Contents

Spring 2025, taught by Professor Kannan Ramchandran.

Chapter 1

CTMC

$$\pi Q = 0$$

where Q is a rate matrix where each row sums up to 0.

We show reversibility using DBE: $\pi_i q_{ij} = \pi_j q_{ji} \quad \forall i, j$.

Jump chain (embedded DTMC) does not have self-loops.

$$\pi_{CTMC}(x) = \frac{\frac{1}{Q(x)}\pi_{DTMC}(x)}{\sum_y \frac{1}{Q(y)}\pi_{DTMC}(y)} \tag{1.1}$$

Problem: [Sp20 Q7 Chair Game] Sean independently stands up/down at rate of 2, Will at rate of 3.

Solution: Split into cases of $(0,0)_A, (1,1)_D, (0,1)_B, (1,0)_C$:

$$\beta_A = \frac{1}{5} + \frac{3}{5}\beta_B + \frac{2}{5}\beta_C \tag{1.2}$$

1 Uniformization

Pick $q \geq \max_x Q(x)$. Then, we have:

$$P = I + \frac{1}{q}Q \tag{1.3}$$

Solving for $\pi P = \pi$ yields $\pi_{\text{uniformized}} = \pi_{CTMC}$.

Chapter 2

Random graphs

Erdős–Rényi graph $G \sim \mathcal{G}(n,p)$ has n vertices, where each edge appears with probability p.

G_0 is some graph with n vertices and m edges:

$$\mathbb{P}(G = G_0) = p^m (1-p)^{\binom{n}{2}-m}$$

The distribution of D, a degree of an arbitrary vertex, is $\text{Bin}(n-1,p)$.

The probability that any vertex is isolated is $(1-p)^{n-1}$.

Fact: Poisson approximation: $\text{Bin}(n,p) \approx \text{Poisson}(np)$.
Stirling's approximation: $n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$ and $\ln n! \approx n \ln n - n$.

Theorem 1 (Sharp threshold)**.** *Let $p(n) := \lambda \frac{\ln n}{n}$ for a constant $\lambda > 0$.*

(1) if $\lambda < 1$, then $\mathbb{P}(G$ is connected$) \to 0$

(2) if $\lambda > 1$, then $\mathbb{P}(G$ is connected$) \to 1$

i.e. the graph is connected with high probability if $p(n) >> \frac{\ln n}{n}$.

Fact: Taylor's expansion: $\ln(1-x) \approx -x$ for small x.

Proof. Assume $\lambda < 1$. Let X_n be the number of isolated nodes in G.

It is sufficient to show that $\mathbb{P}(X_n > 0) \to 1$ as $n \to \infty$.

Let $q := (1-p)^{n-1}$ be the probability a node is isolated.

$\mathbb{E}[X_n] = n(1-p)^{n-1} = nq$: deifne $X_n = \sum_{i=1}^n I_i$ where I_i indicates whether vertex i is isolated.

$$\ln \mathbb{E}[X_n] = \ln n + (n-1)\ln(1-p) \approx \ln n - (n-1)\lambda \frac{\ln n}{n} \to \infty \tag{2.1}$$

$\text{var}(X_n) = n\,\text{var}(I_i) + n(n-1)\,\text{cov}(I_1, I_2)$.

Note that $\mathbb{E}[I_1 I_2] = \mathbb{P}(\text{nodes 1, 2 are isolated}) = (1-p)^{2n-3} = \frac{q^2}{1-p}$ and $\text{cov}(I_1, I_2) = \mathbb{E}[I_1 I_2] - q^2 = \frac{pq^2}{1-p}$ Use the *second-moment method*:

$$\begin{aligned} \mathbb{P}(X_n = 0) &\le \mathbb{P}(|X_n - \mathbb{E}[X_n]| \ge \mathbb{E}[X_n]) \\ &\le \frac{\text{var}(X_n)}{\mathbb{E}[X_n]^2} \\ &= \frac{nq(1-q) + n(n-1)\frac{pq^2}{1-p}}{n^2 q^2} = \frac{1-q}{nq} + \frac{n-1}{n}\frac{p}{1-p} \to 0 \end{aligned} \tag{2.2}$$

□

Problem: [Fa22 Q1 Random Cut of a Random Graph] Let $G \sim \mathcal{G}(100, 1/4)$ where a random cut of G contains vertex with probability $1/3$. Find expected number of edges in a cut.

