MATH 7241: Probability 1 FALL 2022 TEST 1

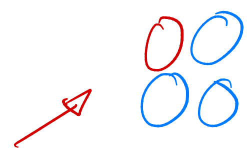
NAME: SOLUTIONS

1) A box contains four balls, each colored either Red or Blue. At each step a ball is randomly chosen from the box, and a fair coin is tossed. If the coin comes up Heads the ball is put back into the box; if the coin comes up Tails then a ball of the other color is put back into the box. Let X_n be the number of Red balls in the box after n steps.

a). Find the transition matrix of the chain $X_0, X_1, \ldots$.

$X_n \in \{0, 1, 2, 3, 4\}$.

$$P = \begin{array}{cc} \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{8} & \frac{1}{2} & \frac{3}{8} & 0 & 0 \\ 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 \\ 0 & 0 & \frac{3}{8} & \frac{1}{2} & \frac{1}{8} \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix} & \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \\ \begin{array}{ccccc} 0 & \;\;1 & \;\;2 & \;\;3 & \;\;4 \end{array} & \end{array}$$



b). Either compute $\lim_{n\to\infty} \mathrm{P}(X_n = 2)$, or explain why the limit does not exist.

Chain is irreducible, aperiodic $\Rightarrow \lim_{n\to\infty} \mathbb{P}(X_n = 2) = w_2$

$$w = \frac{1}{16}(1, 4, 6, 4, 1)$$

$$\Rightarrow \lim_{n\to\infty} \mathbb{P}(X_n = 2) = \frac{6}{16} = \frac{3}{8}$$

2) For a branching process calculate the probability of extinction when the offspring probabilities are $p_0 = 1/4$, $p_1 = 1/4$, $p_2 = 3/8$, $p_3 = 1/8$.

$$\mathbb{E}[Z] = \frac{1}{4} + 2\left(\frac{3}{8}\right) + 3\left(\frac{1}{8}\right) > 1 \quad \Rightarrow \quad \rho < \underline{1}$$

solve

$$s = \phi(s) = \frac{1}{4} + \frac{1}{4}s + \frac{3}{8}s^2 + \frac{1}{8}s^3$$

$$\Leftrightarrow \quad s^3 + 3s^2 - 6s + 2 = 0$$

$$\Leftrightarrow \quad (s-1)(s^2 + 4s - 2) = 0$$

$$\downarrow$$

$$s = \frac{-4 \pm \sqrt{16+8}}{2}$$

$$= -2 \pm \sqrt{6}.$$

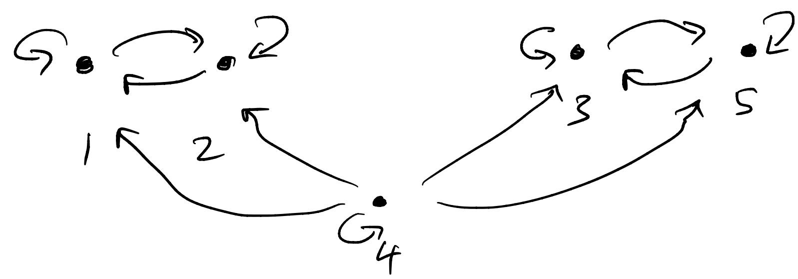
Now $0 < \rho < 1 \Rightarrow \rho = -2 + \sqrt{6}.$

3) Consider the following transition probability matrix for a Markov chain on 5 states:

$$P = \begin{pmatrix} 1/4 & 3/4 & 0 & 0 & 0 \\ 3/4 & 1/4 & 0 & 0 & 0 \\ 0 & 0 & 1/3 & 0 & 2/3 \\ 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \\ 0 & 0 & 2/3 & 0 & 1/3 \end{pmatrix}$$

Number the states $\{1, 2, 3, 4, 5\}$ in the order presented.

a). Write down two different stationary distributions for the chain.



$\{4\}$ transient

$\{1,2\}$ irreducible, closed $\Rightarrow$ $w = \frac{1}{2}(1,1,0,0,0)$

$\{3,5\}$ irreducible closed $\Rightarrow$ $w = \frac{1}{2}(0,0,1,0,1)$.

b). Starting from state 4, find the probability to eventually reach state 1.

By symmetry, since $\{4\}$ is transient

$\mathbb{P}(\text{eventually reach } \{1,2\} \mid X_0 = 4) = \frac{1}{2}$.

Since $\{1,2\}$ is irreducible $\Rightarrow$ if chain enters $\{1,2\}$, it must reach 1.

