•If $\{X_n\}$ is an irreducible chain with period d, then $\{Y_n = X_{dn}\}$ is an aperiodic chain and it then follows that

$$P_{ii}^{nd} = \mathbb{P}(X_{nd} = i | X_0 = i) \to \frac{d}{m_i}, \quad \text{as} \quad n \to \infty$$

Simple Walk with Jumps at End-points

Consider a Markov Chain with transition matrix

$$P = \begin{bmatrix} t_1 & \frac{1}{2}+t_2 & t_3 & t_4 & \cdots & t_{n-1} & t_n \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & \cdots & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{1}{2} & 0 & \frac{1}{2} \\ s_1 & s_2 & s_3 & \cdots & s_{n-2} & \frac{1}{2}+s_{n-1} & s_n \end{bmatrix}$$

where

$$\sum_{i=1}^{n} t_i = \sum_{i=1}^{n} s_i = \frac{1}{2}, \qquad t_i, s_i \geq 0$$

•One can find the stationary distribution explicitly based on probabilistic interpretation of Green's function.

•Can you tell that all eigenvalues of P are **real**? For related work in the Brownian motion setting, see the paper "Spectral analysis of Brownian motion with jump boundary" by Leung, Li and Rakesh (2008), *Proceedings of the American Mathematical Society*, **136**, 4427-4436.

Project II: Due April 7, 2009

Find the stationary distribution explicitly in each of the following cases.

(a). $t_m = s_m = 1/2$, for fixed $1 < m < n$. The answer is a Green's function.

(b). $t_m = s_k = 1/2$, for fixed $1 < m, k < n$. The answer is a two-point mixed Green's function.

(c). The general case with t_i and s_i. The answer is a "mixed" Green's function.

One Page Introduction to Green's Function

•From Potential point of view, the Green's function $G^D(x,y)$ should be considered as one-half the potential at the point x due to a unit positive electrical charge placed at y inside the grounded boundary ∂D.

•From PDE point of view, the Dirichlet Green's function $G^D(x,y)$ is the solution of

$$\frac{1}{2}\Delta_x G^D(x,y) = -\delta(x-y), \qquad x \in D \quad G^D(x,y) = 0, \qquad x \in \partial D.$$

•From the probabilistic point of view, $G^D(x,y)$ is the expected time that the process starting at x spends in y before hitting ∂D.

PROJECTS

Try your best to work on at least FOUR projects assigned during the semester. It is a good idea to find and hand-in related papers, Web printouts, simulations, etc. Partial results and special cases are also important.

•Several problems require you to carry out the entire modelling process: Setting up the model, establishing the parameters for it, doing some calculations, and finally, drawing conclusions. It will require you to think and give you a chance to practice your modelling and analysis skills.

C3.24. A coin, having probability p of landing heads and probability $q = 1 - p$ of landing tails, is continually flipped until at least one head and one tail have been flipped.
(a). Find the expected number of flips needed.
(b). Find the expected number of flips that lands on heads.
(c). Find the expected number of flips that lands on tails.
(d). Repeat part (a) in the case where flipping is continued until a total of at least two heads and one tail have been flipped.

Sol: Denote by N the number of flips asked.
(a). Conditioning on the first come and then the second, we have

$$\begin{aligned} \mathbb{E}\, N &= \mathbb{E}\,(N|H) \cdot p + \mathbb{E}\,(N|T) \cdot q \\ \mathbb{E}\,(N|H) &= \mathbb{E}\,(N|HH) \cdot p + \mathbb{E}\,(N|TH) \cdot q = (1 + \mathbb{E}\,(N|H)) \cdot p + 2q \\ \mathbb{E}\,(N|T) &= \ldots \end{aligned}$$

which implies $\mathbb{E}\, N = 1 + \frac{p}{q} + \frac{q}{p} = \frac{1}{pq} - 1 = \frac{1}{p} + \frac{1}{q} - 1$.
(b). Note that $\mathbb{E}\, N$ in this case is a conditional expectation, namely, the expected number of heads given we stopped with at least one head and one tail. Conditioning again,

$$\begin{aligned} \mathbb{E}\, N &= \mathbb{E}\,(N|H) \cdot p + \mathbb{E}\,(N|T) \cdot q = \mathbb{E}\,(N|H) \cdot p + q \\ \mathbb{E}\,(N|H) &= \mathbb{E}\,(N|HH) \cdot p + \mathbb{E}\,(N|TH) \cdot q = (1 + \mathbb{E}\,(N|H)) \cdot p + q \end{aligned}$$

which implies $\mathbb{E}\, N = q + \frac{p}{q}$.

