1}{K}\sum_{k=1}^{K}|Y_k - a - bX_k|^2$

$$\mathbb{E}[|Y - a - bX|^2] = \frac{1}{K}\sum_{k=1}^{K}|Y_k - a - bX_k|^2$$

As $K \to \infty$, linear regression approaches LLSE, assuming (X_k, Y_k) are i.i.d.

MMSE (Minimum Mean Squared Error)

$$g(X) = \mathbb{E}[Y|X]$$

(Orthogonality)

$$g(X) = \mathbb{E}[Y|X]$$
$$\Leftrightarrow$$
$$\mathbb{E}[(Y - g(X))\Phi(X)] = 0 \;\forall\; \Phi(X)$$

Markov Chain

P: transition matrix, π_0: initial distribution vector

$\pi_n = P^n\pi_0$	$\pi = P\pi$

<u>Irreducibility</u>: can go from every state i to every other state j in finite moves.

<u>Theorem</u>: For finite irreducible Markov chain, for any π_0, exists unique invariant distribution π s.t.

$$\lim_{n\to\infty}\frac{1}{n}\sum_{m=0}^{n-1}\mathbb{P}[X_m = i] = \pi(i)$$

Probabilistic Bounding

<u>Definitions</u>:

$$\mathbb{P}[|\hat{p} - p| \geq \varepsilon] \leq 1 - \delta$$

ε: error / accuracy / tolerance; $1 - \delta$: confidence

<u>Toolbox</u>:

- [Markov] Nonnegative RV X, finite mean
$$\mathbb{P}[X \geq c] \leq \frac{\mathbb{E}[X]}{c},\ \ c > 0$$
- [Generalized Markov] Y not necessarily nonnegative, finite mean; $c, r > 0$
$$\mathbb{P}[|Y| \geq c] \leq \frac{\mathbb{E}[|Y|^r]}{c^r}$$
- [Extended Markov] X not necessarily nonnegative; $\Phi(X)$ nonnegative function, monotonically increasing for $x > 0$; $\alpha > 0$
$$\mathbb{P}[X \geq \alpha] \leq \frac{\mathbb{E}[\Phi(X)]}{\Phi(\alpha)}$$
- [Chebyshev] $c > 0$
$$\mathbb{P}[|X - \mu| \geq c] \leq \frac{\mathrm{Var}[X]}{c^2}$$
$$\mathbb{P}[|X - \mu| \geq k\sigma] \leq \frac{1}{k^2}$$
- [Cantelli] $\alpha > 0$
$$\mathbb{P}[X - \mathbb{E}[X] \geq \alpha] \leq \frac{\sigma^2}{\alpha^2 + \sigma^2}$$
- [Law of Large Numbers] $X_1, \dots, X_n$ i.i.d. RV with $\mathbb{E}[X_i] = \mu < \infty$. Define $S_n = X_1 + \cdots + X_n$
$$\forall\varepsilon\ \lim_{n\to\infty}\mathbb{P}\left[\left|\frac{1}{n}S_n - \mu\right| < \varepsilon\right] = 1$$
- [Central Limit Theorem] Distribution of sample average $\frac{S_n}{n}$ approaches a **normal distribution** with mean μ and variance $\frac{\sigma^2}{n}$.
$$\frac{\frac{S_n}{n} - \mu}{\sqrt{\sigma^2/n}} = \frac{S_n - n\mu}{\sigma\sqrt{n}} \sim N(0,1)$$
$$\mathbb{P}\left[\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq c\right] \to \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{c} e^{-\frac{x^2}{2}}\mathrm{d}x$$

Continuous Probability

<u>PDF</u>: $f: \mathbb{R} \to \mathbb{R}$, nonnegative, normalized

$$\mathbb{P}[a \leq X \leq b] = \int_a^b f(x)\,\mathrm{d}x$$

<u>CDF</u>:

$$F(x) = \mathbb{P}[X \leq x] = \int_{-\infty}^{x} f(z)\,\mathrm{d}z$$

<u>Joint Density</u>: $f: \mathbb{R}^2 \to \mathbb{R}$ that is nonnegative and normalized. $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}x\,\mathrm{d}y = 1$.

$$\mathbb{P}[x \leq X \leq x + \mathrm{d}x, y \leq Y \leq y + \mathrm{d}y] = f(x,y)\,\mathrm{d}x\,\mathrm{d}y$$

Definition:
For $i \in \mathcal{H}$, define:
$d(i) := \gcd\{n > 0 \mid P^n(i,i) = \mathbb{P}[X_n = i \mid X_0 = i] > 0\}$
An *irreducible* Markov chain is **aperiodic** if $d = 1$, else **periodic** with period d.