Solution: Expected number of edges in a cut of size K is $k(n-K)p$:

$$\begin{aligned} \sum_{k=0}^{n} \binom{n}{k} q^k (1-q)^{n-k} k(n-k)p = p\,\mathbb{E}[K(n-K)] \quad K \sim \text{Bin}(n, q) \\ = p(n\,\mathbb{E}[K] - \mathbb{E}[K^2]) \\ = p(n \cdot nq - (nq(1-q) + n^2 q^2)) \\ = pqn(n - 1 + q - nq) = pqn(n-1)(1-q) \end{aligned} \tag{2.3}$$

Problem: [HW 11 Subcritical Forest] Let $G \sim \mathcal{G}(n, p(n))$, where $p(n) = o(\frac{1}{n})$, which is called **subcritical phase**.

(i) Let X_n be the number of cycles in the graph. Show that $\mathbb{E}[X_n] \to 0$.

(ii) Show that $\mathbb{P}(G \text{ is a forest}) \to 1$ as $n \to \infty$.

Solution:

(i) *Proof.* Let Y_k be the number of cycles of length k, where $\mathbb{E}[Y_k] = \binom{n}{k} p(n)^k k! \frac{1}{k}\frac{1}{2}$ ($k!$ possible orderings, k possible starting vertices, undirected).

$$\mathbb{E}[X_n] = \sum_{k=3}^{n} \mathbb{E}[Y_k] = \sum_{k=3}^{n} \binom{n}{k} p(n)^k \frac{(k-1)!}{2} = \sum_{k=3}^{n} \frac{(np(n))^k}{2k} \le \sum_{k=3}^{n} (np(n))^k \to 0 \tag{2.4}$$

□

(ii) *Proof.* Using Markov's inequality:

$$\mathbb{P}(X_n > 0) = \mathbb{P}(X_n \geq 1) \leq \mathbb{E}[X_n] = 0 \tag{2.5}$$

□

Chapter 3

Hypothesis testing

Likelihood:

$$L(y) = \frac{f_{Y|H_1}(y)}{f_{Y|H_0}(y)} \tag{3.1}$$

Decision rule: accept H_1 if $L(y) > c$, accept H_1 w.p. γ if $L(y) = c$.

1 Neyman-Pearson rule

Intuition: we want to choose to accept or reject hypothesis given a single observation.

$$\begin{aligned} &\max_{\hat{X}} PCD := \mathbb{P}(\hat{X} = 1|X = 1) = \sum_{y} \mathbb{P}(\hat{X} = 1|Y = y) \cdot \mathbb{P}(Y = y|X = 0) \\ &\text{s.t. } PFA := \mathbb{P}(\hat{X} = 1|X = 0) \leq \beta \end{aligned} \tag{3.2}$$

for some fixed $\beta \in [0, 1]$.

Key: evaluate for each value of y.

Probability of false alarm (PFA): $\mathbb{P}(\hat{H} = 1|H = 0)$.

Probability of correct detection (PCD): $\mathbb{P}(\hat{H} = 1|H = 1)$.

2 MAP, MLE

MLE assumes uniform prior distribution, while MAP incorporates some information about argument.

$$\begin{aligned} \hat{\theta}_{MLE} &= \arg\max_{\theta} \mathbb{P}(\text{data}|\theta) \\ \hat{\theta}_{MAP} &= \arg\max_{\theta} \mathbb{P}(\text{data}|\theta)\, \mathbb{P}(\theta) \end{aligned} \tag{3.3}$$

Example: [German Tank Problem] Estimate total number of German tanks N given any two serial numbers X_1 and X_2.

$$\hat{N}_{MLE} = \arg\max_{n \geq \max(x_1,x_2)} \mathbb{P}(X_1 = x_1, X_2 = x_2 | N = n) = \arg\max_{n \geq \max(x_1,x_2)} \frac{1}{\binom{n}{2}} = \arg\min_{n \geq \max(x_1,x_2)} \binom{n}{2} = \max(x_1, x_2)$$

since those two samples could be any unordered pair.