(c). Interchanging p and q in (b) gives result: $\mathbb{E}\, N = p + \frac{q}{p}$.

(d). Conditioning, we have $\mathbb{E}\, N = \mathbb{E}\,(N|H) \cdot p + \mathbb{E}\,(N|T) \cdot q$. From part (a), we know $\mathbb{E}\,(N|H) = 1 + (\frac{1}{p} + \frac{1}{q} - 1)$. We can also find it by conditioning and using mean formula for geometric r.v. That is,

$$\mathbb{E}\,(N|H) = \mathbb{E}\,(N|HH) \cdot p + \mathbb{E}\,(N|TH) \cdot q = (2 + \frac{1}{q}) \cdot p + (2 + \frac{1}{p}) \cdot q.$$

Similarly, $\mathbb{E}\,(N|T) = 1 + \frac{2}{p}$ via means of negative binomial r.v, or by conditioning,

$$\mathbb{E}\,(N|T) = \mathbb{E}\,(N|TH) \cdot p + \mathbb{E}\,(N|TT) \cdot q = (2 + \frac{1}{p})p + (1 + \mathbb{E}\,(N|T)) \cdot q.$$

The final answer is $\mathbb{E}\, N = \frac{2}{p} + \frac{1}{q} - 2 + q$.

Project I: Due April 7, 2009

A coin, having probability p of landing heads and probability $q = 1-p$ of landing tails, is continually flipped until at least m head and n tail have been flipped. Find the expected number of flips needed.

C3.26. Sol: Let N_A and N_B denote the number of games needed given that you start with A and given that you start with B, respectively. Conditioning on the outcome of the first game gives $\mathbb{E}\,N_A = \mathbb{E}\,(N_A|W_A)\cdot p_A + \mathbb{E}\,(N_A|L_A)\cdot(1-p_A)$ Conditioning on the outcome of the next game gives

$$\begin{aligned}\mathbb{E}\,(N_A|W_A) &= \mathbb{E}\,(N_A|W_AW_B)\cdot p_B + \mathbb{E}\,(N_A|W_AL_B)\cdot(1-p_B)\\ &= 2p_B + (2+\mathbb{E}\,N_A)(1-p_B).\end{aligned}$$

As $\mathbb{E}\,(N_A|L_A) = 1+\mathbb{E}\,N_B$, we obtain that

$$\begin{aligned}\mathbb{E}\,N_A &= (2+(1-p_B)\mathbb{E}\,N_A)\cdot p_A + (1+\mathbb{E}\,N_B)\cdot(1-p_A)\\ &= 1+p_A+p_A(1-p_B)\mathbb{E}\,N_A + (1-p_A)\mathbb{E}\,N_B.\end{aligned}$$

Similarly,

$$\mathbb{E}\,N_B = 1+p_B+p_B(1-p_A)\mathbb{E}\,N_B + (1-p_B)\mathbb{E}\,N_A.$$

Subtracting gives

$$\mathbb{E}\,N_A - \mathbb{E}\,N_B = p_A - p_B + (p_A-1)(1-p_B)\mathbb{E}\,N_A + (1-p_B)(1-p_A)\mathbb{E}\,N_B$$

or $(1+(1-p_A)(1-p_B))(\ EN_A - \mathbb{E}\,N_B) = p_A - p_B$ Hence, if $p_B > p_A$ then $\mathbb{E}\,N_A - \mathbb{E}\,N_B < 0$, showing that playing A first is better.

C3.27. Sol: Condition on the outcome of the first flip to obtain

$$\mathbb{E}\,X = \mathbb{E}\,(X|H) \cdot p + \mathbb{E}\,(X|T) \cdot (1-p) = (1 + \mathbb{E}\,X)p + \mathbb{E}\,(X|T)(1-p)$$

Conditioning on the next flip gives

$$\mathbb{E}\,(X|T) = \mathbb{E}\,(X|TH)p + \mathbb{E}\,(X|TT)(1-p) = (2+\mathbb{E}\,X)p + (2+1/p)(1-p)$$

where the final equality follows since given that the first two flips are tails the number of additional flips is just the number of flips needed to obtain a head. Putting the preceding together yields

$$\mathbb{E}\,X = (1 + \mathbb{E}\,X)p + (2 + \mathbb{E}\,X)p(1-p) + (2 + 1/p)(1-p)^2$$

or $\mathbb{E}\,[X] = \frac{1}{p(1-p)^2}$.