Theorem: For aperiodic Markov Chain, $\mathbb{P}[X_n = i] \to \pi(i)$ as $n \to \infty$.

Classic Problems:

Expected Steps	$state_A$ before $state_B$
$\beta(state_1) = 1 + \sum \beta(state_i)$	Set $\alpha(B) = 0$ and $\alpha(A) = 1$ and solve.

Techniques:
- Additional start and end state
- Clumping of equivalent states
- Redefine states/transitions
- Check all outgoing edges sum up to 1.

Countability

Techniques
1. Bijection
2. $|S_1| \le |S_2|$ and $|S_2| \le |S_1|$
3. Diagonalization, show $\in S$
4. Subset of countable
5. Superset of uncountable
6. Reduction (to solving Turing)
7. Self-referencing contradictions

Countable	Uncountable
• $\mathbb{N}, \mathbb{Q}, \mathbb{Z}, \mathbb{N} \times \mathbb{N}$ • Binary strings • Subset T of countable set S	• $\mathbb{R}$ • $x \in [0,1], \mathbb{R}$ • $\mathcal{P}(\mathbb{N})$

Final Checks
- Define all random variables.
- Check the domain of PMF, PDF, CDF.
- For PMF diagrams, draw 0 for "elsewhere".

Last Resorts
- PIE (Midterm 1 horror)
- Difference method, hockey stick theorem
- If PDF method fails, work with CDF
- Indicator variable approach + algebra
- Consider other forms of indicators
- Union bound

$$\mathbb{P}\left[\bigcup A_i\right] \le \sum \mathbb{P}[A_i]$$

Independence: X, Y are independent if $\forall\, a, b, c, d$:

$$\mathbb{P}[a \le X \le b, c \le Y \le d] = \mathbb{P}[a \le X \le b] \cdot \mathbb{P}[c \le Y \le d]$$

The joint density becomes separable:

$$f(x,y) = f_X(x) f_Y(y)$$

$f(x) = \dfrac{dF(x)}{dx}$	$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x)\,dx$
$\mathrm{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \int_{-\infty}^{\infty} x^2 f(x)dx - \left(\int_{-\infty}^{\infty} x f(x) dx\right)^2$	

Computability

```
HALT(P, I) # True if P(I) halts, else False
TURING(P):
    if HALT(P, P) == "halts": ∞
    else: "halts"
```

Un-computable problems:
- Halting: Does program P halt on input I?
- Variant: Does program P halt on 0? (or any other input for that matter)

```
HALT(P, x):
    def P'(y):
        return P(x)
    return  VARIANT_HALT(P')
```

- Variant: A program $P(F, x, y)$ that returns `true` if $F(x) = y$

```
HALT(F, x):
    def Q(y):
        F(x)
        return 0
    return P(Q, x, 0)
```

- Variant: A program $P(F, G)$ that returns `true` if F, G halt on same set of input

```
def HALT(F, x):
    def Q(y): loop
    def R(y):
        if y == x: return F(x)
        else: loop
    return not P(Q, R)
```

- Variant: A program $E2(P, x)$ that returns `true` if P runs an even number of lines on x.

```
def TuringE(P):
    if E2(P, P):
        print("Filler")
    return
```

Logic and Function

Implies: $P \Rightarrow Q \equiv \neg(P \wedge \neg Q) \equiv \neg P \vee Q$
Converse: $Q \Rightarrow P$
Inverse: $\neg P \Rightarrow \neg Q$
Contrapositive: $\neg Q \Rightarrow \neg P$
De Morgan's Law:

$$\neg(P \wedge Q) \equiv (\neg P) \vee (\neg Q)$$
$$\neg(P \vee Q) \equiv (\neg P) \wedge (\neg Q)$$
$$\neg(\forall x\, P) \equiv \exists x\, (\neg P)$$
$$\neg(\exists x\, P) \equiv \forall x\, (\neg P)$$

Final Checks

- *Check if it is Stable Matching or Propose-and-Reject problem.*
- Polynomials in GF must mod coefficients
- RSA: write N, e, d explicitly to avoid errors
- $0 \in \mathbb{N}$ for this class
- Be careful of the bound in vertex coloring
- Be careful of base cases for graph
- Counting: rotations / inversions included?