$\Rightarrow \mathbb{P}(\text{eventually reach } 1 \mid X_0 = 4) = \frac{1}{2}$

4) Determine whether or not the following chain is reversible:

$$P = \begin{array}{cc} \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/2 & 0 & 1/2 \\ 0 & 1/2 & 1/2 \end{pmatrix} & \begin{matrix} 1 \\ 2 \\ 3 \end{matrix} \\ \begin{matrix} 1 & \quad 2 & \quad 3 \end{matrix} & \end{array}$$

$w_i P_{ij} = w_j P_{ji}$

all i, j

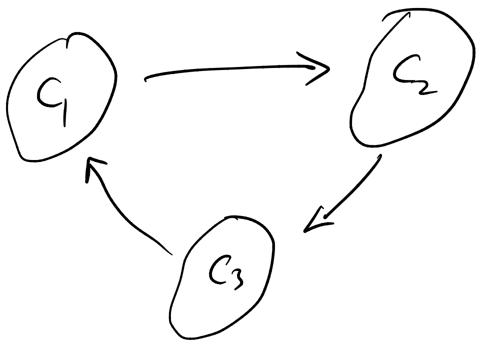
$P_{13} > 0, \quad P_{31} = 0$

$\Rightarrow$ cannot satisfy $w_1 P_{13} = w_3 P_{31}$

$\Rightarrow$ not reversible

5) Either give an example of a finite reversible Markov chain with period equal to 3, or show that that this cannot happen.

Period $= 3 \Rightarrow$ can divide states into 3 classes C_1, C_2, C_3 so that



There is some state $i \in C_1$, and $j \in C_2$

so that $P_{ij} > 0$

But $P_{ji} = 0$.

$\Rightarrow$ cannot be reversible.

6) Consider the following transition probability matrix for a Markov chain on 3 states:

$$\begin{array}{cc} P = \begin{pmatrix} 1/2 & 1/3 & 1/6 \\ 3/4 & 0 & 1/4 \\ 0 & 1 & 0 \end{pmatrix} & \begin{array}{c} 1 \\ 2 \\ 3 \end{array} \end{array}$$

Let X_n and Y_n be independent copies of this chain. Compute the long-run fraction of time when both chains X_n and Y_n are in the same state.

$$\mathbb{P}(X_n = Y_n) = \sum_i \mathbb{P}(X_n = Y_n = i)$$

$$= \sum_i \mathbb{P}(X_n = i)\, \mathbb{P}(Y_n = i)$$

Chain is irreducible, aperiodic $\Rightarrow$ unique stationary distribution w and

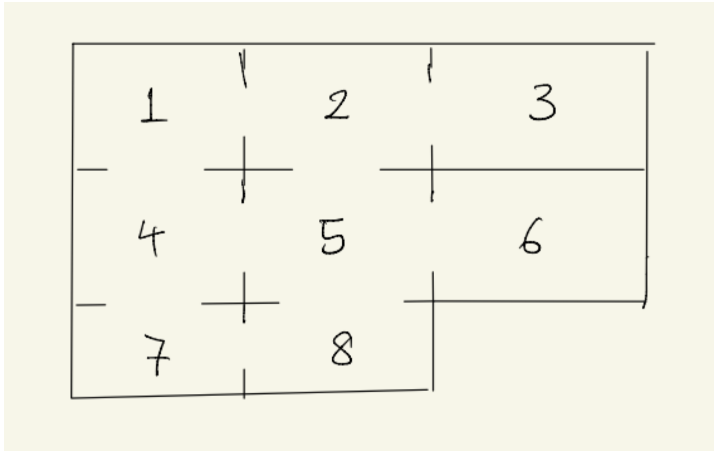
$$\lim_{n \to \infty} \mathbb{P}(X_n = i) = w_i$$

$$\Rightarrow \lim_{n \to \infty} \mathbb{P}(X_n = Y_n) = \sum_i w_i^2$$

Compute $w = \frac{1}{6}(3, 2, 1)$

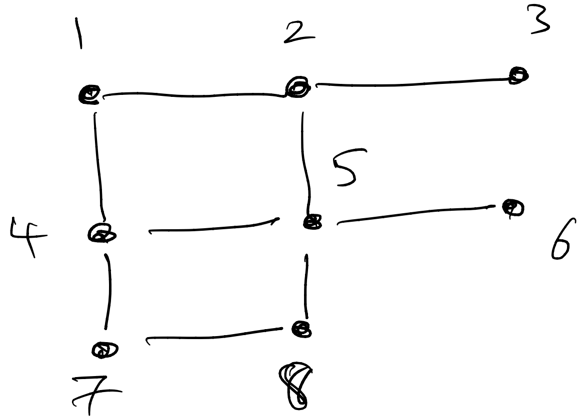
$$\Rightarrow \lim_{n \to \infty} \mathbb{P}(X_n = Y_n) = \frac{1}{36}(9 + 4 + 1) = \frac{7}{18}$$

7)



A rat moves through the 8 rooms in the maze shown above. At each step the rat either (a) with probability 1/2 stays in its current room, or (b) with probability 1/2 randomly selects a door and goes into the neighboring room. Let X_n denote the room where the rat is found after n steps.

a) Write down the transition matrix of the Markov chain $\{X_n\}$.