Finite-state Irreducible Chains

We follow the notes of Aldous and Fill. Let

$$T_i = \min\{n \geq 0 : X_n = i\}$$

be the first hitting time on state i, and write

$$T_i^+ = \min\{n \geq 1 : X_n = i\}.$$

Of course $T_i^+ = T_i$ unless $X_0 = i$, in which case T_i^+ is called the first return time to state i. More generally, a subset A of states has first hitting time

$$T_A = \min\{n \geq 0 : X_n \in A\}.$$

Identities for mean hitting times and occupation times

Key Identity: Consider the chain stated at stat i. Let $0 < T < \infty$ be a stopping time such that $X_T = i$ and $\mathbb{E}_i T < \infty$. Let j be an arbitrary state. Then

$$\mathbb{E}_i(\text{\# of visits to } j \text{ before time } T) = \pi_j \mathbb{E}_i T.$$

In the phrase "number of ... before time t", the convention is to include time 0 but exclude time t.

•We can lift point measure to general measure and obtain an occupation measure identity: Let μ be a probability distribution on S. Let $0 < T < \infty$ be a stopping time such that $\mathbb{P}_\mu(X_T \in \cdot) = \mu(\cdot)$ and $\mathbb{E}_\mu T < \infty$. Let j be an arbitrary state. Then

$$\mathbb{E}_\mu(\text{\# of visits to } j \text{ before time } T) = \pi_j \mathbb{E}_\mu T.$$

Mean hitting time and related formulas

Define the fundamental matrix $Z = (Z_{ij})$ where

$$Z_{ij} = \sum_{n=0}^{\infty} \left(P_{ij}^{(n)} - \pi_j \right).$$

Note that knowing Z is equivalent as knowing the mean hitting time matrix $(\mathbb{E}_i T_j)$ since $Z_{ij} = \pi_j(\mathbb{E}_\pi T_j - \mathbb{E}_i T_j)$.

•For all $i \in S$, $\sum_{j\in S} Z_{ij} = 0$.

•The Z is a "generalized inverse" of $\mathbf{I} - \mathbf{P}$ in the sense that

$$(\mathbf{I} - \mathbf{P})\mathbf{Z} = \mathbf{Z}(\mathbf{I} - \mathbf{P}) = \mathbf{I} - \mathbf{\Pi}$$

where $\mathbf{\Pi}$ is the matrix with (i,j)-entry π_j.

•For any i, the first return time $m_i = \mathbb{E}_i T_i^+ = 1/\pi_i$.

•For any i,

$$\mathbb{E}_i N_j(T_i^+) = \mathbb{E}_i(\#\text{ of visits to } j \text{ before time } T_i^+) = \pi_j/\pi_i.$$

•For any $j \neq i$,

$$\mathbb{E}_i N_i(T_j) = \mathbb{E}_i(\text{\# of visits to } i \text{ before time } T_j) = \pi_i(\mathbb{E}_i T_j + \mathbb{E}_j T_i).$$

•For any $j \neq i$,

$$\mathbb{P}_i(T_j < T_i^+) = \frac{1}{\pi_i(\mathbb{E}_i T_j + \mathbb{E}_j T_i)}.$$

•For any $l \neq i$ and arbitrary j,

$$\mathbb{E}_i(\text{\# of visits to } j \text{ before time } T_l) = \pi_j(\mathbb{E}_i T_l + \mathbb{E}_l T_j - \mathbb{E}_i T_j).$$

•For any $l \neq i$ and $l \neq j$,

$$\mathbb{P}_i(T_j < T_l) = \frac{\mathbb{E}_i T_l + \mathbb{E}_l T_j - \mathbb{E}_i T_j}{\mathbb{E}_j T_l + \mathbb{E}_l T_j}.$$

•$Z_{ii} = \pi_i \mathbb{E}_\pi T_i$.

•$\pi_j \mathbb{E}_i T_j = Z_{jj} - Z_{ij}$.

•For any i, $\sum_j \pi_j \mathbb{E}_i T_j = \sum_j Z_{jj}$ which is called the random target lemma.