Stable Matching

When a candidate does not immediately reject a job, the job is still assumed to "propose" to the candidate on the next day.

[Improvement Lemma] Candidate's matching can only improve. (exchange argument)

Job-Propose and Reject always terminate with matching (contradiction), gives job-optimal and candidate-pessimal (contradiction).

Job	I	II	III	C	I	II	III
A	1	2	3	1	B	C	A
B	2	3	1	2	C	A	B
C	3	1	2	3	A	B	C

$\{(A,1),(B,2),(C,3)\}$, $\{(A,3),(B,1),(C,2)\}$, $\{(A,2),(B,3),(C,1)\}$ are all stable.

Stable Matching Trivia

- Always exists a candidate who is not proposed to until the last day.
- Propose-and-reject algorithm must terminate in at most $(n-1)^2+1$ days.
- For even $n \geq 2$, exists instance of stable matching of n jobs and candidates with at least $2^{n/2}$ distinct stable matching. (induct on n)

Graph Theory (Definition)

Path: a sequence of edges, vertices distinct.
Cycle: a path (distinct vertices) with $v_1 = v_n$
Walk: a path without distinct vertices condition
Tour: a walk with $v_1 = v_n$

A cycle is a walk. A tour is a walk.

Eulerian walk: uses all edge exactly once.
Eulerian tour: walk that ends at start vertex
Hamiltonian walk/cycle: a walk/cycle that visits all vertices *exactly once*.
Hypercube ($\dim N$): 2^N nodes, $N2^{N-1}$ edges

Function

$$f(X) = \{y \mid \exists x \in X \text{ s.t. } y = f(x)\}$$

$$f^{-1}(Y) = \{x \mid f(x) \in Y\}$$

Graph Theory

Lines of Attack: Induction on $|V|$, $|E|$, tree-shaving (removal of leaf node), Eulerian tours, pigeonhole,

Euler's Theorem: Planar graphs with $v \geq 3$ satisfy $v + f = e + 2$
Corollary: All planar graphs satisfy $e \leq 3v - 6$
$K_{3,3}$ **Variant**: $e \leq 2v - 4$
Kuratowski's Theorem: A graph is planar iff it doesn't contain K_5 or $K_{3,3}$

Coloring

- A graph with max degree k is $k+1$ colorable. (induct on $|V|$)
- A connected graph of max degree $d \geq 2$ can be vertex colored with d colors so long as $\exists$ vertex with degree $< d$. ($|V|$)
- Graph with max degree $d \geq 1$ can be edge colored in $2d-1$ colors. (induct $|E|$)

Error Correcting Codes

Message of n packets $(m_1, m_2, \ldots, m_n)$ where $m_i = P(i)$ for some polynomial P of at most degree $n-1$.

Bounding of $GF(q)$, q prime:

$$q \geq \max(m_i + 1, n + k)$$
$$q \geq \max(m_i + 1, n + 2k)$$

Error Correction:

$$Q(x) = P(x)E(x)$$

• In a job propose algorithm, jobs can't lie to improve their own outcomes, but can to improve others. • If candidate rejects a job in JPA, there is no stable matching in which the candidate and job is paired. • If a candidate misbehaves (rejects falsely), then it is the only candidate that can be in a rogue couple.	$$Q(x_i) = r_i E(x_i)$$ $$E(x) = (x - e_1) \dots (x - e_k)$$ Fractional variant: $$n'(1-\alpha) = n \Longrightarrow n' = \frac{n}{1-\alpha}$$ $$n'(1-2\alpha) = n \Longrightarrow n' = \frac{n}{1-2\alpha}$$
RSA	Secret Sharing
Key (N, e, d). (N, e) is public. d is private. $N = pq$ where p, q are large primes. p, q must be secret, but if forgotten it's fine. Only requires d to decode. $$\big(e, (p-1)(q-1)\big) = 1$$ $$d^{-1} \equiv e \pmod{(p-1)(q-1)}$$ $$E(x) = x^e \pmod{N}$$ $$D(x) = x^d \pmod{N}$$ Security relies on the computational intractability of obtaining x in $y = x^e \pmod{N}$	Bounding of $GF(q)$, q prime: Secret sharing among m people (the $+1$ comes from the secret): $$q \geq \max(s+1, m+1)$$ Can delegate sub-polynomials for hierarchy. Spy variants: spies can corrupt messages.