$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{4} & 0 & \frac{1}{4} & 0 & 0 & 0 & 0 \\ \frac{1}{6} & \frac{1}{2} & \frac{1}{6} & 0 & \frac{1}{6} & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{6} & 0 & 0 & \frac{1}{2} & \frac{1}{6} & 0 & \frac{1}{6} & 0 \\ 0 & \frac{1}{8} & 0 & \frac{1}{8} & \frac{1}{2} & \frac{1}{8} & 0 & \frac{1}{8} \\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{4} & 0 & 0 & \frac{1}{2} & \frac{1}{4} \\ 0 & 0 & 0 & 0 & \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{2} \end{pmatrix}$$

b) The rat starts in room 1. Find the mean number of steps until the rat first returns to room 1.

d_i = # doors in room i

$P_{ij} = \frac{1}{2}\frac{1}{d_i}$ if $i - j$

Reversible equations: $w_i P_{ij} = w_j P_{ji}$

$\Leftrightarrow w_i \frac{1}{2}\frac{1}{d_i} = w_j \frac{1}{2}\frac{1}{d_j}$

$\Rightarrow w_i = \dfrac{d_i}{\sum d_j} \Rightarrow w_1 = \dfrac{2}{2+3+1+3+4+1+2+2}$

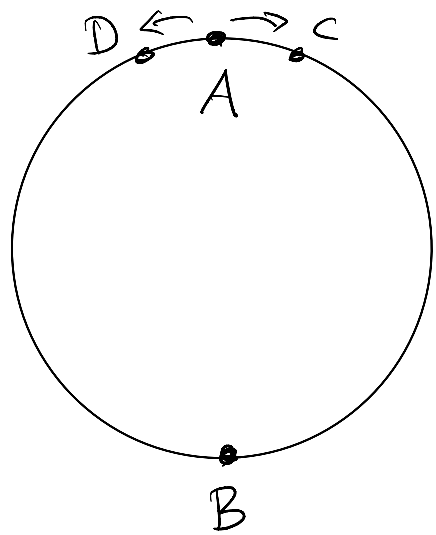
$$\Rightarrow w_1 = \frac{2}{18} = \frac{1}{9}$$

$\Rightarrow$ mean return ~~time~~ ~~to~~

room 1 $= 9$.

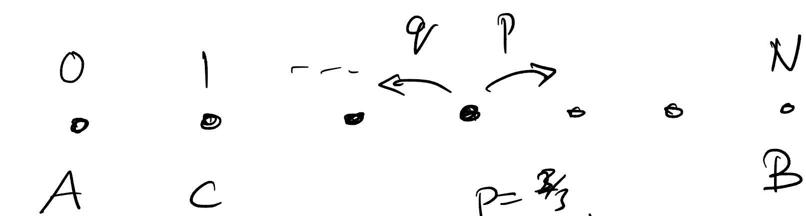
CHALLENGE: only attempt this if you are bored!!

8) A particle moves between $2N$ points which are equally spaced around a circle. At each step the particle either moves one point clockwise, with probability $2/3$, or moves one point counterclockwise, with probability $1/3$. Points A and B are N points apart on opposite sides of the circle. Starting at point A, find the probability that the particle visits point B before returning to A.



$$\mathbb{P}(\text{visit } B \mid X_0 = A) = \frac{2}{3}\,\mathbb{P}(\text{visit} \mid X_1 = C) + \frac{1}{3}\,\mathbb{P}(\text{visit } B \mid X_1 = D).$$

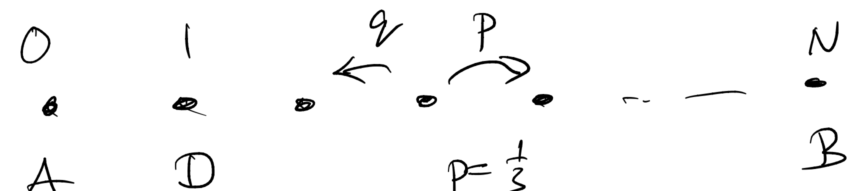
If $X_1 = C$:



$p = 2/3$
$q = \frac{1}{3}$

$$\mathbb{P}(\text{visit } B \text{ before } A \mid X_0 = C) = \frac{1 - q/p}{1 - (q/p)^N} = \frac{1 - \frac{1}{2}}{1 - \frac{1}{2^N}}$$

If $X_1 = D$:



$p = \frac{1}{3}$
$q = 2/3$

$$\Rightarrow \mathbb{P}(\text{visit } B \text{ before } A \mid X_0 = D) = \frac{1 - 2}{1 - 2^N} = \frac{1}{2^N - 1}$$

$\Rightarrow P(\text{visit } B \text{ before return to } A \mid X_0 = A)$

$$= \frac{2}{3} \frac{1 - \frac{1}{2}}{1 - \frac{1}{2^N}} + \frac{1}{3} \frac{1}{2^N - 1}$$

$$= \frac{1}{3} \frac{1}{1 - 2^{-N}} + \frac{1}{3} \frac{1}{2^N - 1}$$

$$= \frac{1}{3} \frac{2^N}{2^N - 1} + \frac{1}{3} \frac{1}{2^N - 1}$$

$$= \frac{1}{3} \frac{2^N + 1}{2^N - 1}$$

$$= \frac{1}{3} \frac{1 + 2^{-N}}{1 - 2^{-N}} \simeq \frac{1}{3} \quad \text{as } N \to \infty.$$