•For any $i, j \in S$,

$$\mathbb{E}_\pi(\# \text{ of visits to } j \text{ before time } T_i) = \frac{\pi_j}{\pi_i} Z_{ii} - Z_{ij}.$$

•It maybe better to use the notation

$$N_j(T) = \# \text{ of visits to } j \text{ before time } T$$

Project III: More Identities

Find expression for (a) $\mathbb{E}_i \min(T_k, T_l)$;
(b) $\mathbb{E}_i$ (# of visits to j before time $\min(T_k, T_l)$);
(c) $\mathbb{P}_i$(hit j before time $\min(T_k, T_l)$).

Open Problem: Portmanteau theorem for occupation times. See page 31 of Aldous-Fill. Can the results listed be formulated as a single theorem?

To explain the goal by analogy, consider the use of Feynman diagrams to calculate quantities such as $\mathbb{E}\, X_1^3 X_2 X_3^2$ for dependent mean-zero Gaussian (X_1, X_2, X_3), ...

Key point: We seek a general rule which associates an expression like

$$\mathbb{E}_i(\text{\# of visits to } j \text{ before time } \min(T_k, T_l))$$

with a combinatorial structure involving indexes $\{i, j, k, l\}$; then associates with the combinatorial structure some function of variables $\{p_v, z_{vw}, v, w \in \{i, j, k, l\}\}$; then shows that the value of the expression applied to a finite Markov chain equals the function of $\{pi_v, Z_{vw}, v, w \in \{i, j, k, l\}\}$.

Project IV : Time Saved in Table Tennis

The old rule for table tennis (ping-pong) game is best of five. The game is over when the winner wins three sets. For each set, the first to get 21 points wins the set if the other scores no more than 19 points. If the score is tied at 20 or beyond, then the winner of the set needs to score two consecutive points to have a set score such as 24 : 22. The new rule is best of seven and the set score is 11 points. It is claimed that the main reason for the rule change is the reduction of the total game time. Assume two players are equally matched, i.e. each point in a set is scored equally likely. What is the percentage of time saved under the new rule on the average?

- Stochastic modelling and analysis
- Stochastic simulation

Time Reversible Markov Chain

Given a stationary Markov chain $X_1, \cdots, X_n, X_{n+1}, \cdots$ with

$$P_{ij} = \mathbb{P}(X_{n+1} = j | X_n = i), \quad \mathbb{P}(X_n = i) = \pi_i$$

then the reversed process $X_n, X_{n-1}, \cdots$ is also a Markov chain with one-step transition probabilities

$$Q_{ij} = \mathbb{P}(X_m = j | X_{m+1} = i) = P_{ji} \cdot \frac{\pi_j}{\pi_i}$$

Def: An ergodic Markov chain is *time reversible* if $Q_{ij} = P_{ij}$ for all i, j, i.e.

$$\pi_i P_{ij} = \pi_j P_{ji}$$

- The rate from i to j is

$$\lim_{n \to \infty} \mathbb{P}(X_{n+1} = j, X_n = i) = \pi_i P_{ij}.$$

Thus for time reversible Markov chain, the rate from i to j = the rate from j to i.

• The long run (expected) proportion of one-step transitions from i to j is

$$\lim_{n\to\infty} \mathbb{E} \frac{1}{n} \sum_{k=1}^{n} 1_{\{X_{n+1}=j, X_n=i\}} = \pi_i P_{ij}.$$

Thus for time reversible Markov chain, they are the same for i to j and j to i.

• If there exist $x_i \geq 0$, $\sum x_i = 1$, such that

$$x_i P_{ij} = x_j P_{ji}, \quad \forall i, j,$$

then the Markov chain is time reversible and $x_i = \pi_i$. This provides a way to find π_i.

• **Kolmogorov's criterion for reversibility.** An ergodic Markov chain for which $P_{ij} = 0$ whenever $P_{ji} = 0$ is time reversible if and only if starting in state i, any path back to i has the same probability as the reversed path. That is, if

$$P_{i,i_1} P_{i_1,i_2} \cdots P_{i_k,i} = P_{i,i_k} P_{i_k,i_{k-1}} \cdots P_{i_1,i}$$

for all states $i, i_1, \cdots, i_k$.