Problem: [Disc 11 Q1] Decision rule: accept $X = 1$ if $Y > t$. Else, accept $X = c$.

Note that $Y \sim \text{Exp}(X)$.

$$\begin{aligned} &L(y) = \frac{c}{e^{y(c-1)}} \quad \text{decreasing in } y \\ &\mathbb{P}(\hat{X} = 1 | X = c) = \mathbb{P}(Y > t | X = c) = e^{-ct} \leq 0.05 \end{aligned} \tag{3.4}$$

so $-ct = \log \frac{1}{20}$ and $t = \frac{\log 20}{c}$.

Chapter 4

Hilbert space of RVs

Fact: None of Hilbert space conditions are strong enough to imply independence, including orthogonality!

1. $\mathbb{E}[XY] = \langle X, Y \rangle = \text{cov}(X, Y)$ if X, Y zero-mean
2. $\mathbb{E}[XY] = 0 \iff X \perp Y$
3. Orthogonality principle: $\mathbb{E}[(Y - \mathbb{L}[Y|X])X] = 0$, or $X \perp Y - \mathbb{L}[Y|X]$
4. Projection: $\hat{Y} = \mathbb{L}[Y|X] = \mathbb{E}[Y] + \frac{\text{cov}(X,Y)}{\text{var}(X)}(X - \mathbb{E}[X])$
5. Norm: $\|X\| = \sqrt{\langle X, X \rangle} = \sqrt{\mathbb{E}[X^2]} = \sqrt{\text{var}(X) + \mathbb{E}[X]^2}$

The expectation of a RV X always minimizes MSE, where we are projecting into 1:

$$\mathbb{E}[X] = \arg\min_{x \in \mathbb{R}} \mathbb{E}[(X - x)^2] \tag{4.1}$$

Fact: $Y - \mathbb{L}[Y|X]$ is called *innovation*, since it represents the new information that was not predictable from previous observations.

1 Minimum Mean Square Error, Linear Least Square Estimator

$$\begin{aligned} MMSE[Y|X] &= \mathbb{E}[Y|X] \\ LLSE[Y|X] &= \mathbb{E}[Y] + \frac{\text{cov}(X,Y)}{\text{var}(X)}(X - \mathbb{E}[X]) = \mathbb{L}[Y|X] = aX + b \text{ is the best linear approximation} \end{aligned} \tag{4.2}$$

Note that $\mathbb{L}[Y|X]$ is orthogonal to all *linear* functions of X, but not all functions of X in general.

Theorem 2 (Orthogonal LLSE update)**.**

$$\mathbb{L}[Y|X,Z] = \mathbb{L}[Y|X] + \mathbb{L}[Y|Z - L[Z|X]] \tag{4.3}$$

where $X \perp Z - \mathbb{L}[Z|X]$*.*

Fact: $\mathbb{L}[Y|X,Z] = \mathbb{L}[Y|X] + \mathbb{L}[Y|Z]$ iff X, Y, Z are zero-mean and $X \perp Z$.

Lemma 1. *(a)* $\mathbb{E}[(X - \mathbb{E}[X|Y])\phi(Y)] = 0 \quad \forall$ *function* $\phi(\cdot)$

(b) if there exists a function $g(Y)$ *s.t.* $\mathbb{E}[(X - g(Y))\phi(Y)] = 0 \quad \forall\phi(\cdot)$*, then* $g(Y) = \mathbb{E}[X|Y]$

Lemma 2. *If* $MMSE[Y|X] = \mathbb{E}[Y|X]$ *is linear, it is equal to* $LLSE[Y|X]$*.*

Chapter 5

Jointly Gaussian

Random variables X, Y are jointly Gaussian iff any their linear combination $aX + bY$ is Gaussian.

1. Jointly Gaussian RVs X, Y are independent if $\text{cov}(X, Y) = 0$ (sufficient condition).
2. Any linear transformation of JG RVS is also JG, i.e. if (X, Y) is JG then $(aX + bY, cX + dY)$ is JG.
3. If X, Y are JG, then $MMSE = LLSE$.
4. If X, Y are JG, they have marginal Gaussian distributions. The converse is not true.