Dirichlet distribution

Define the probability simplex in $\mathbb{R}^K$: triangle when $K=3$

$$\Delta_K = \{x = (x_1, \ldots, x_k) \in \mathbb{R}^K \,|\, 0 \le x_i \le 1,\ \sum_{j=1}^{K} x_j = 1\}$$

Every point in Δ_K is a probability distribution for a set of size K. We will describe a type of probability distribution on Δ_K called a Dirichlet distribution. Let $\alpha = (\alpha_1, \ldots, \alpha_K) \in \mathbb{R}^K$ where $\alpha_i \ge 0$ for all $i = 1, \ldots, K$. The Dirichlet distribution defined by α is

$$\pi_\alpha(x) = N(\alpha) \prod_{j=1}^{K} x_j^{\alpha_j - 1}, \quad x \in \Delta_K$$

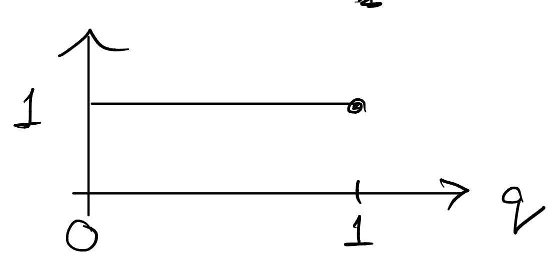
The number $N(\alpha)$ is the normalizing constant which makes this a pdf. We won't really need it, but the value is

$$N(\alpha) = \frac{\Gamma(\sum_{j=1}^{K} \alpha_j)}{\prod_{j=1}^{K} \Gamma(\alpha_j)}$$

where Γ is the gamma-function. The particular case $K = 2$ corresponds to coin tossing.

Recall: considered a coin with bias q, where we know nothing about $q := P(\text{Heads})$.
Our prior distribution for q is uniform on $[0,1]$.

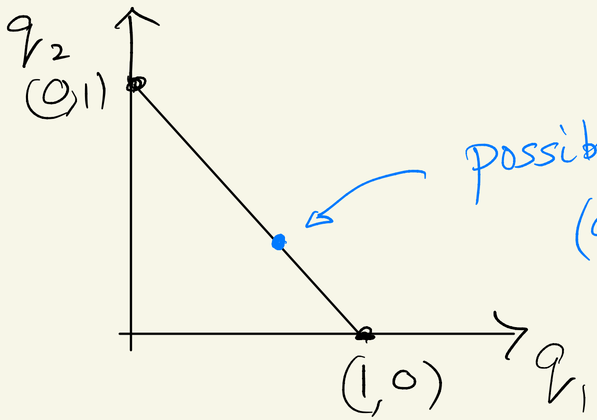
Prior: $$f_0(q) = \begin{cases} 1 & \text{for } 0 \le q \le 1 \\ 0 & \text{else} \end{cases}$$



$\mathbb{P}(H) = q, \quad \mathbb{P}(T) = 1-q$

$\mathbb{P}(H) = q_1 \qquad \mathbb{P}(T) = q_2$

where $q_1 + q_2 = 1$.



possible value for (q_1, q_2) where $q_1 = \mathbb{P}(H)$, $q_2 = \mathbb{P}(T)$.

Prior pdf f_0 is uniform over the diagonal line $q_1 + q_2 = 1$.

$$f_0(q_1, q_2) = \begin{cases} c & \text{for } q_1 + q_2 = 1, \\ & 0 \le q_1, q_2 \le 1 \\ 0 & \text{else.} \end{cases}$$

whee c is constant.

What is c?