• For an irreducible Markov chain with transition probabilities P_{ij}, if there exist $x_i \geq 0$, $\sum x_i = 1$, and a transition probability matrix $\mathbf{Q} = [Q_{ij}]$ such that for all i, j,

$$x_i P_{ij} = x_j Q_{ji},$$

then the Q_{ij} are the transition probabilities for the reversed chain and the π_i are the stationary probabilities both for the original and reversed chain. This also provides a way to find π_i.

Ex: King's Problem.

Ex: Random Walk on a Connected Graph with Edge Weights.

Ex: A List Model: Front of the List Rule vs One-closer Rule.

Ex: The Ehrenfest Urn Model.

Perron-Frobenius Theorem: If $\mathbf{P}$ is the transition matrix of a finite irreducible chain with period d, then
(a) $\lambda_1 = 1$ is an eigenvalue of $\mathbf{P}$;
(b) The d complex of unity $\lambda_j = e^{(j-1)2\pi \mathtt{i}/d}$, $1 \leq j \leq d$, are eigenvalue of $\mathbf{P}$;
(c) the reminding eigenvalues λ_j, $d+1 \leq j \leq |S|$, satisfy $|\lambda_j| < 1$.

•If the eigenvalues $\lambda_1, \cdots, \lambda_N$ of $\mathbf{P}$ are distinct then there exists a matrix L such that $\mathbf{P} = L^{-1}\Lambda L$ where Λ is the diagonal matrix with entries $\lambda_1, \cdots, \lambda_N$. Thus

$$\mathbf{P}^n = L^{-1}\Lambda^n L = L^{-1} diag(\lambda_1^n, \cdots, \lambda_N^n) L$$

The rows of L are left eigenvectors of $\mathbf{P}$. We can use Perron-Frobenius theorem to obtain properties of $\mathbf{P}^n$ for large n.

•If $d = 1$, then

$$\mathbf{P}^n \to L^{-1}\text{diag}(1, 0, \cdots, 0)L = \text{col}(\pi_1, \cdots, \pi_{|S|})$$

•When the eigenvalues of $\mathbf{P}$ are not distinct, then $\mathbf{P}$ can not always be reduced to the diagonal form in this way. The best that one can do is to rewrite $\mathbf{P}$ in its "Jordan canonical form".

•In typical linear algebra, we have $AV = V\Lambda$ where V are formed by eigenvectors, the standard one from right. Thus we have $L = V^{-1}$.

•Second eigenvalue and speed of convergence.

Renyi's Connection between Information and Reversible Markov Chains

For any two finite measures $\mu = (\mu_1, \mu_2, \cdots, \mu_N)$ and $\nu = (\nu_1, \nu_2, \cdots, \nu_N)$, the *relative entropy*, also called *Kullback-Leibler divergence* or just *divergence* of μ and ν is

$$I(\mu, \nu) = \sum_{i=1}^{N} \mu_i \log_2(\mu_i/\nu_i).$$

• q_{ij} is the "long-run Bayes rule" probability that "you are in state i, given that on the next step you go to j". Note that by the stationarity of π that for all j that $Q = (q_{ij})$ has column sums equal to one, i.e. $\sum_{i \in S} q_{ij} = 1$.

Entropy Characterization: If $I(\mu, \nu) = 0$, then $\mu = \nu$. This can be easily seen by using the condition for equality in Jensen's inequality.

Averaging Steps: Note that

$$p_{ij}^{(n+1)} = \sum_{k \in S} p_{ik}^{(n)} \cdot p_{kj} = \sum_{k \in S} p_{ik}^{(n)} \cdot \pi_j q_{kj}/\pi_k$$

and hence we have a central *averaging* relation:

$$p_{ij}^{(n+1)}/\pi_j = \sum_{k \in S} q_{kj} \cdot (p_{ik}^{(n)}/\pi_k).$$

Decreasing Entropy: We write $\mu_i^{(n)}$ for the probability distribution (measure) of the chain that starts in state i and takes n steps, i.e. for any $A \subset S$,

$$\mu_i^{(n)}(A) = \mathbb{P}(X_n \in A \,|\, X_0 = i).$$

Then for all $i \in S$, by Jensen's inequality,

$$\begin{aligned}
I(\mu_i^{(n+1)}|\pi) &= \sum_{j\in S} p_{ij}^{(n+1)} \log_2(p_{ij}^{(n+1)}/\pi_j) \\
&= \sum_{j\in S} \pi_j \left(\sum_{k\in S} (p_{ik}^{(n)}/\pi_k) q_{kj}\right) \log_2 \left(\sum_{k\in S} (p_{ik}^{(n)}/\pi_k) q_{kj}\right) \\
&\le \sum_{j\in S} \pi_j \sum_{k\in S} \Big((p_{ik}^{(n)}/\pi_k) \log_2(p_{ik}^{(n)}/\pi_k)\Big)\, q_{kj} \\
&= \sum_{j\in S} \pi_j \sum_{k\in S} \Big((p_{ik}^{(n)}/\pi_k) \log_2(p_{ik}^{(n)}/\pi_k)\Big)\, (p_{kj}/\pi_j)\cdot \pi_k \\
&= I(\mu_i^{(n)}|\pi).
\end{aligned}$$

The Limits: For each $i \in S$, the limit

$$\lim_{n\to\infty} I(\mu_i^{(n)}|\pi) = L_i = 0, \quad \text{and} \quad \lim_{n\to\infty} p_{ij}^{(n)} = \pi_j.$$

The proof is based on a very useful "subsequence and characterization" method.

Subsequences: Due to the finiteness of the state spaces, we can find α_{ij} and a subsequence $n_1 < n_2 < \cdots$ such that for all i and j,

$$\lim_{k\to\infty} p_{ij}^{(n_k)} = \alpha_{ij}.$$

Note that $\sum_{j\in S} \alpha_{ij} = 1$ for all $i \in S$. The "clever" or "mixing" step is to consider a new sequence $\beta_{ij} = \sum_{k\in S} \alpha_{ik} p_{kj}$. By the convergence of $I(\mu_i^{(n)}|\pi)$ to L_i, we have

$$\lim_{k\to\infty} I(\mu_i^{(n_k)}|\pi) = L_i = I(\alpha_i|\pi)$$

and

$$\lim_{k\to\infty} I(\mu_i^{(n_k+1)}|\pi) = L_i = I(\beta_i|\pi).$$

Characterization: By the definition of β_{ij} and same Jensen calculation used before, we have $I(\beta_i|\pi) \le I(\alpha_i|\pi)$. Since equality holds above iff $\alpha_{ik} = c\pi_k$ for some constant c and all $k \in S$, we see $c = 1$ by summing on $k \in S$. Hence $\alpha_{ij} = \pi_j$ for all $j \in S$. To complete the proof, we just have to note that we have already proved that given *any* initial sequence n_m we can find a further subsequence m_k such that $\mu_i^{(m_k)}$ convergence to π. This suffices to prove the result.

• This proof requires strick positivity of the (p_{ij}). One should chase down where it was used.

• This proof showed that each step of the transition matrix strictly decreases the divergence unless $\mu_i^{(n)}$ is already exactly equal to π.

• This proof of the convergence of $\mu_i^{(n)}$ did not prove the existence of the stationary measure. In fact, the proof needed π even to get started talking about divergence. One can show the existence of π by establishing a easier limit result that $\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} p_{ij}^{(k)}$ exists.

• This relative entropy technique is of importance in the proof of asymptotic consistency of maximum likelihood estimates of Hidden Markov Models, Mixture models, and other models.

• There is an extension of this method to denumerable M.C. in: Kendall, D. (1963). Information theory and the limit-theorem for Markov chains and processes with a countable infinity of states. *Ann. Inst. Statist. Math.* **15**, 137–143.

• **Ref:** Rényi, A. (1961). On measures of entropy and information. *Proc. 4th Berkeley Sympos. Math. Statist. and Prob.*, Univ. California Press, Berkeley, Vol. I, 547–561.

Finite M.C with Transient and Absorbing States

Related questions for finite M.C with transient and absorbing states:

(a). Where does absorption take place?

(b). How long does it take on average?

(c). What is the mean number of visits to a non-absorbing state prior to absorption?

(d). What is the probability that the M.C never visits a non-absorbing state?

In general, consider

$$\mathbf{P}_{0\le i,j\le N} = \begin{pmatrix} \mathbf{Q}_{r\times r} & \mathbf{R}_{r\times(N-r+1)} \\ \mathbf{0}_{(N-r+1)\times r} & \mathbf{I}_{(N-r+1)\times(N-r+1)} \end{pmatrix}$$

where states $0, 1, \cdots, r-1$ are transient ($P_{i,j}^{(n)} \to 0$ as $n \to \infty$ for $0 \le i, j \le r-1$) and states $r, r+1, \cdots, N$ are absorbing ($P_{ii} = 1$ for $r \le i \le N$).