$$\iint_{\{q_1+q_2=1\}} f_0(q_1,q_2)\, dq_1\, dq_2 = 1.$$

$$\Rightarrow \iint_{\{q_1+q_2=1\}} c \; dq_1\, dq_2$$

$$= c \iint_{\{q_1+q_2=1\}} dq_1\, dq_2$$

$$= c\sqrt{2}. \quad = 1$$

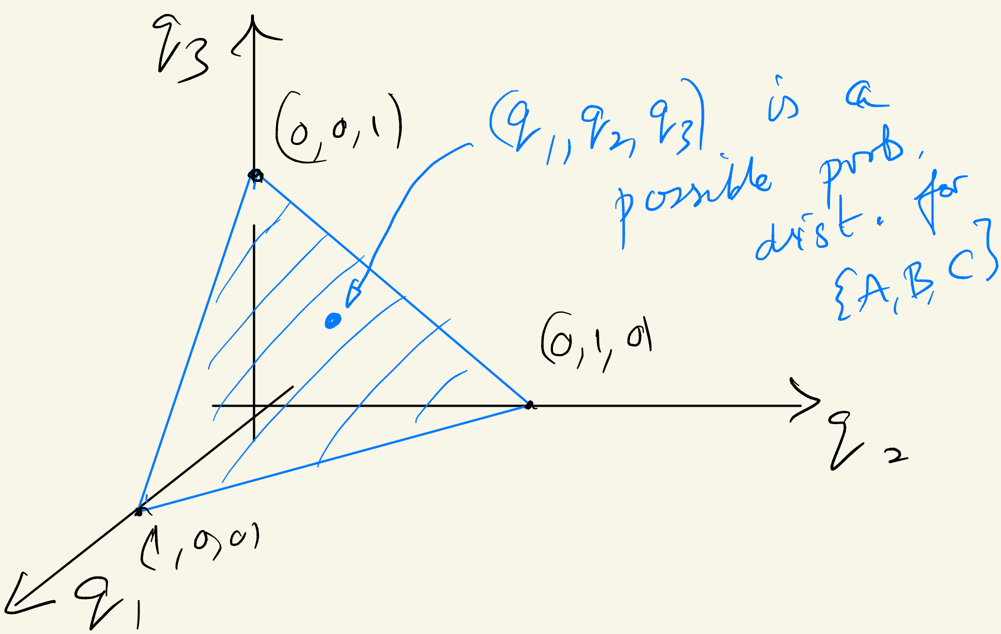
$$c = \frac{1}{\sqrt{2}}.$$

Extend this to a "block" with 3 sides $\{A, B, C\}$:

$$\left.\begin{array}{l} \mathbb{P}(A) = q_1 \\ \mathbb{P}(B) = q_2 \\ \mathbb{P}(C) = q_3 \end{array}\right\} \quad q_1 + q_2 + q_3 = 1.$$

(q_1, q_2, q_3)

$0 \leq q_i \leq 1.$



What is the uniform prior f_0 for (q_1, q_2, q_3)?

f_0 = constant on the shaded triangle.

$$f_0(q_1, q_2, q_3) = \begin{cases} c & \text{if } q_1 + q_2 + q_3 = 1 \\ & \underline{\text{and}} \quad 0 \leq q_i \leq 1 \\ 0 & \text{else.} \end{cases}$$

<u>Generalize this:</u>

K outcomes $\{1, 2, \dots, K\}$.

$$q_i = \mathbb{P}(X = i) \qquad i = 1, \dots, K.$$

$$\sum_{i=1}^{K} q_i = 1$$

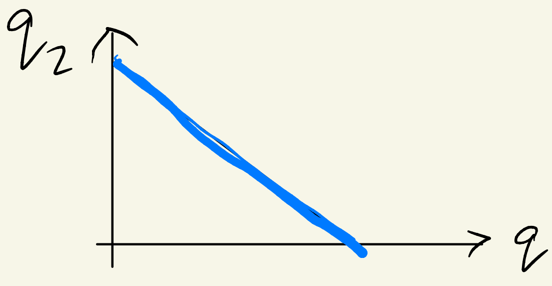
$q = (q_1, q_2, \dots, q_K)$ point in $\mathbb{R}^K$.

lies on hyperplane $\sum_{i=1}^{K} q_i = 1$.

Uniform prior for q is

$$f_0(q_1, q_2, \dots, q_K) = \begin{cases} c & \text{for } \sum_{i=1}^{K} q_i = 1 \\ & \underline{\text{and}} \quad 0 \le q_i \le 1 \\ 0 & \text{else.} \end{cases}$$

$K = 2$ $\qquad q = (q_1, q_2)$



$f_0(q) = $ uniform.

Toss coin n times, get k Heads.

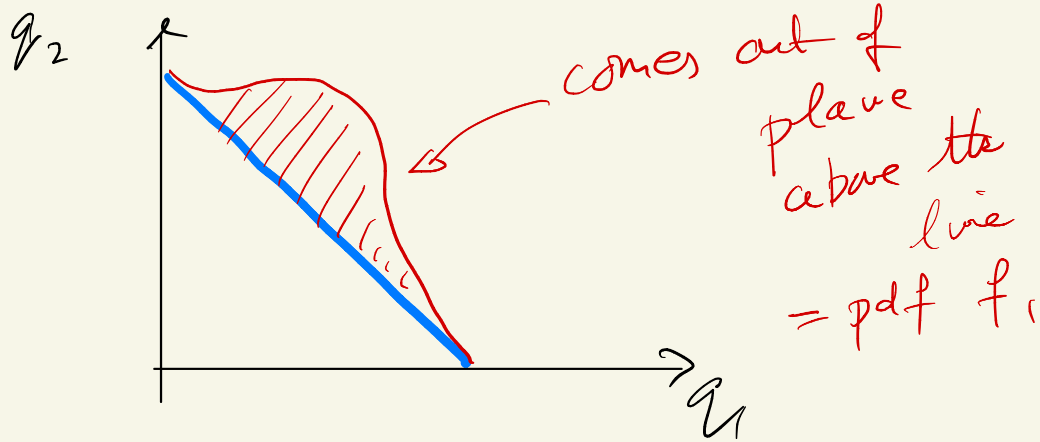
$$f_1(q \mid \mathcal{D}) = \frac{\mathbb{P}(\mathcal{D} \mid q)\, f_0(q)}{\mathbb{P}(\mathcal{D})}.$$

$$\mathbb{P}(\mathcal{D} \mid q_1, q_2) = \binom{n}{k} q_1^{k}\, q_2^{n-k}$$

<u>Posterior pdf:</u>

$$f_1(q \mid \mathcal{D}) = \begin{cases} N\, q_1^{k}\, q_2^{n-k} & \text{for } q_1 + q_2 = 1,\ 0 \le q_i \le 1 \\ 0 & \text{else.} \end{cases}$$