Let $T = \min\{n \ge 0 : X_n \ge r\}$ be the absorption time.

(a). Let $u_{ik} = \mathbb{P}(X_T = k | X_0 = i)$, $k = r, \cdots, N$, be the probability of absorption at state k, starting from state i. Then u_{ik} can be solved from

$$u_{ik} = P_{ik} + \sum_{j=0}^{r-1} P_{ij} \cdot u_{jk}, \qquad i = 0, 1, \cdots, r-1.$$

(b). Let $v_i = \mathbb{E}(T | X_0 = i)$, $i = 0, 1, \cdots, r-1$, be the mean time to absorption, stating from state i. Then v_i can be solved from

$$v_i = 1 + \sum_{j=0}^{r-1} P_{ij} \cdot v_j, \qquad i = 0, 1, \cdots, r-1.$$

(c). Let $s_{ik} = \mathbb{E}\left(\sum_{n=0}^{T-1} 1_{(X_n = k)} | X_0 = i\right)$, $k = 0, 1, \cdots, r-1$, be the average number of visit to k before absorption, starting from state i. Then s_{ik} can be solved from

$$s_{ik} = \delta_{ik} + \sum_{j=0}^{r-1} P_{ij} \cdot s_{jk}, \qquad i = 0, 1, \cdots, r-1.$$

(d). Let $f_{ik} = \mathbb{P}(X_n = k \text{ for some } n | X_0 = i)$, $k = 0, 1, \cdots, r-1$, be the probability of ever enter k, starting from state i. Then

$$f_{ik} = (s_{ik} - \delta_{ik})/s_{kk},$$

Thus the probability of never enter state k is $1 - f_{ik}$.

In matrix notation,

$$(s_{ik})_{0 \le i,k \le r-1} = \mathbf{S}_{r \times r} = (\mathbf{I} - \mathbf{Q})^{-1}$$

and it is called fundamental matrix of $\mathbf{Q}$. Note that in the textbook, $\mathbf{Q}$ is represented as $\mathbf{P_T}$.

We have

$$(u_{ik})_{0 \le i \le r-1, r \le k \le N} = \mathbf{U}_{r \times (N-r+1)} = \mathbf{S} \cdot \mathbf{R}.$$

and

$$v_i = \sum_{k=0}^{r-1} s_{ik}.$$

M.C. Monte Carlo Methods

MCMC is a computer simulation method with broad applicability in physics, molecular biology and statistics. The two main challenges in MCMC are developing techniques to analyze the mixing time, or convergence rates, of MC on large state spaces and designing fast Monte Carlo algorithms to address specific applications.

Basic Ideas: Let $\mathbf{\Pi} = (\pi_0, \pi_1, \cdots, \pi_M)$ denote a distribution with $\pi_i > 0$ for $i \in S$, and $\sum_{i=0}^{M} \pi_i = 1$. We wish to estimate

$$I = \mathbb{E}_{\mathbf{\Pi}}(f) = \sum_{i \in S} f(i)\pi_i$$

where f is a function on state space S.

We may do this by constructing a transition matrix $\mathbf{P} = (P_{ij})$ so that $\mathbf{\Pi}$ is the unique stationary distribution of $\mathbf{P}$, i.e. $\mathbf{\Pi} = \mathbf{\Pi}\mathbf{P}$. We then simulate a MC $\{X_n\}$ with state space S and transition matrix $\mathbf{P}$ for time $n = 0, 1, \cdots$. Under fairly general conditions,

$$\hat{I}_N = \frac{1}{N} \sum_{i=1}^{N} f(X_i) \to I$$

with prob. one.

Metropolis-Hastings Algorithm

Let $\mathbf{Q} = (q(i,j))$ be any irreducible MC, a proposed MC. When $X_n = i$, generate a r.v. Y with $\mathbb{P}(Y = j) = q(i,j)$, i.e. sampling a candidate point Y from the proposed distribution $q(i,j)$.