$q = (q_1, q_2)$



This is a Dirichlet distribution. It is a 2-parameter family of distributions:

$$\pi_{\alpha}(q_1, q_2) = N(\alpha)\, q_1^{\alpha_1 - 1}\, q_2^{\alpha_2 - 1}$$

normalization = number depending on α.

α_1 = real number

α_2 = real number

2 parameters of the pdf

The mean of the Dirichlet is

$$\mathbb{E}[x_i] = \int_{\Delta_K} x_i \, \pi_\alpha(x)\, dx = \frac{\alpha_i}{\sum_{i=1}^K \alpha_j}$$

The special case $\alpha = (1, 1, \ldots, 1)$ gives the uniform distribution on Δ_K. Also the product of two Dirichlet distributions is (up to normalization) again Dirichlet:

$$\begin{aligned} \pi_\alpha(x)\, \pi_\beta(x) &= N(\alpha)\, N(\beta) \prod_{j=1}^K x_j^{\alpha_j+\beta_j-2} \\ &\propto \pi_{\alpha+\beta-1} \end{aligned}$$

Mean values:

$$f_1(q.|\mathcal{D}) = \pi_\alpha(q) = N(\alpha)\, q_1^{\alpha_1-1} q_2^{\alpha_2-1} \cdots q_K^{\alpha_K-1}.$$

suppose posterior is this

Point estimates for q:

Mean: $$(q_i)_{mean} = \mathbb{E}_{\pi_\alpha}[q_i] = \frac{\alpha_i}{\sum_{j=1}^K \alpha_j}$$

MLE: $$(q_i)_{MLE} = \frac{\alpha_i - 1}{\sum_{j=1}^K (\alpha_j - 1)}$$

Returning to the multinomial we see that the likelihood function is a Dirichlet distribution. It is common practice to also use a Dirichlet distribution for the prior, so that

$$f_0(Q) = \pi_\alpha(Q)$$

for some $\alpha = (\alpha_1, \ldots, \alpha_K)$. In this case the posterior is again a Dirichlet:

$$\begin{aligned} f_1(Q|\mathcal{D}) &\propto \prod_{i=1}^{n}\prod_{j=1}^{K} q_j^{Y_{ij}} \prod_{j=1}^{K} q_j^{\alpha_j - 1} \\ &\propto \prod_{j=1}^{K} q_j^{\sum_{i=1}^{n} Y_{ij} + \alpha_j - 1} \\ &\sim \text{Dirichlet}\left(\alpha_1 + \sum_{i=1}^{n} Y_{i1}, \ldots, \alpha_K + \sum_{i=1}^{n} Y_{iK}\right) \end{aligned}$$

So for example we immediately get the posterior means:

$$\mathbb{E}[q_j|\mathcal{D}] = \frac{\alpha_j + \sum_{i=1}^{n} Y_{ij}}{\sum_{l=1}^{K} \alpha_l + n}$$

Example: Suppose $K = 3$, so we have a multinomial distribution for three types, with probabilities $q = (q_1, q_2, q_3)$, where $q_1 + q_2 + q_3 = 1$. Assume that the prior distribution is uniform on the simplex, so this is Dirichlet with $\alpha = (1, 1, 1)$. The pdf is

$$f_0(q) = \begin{cases} 2 & \text{for } q_1 + q_2 + q_3 = 1 \\ 0 & \text{else} \end{cases}$$

→ Dirichlet with $\alpha = (1, 1, 1)$.

Suppose we make 10 measurements and get the results

$$\mathcal{D} = (X_1, \ldots, X_{10}) = (1, 3, 3, 2, 3, 1, 1, 2, 1, 3).$$

So we have for example

$$Y_{11} = 1, Y_{12} = 0, Y_{13} = 0, Y_{21} = 0, Y_{22} = 0, Y_{23} = 1, \ldots$$

Then the posterior is again a Dirichlet distribution with parameters

$$\alpha' = \left(1 + \underbrace{\sum_{i=1}^{10}}_{=4} Y_{i1}, 1 + \underbrace{\sum_{i=1}^{10}}_{=2} Y_{i2}, 1 + \underbrace{\sum_{i=1}^{10}}_{=4} Y_{i2} \right) = (5, 3, 5)$$

So the posterior pdf is

$$f_1(q \,|\, \mathcal{D}) = N\, q_1^{\alpha_1' - 1}\, q_2^{\alpha_2' - 1}\, q_3^{\alpha_3' - 1} = 8316\, q_1^4\, q_2^2\, q_3^4$$

→ Dirichlet with $\alpha = (5, 3, 5)$

We can calculate the point estimates: the mean is

$$q_{1,mean} = \frac{5}{13}, \quad q_{2,mean} = \frac{3}{13}, \quad q_{3,mean} = \frac{5}{13}$$

and the MLE is

$$q_{1,MLE} = \frac{4}{10}, \quad q_{2,MLE} = \frac{2}{10}, \quad q_{3,MLE} = \frac{4}{10}$$

MTH 7241 Fall 2022: Prof. C. King

Notes 8: Markov Chain Monte Carlo

0.1 The basic idea

Persi Diaconis titled his review article 'The Markov Chain Monte Carlo Revolution', referring to the fact that MCMC techniques have become pervasive and essential in most aspects of Data Analysis and Applied Probability. By using powerful computing resources the MCMC technique allows answers to be found for highly complicated problems involving statistical optimization.

We will describe the basic MCMC algorithm for discrete Markov chains on finite state spaces, and provide several concrete examples. Let Ω be a finite state space, and let π be a probability distribution on Ω. The goal is to generate random samples from π, meaning to generate elements of Ω which are chosen randomly with probability distribution π. The MCMC method proceeds by constructing an ergodic Markov chain on Ω for which π is the stationary distribution. By running the Markov chain we generate a sequence of states $X_1, X_2, \dots$ whose distribution is guaranteed to converge to π. So for any $\epsilon > 0$, there is N sufficiently large such that the sequence $X_N, X_{N+1}, \dots$ has distribution close to π, in the sense that

$$\sum_{j \in \Omega} |\mathbb{P}(X_{N+k} = j) - \pi(j)| \leq \epsilon$$

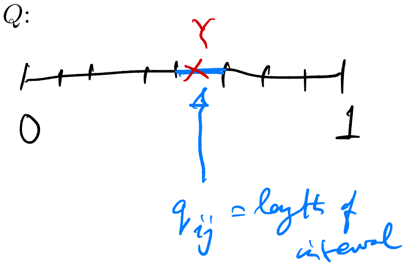
for all $k \geq 0$.

0.2 The basic algorithm

We assume for definiteness that $\pi(j) > 0$ for all $j \in \Omega$. We also assume that $Q = (q_{ij})$ is a known symmetric irreducible aperiodic transition matrix on Ω. Being symmetric, the stationary distribution of Q is uniform on Ω. The MCMC algorithm will 'transform' Q into a new Markov chain with the desired stationary distribution π. The algorithm constructs a sequence $X_0, X_1, X_2, \ldots$ where X_0 is any chosen initial state, and X_{k+1} is constructed from X_k using the following steps:

1) randomly select $Y \in \Omega$ according to the probability distribution Q:

$$\mathbb{P}(Y = j \mid X_k = i) = q_{ij}$$



2) define the acceptance probability

$$\begin{aligned} \alpha &= \min\left\{1, \frac{\pi(Y)}{\pi(X_k)}\right\} \\ &= \min\left\{1, \frac{\pi(j)}{\pi(i)}\right\} \quad \text{if } Y = j,\ X_k = i \end{aligned}$$

3) define

$$X_{k+1} = \begin{cases} Y & \text{with probability } \alpha \\ \\ X_k & \text{with probability } 1 - \alpha \end{cases}$$

(more concretely, we flip a weighted coin with probability of Heads equal to α; if it comes up Heads we set $X_{k+1} = Y$, if Tails we set $X_{k+1} = X_k$).

This protocol defines the off-diagonal entries of the transition matrix $P = (p_{ij})$ for the new Markov chain, namely

$$p_{ij} = \mathbb{P}(X_{k+1} = j \mid X_k = i) = q_{ij} \, \min\left\{1, \frac{\pi(j)}{\pi(i)}\right\} \quad \text{for } i \neq j$$

We then define the diagonal entries in order to make the matrix stochastic:

$$p_{ii} = 1 - \sum_{j \neq i} p_{ij}$$

Theorem 1 *For all $i, j \in \Omega$,*

$$\pi(i) p_{ij} = \pi(j) p_{ji}$$

Proof: It is sufficient to assume that $i \neq j$, in which case

$$\begin{aligned}
\pi(i) p_{ij} &= \pi(i) q_{ij} \, \min\left\{1, \frac{\pi(j)}{\pi(i)}\right\} \\
&= q_{ij} \min\left\{\pi(i), \pi(j)\right\} \\
&= q_{ji} \pi(j) \, \min\left\{1, \frac{\pi(i)}{\pi(j)}\right\} \quad \text{(since } Q \text{ is symmetric)} \\
&= \pi(j) p_{ji}
\end{aligned}$$

$\boxed{\text{QED}}$

Remarks: Theorem 1 shows that the new Markov chain is reversible, and therefore it follows that π is its stationary distribution. So $\{X_k\}$ is the promised Markov chain that converges to π. Note also that the computation of p_{ij} requires only knowledge of the ratio $\pi(j)/\pi(i)$. We will see that there are many applications where these ratios can be easily computed, while the probabilities $\pi(i)$ are themselves difficult to find, because the overall normalization is hard to compute. One important case is Bayesian inference.

Example 1: sampling from a multinomial distribution

Suppose that X is discrete with $\mathrm{Ran}(X) = \{1, 2, \ldots, K\}$, and pdf

$$\mathbb{P}(X = k) = q_k, \qquad k = 1, \ldots, K.$$

We will use MCMC to estimate the expected value $\mathbb{E}[X^2]$. So our stationary pdf for the Markov chain is $\pi = (q_1, \ldots, q_K)$. We choose the reference transition to be uniform on the state space, so

$$q_{ij} = \frac{1}{K} \qquad \text{for all } i, j = 1, \ldots, K$$

Then the transition probability for MCMC is

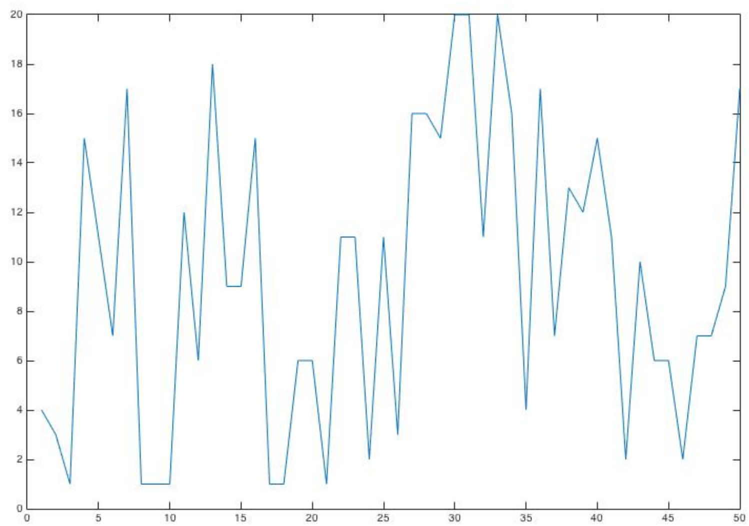
$$p_{ij} = \begin{cases} \frac{1}{K} \min\left(1, \frac{q_j}{q_i}\right) & \text{for all } i \neq j \\ 1 - \sum_{k \neq i} p_{ik} & \text{for } j = i \end{cases}$$

So starting with an initial state X_0, we simulate a sequence $X_1, X_2, \ldots$ using this transition matrix.

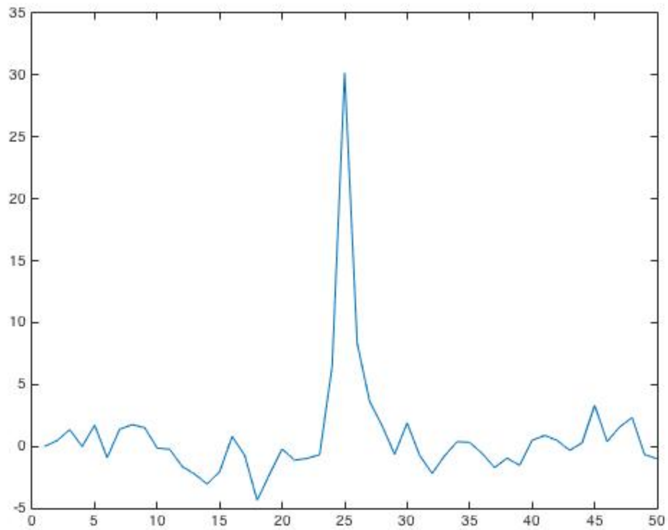
We then estimate the mixing time of the chain – this is the number of steps s such that X_n and X_{n+s} are approximately independent. Then we estimate the expected value as

$$\mathbb{E}[X^2] \simeq \frac{1}{N}\sum_{n=1}^{N} X_{ns}^2$$

Let's see a concrete example. Take $K = 20$, and generate a random distribution $(q_1, \ldots, q_{20})$. Choose initial state $X_0 = 4$, and generate a run $X_1, \ldots, X_{50}$; see Fig 1 for an example.



Now we need to estimate s, the mixing time. Generate 500 independent runs of length 50, and compute the covariance between X_n and X_{n+s} for different values of s. See Fig 2 for a graph of covariance of X_{25} with X_n for $n = 1, \ldots, 50$.



We see that the covariance becomes small for $\mathrm{COV}(X_{25}, X_{30})$, so we take $s = 5$. Now we generate a long run of length 1000, $X_1, \ldots, X_{1000}$. Then our estimate for the expected value is

$$\mathbb{E}[X^2] \simeq \frac{1}{200} \sum_{n=1}^{200} X_{5n}^2$$

A sample run with $K = 20$, random pdf for X, use MCMC to generate $X_1, \ldots, X_{1000}$ gives

$$\begin{aligned} \mathbb{E}[X^2] &\simeq 158.3 \quad \text{(MCMC approximation)} \\ \mathbb{E}[X^2] &= 154.8 \quad \text{(exact)} \end{aligned}$$