If $Y = j$, then set

$$X_{n+1} = \begin{cases} j & \text{with prob. } \alpha(i,j) \\ i & \text{with prob. } 1 - \alpha(i,j) \end{cases}$$

i.e, the candidate point Y is then accepted with prob. $\alpha(i,j)$. Thus

$$P_{ij} = \begin{cases} q(i,j)\alpha(i,j) & \text{if } j \neq i \\ q(i,i) + \sum_{k \neq i} q(i,k)(1 - \alpha(i,k)) & \text{if } j = i \end{cases}$$

The MC $\mathbf{P}$ is time reversible if $\pi_i P_{ij} = \pi_j P_{ji}$, which requires

$$\alpha(i,j) = \min\left(1, \frac{\pi_j q(j,i)}{\pi_i q(i,j)}\right)$$

- To define $\mathbf{P}$, one only need the ratios π_j/π_i, $i,j \in S$.
- It is important to pick a good proposed MC $\mathbf{Q}$.
- History: Metropolis (1953), Hastings(1970).

• Billera, L. and Diaconis, P. (2001), A Geometric Interpretation of the Metropolis-Hastings Algorithm, *Statist. Sci.* **16**, 335-339

Abstract: The Metropolis-Hastings algorithm transforms a given stochastic matrix into a reversible stochastic matrix with a prescribed stationary distribution. We show that this transformation gives the minimum distance solution in an L^1 metric.

• Fishman, G. (1996), *Monte Carlo, Concepts, Algorithms and Applications*, Springer, New York.

• Liu, J. (2001). *Monte Carlo Techniques in Scientific Computing*, Springer, New York.

• Cobb, G. and Chen, Y. (2003), An application of MCMC to community ecology, *AMS Monthly*, **110**, April, 265-288.

• Jones, G. and Hobert, J. (2001), Honest Exploration of Intractable Probability Distributions via Markov Chain Monte Carlo, *Statist. Sci.* **16**, 312-334

Abstract: Two important questions that must be answered whenever a Markov chain Monte Carlo (MCMC) algorithm is used are (Q1) What is an appropriate burn-in? and (Q2) How long should the sampling continue after burn-in? Developing rigorous answers to these questions
presently requires a detailed study of the convergence properties of the underlying Markov chain. Consequently, in most practical applications of MCMC, exact answers to (Q1)and (Q2) are not sought. The goal of this paper is to demystify the analysis that leads to honest answers to (Q1) and (Q2). The authors hope that this article will serve as a bridge between those developing Markov chain theory and practitioners using MCMC to solve practical problems.

Open Questions for the List Model Based on One-closer Rule:

- Find a method for fast sampling.
- Find the rate of convergence to stationary.

Hidden Markov Model

Hidden Markov models (HMMs) have become a tool for modelling sequences of dependent random variables during the last decade. They have been extensively considered in signal processing.

A hidden Markov model (HMM) is a bivariate process $(X, Y) = \{(X_k, Y_k)\}$ such that X is a Markov chain, the Y's are conditionally independent given the X's, and the corresponding conditional distribution of Y_n depends on X_n only. Only Y is available to the observer; the process X is non-observable, or hidden. Often the state space of X is assumed finite, but it can be any compact space. Sometime, HMM is viewed as a Markov chain that is observed with noise. Estimation problems include the transition kernel of X and the conditional law of Y_n given X_n.

• Rabiner, L.(1989), A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE*, **77**, 257-284.

• Elliott, R., Aggoun, L. and Moore, J. (1995), *Hidden Markov models: Estimation and control*, Applications of Mathematics, 29. Springer-Verlag.

• Bickel, P., Ritov, Y. and Ryden, T. (1998), Asymptotic normality of the maximum-likelihood estimator for general hidden Markov models, *Ann. Statist.* **26**, 1614–1635.

• Koski, T. (2001), *Hidden Markov Models for Bioinformatics*, Computational Biology Series, **2**. Kluwer Academic Publishers.

Interpretation of the Fundamental Matrix $(\mathbf{I} - \mathbf{Q})^{-1}$

The (i, j) entry in the fundamental matrix gives the expected number of times that a system which begins in the i-th non-absorbing state will be in the j-th non-absorbing state before it reaches an absorbing state. The sum of the entries in the i-th row gives the expected number of transitions starting in i-th non-absorbing state and continues until it reaches an absorbing state.

•Markov chains have several advantages as mathematical models: They are general enough to provide useful models for many situations in diverse fields; they are special enough to be mathematical tractable and easy to use; and they have been intensive studied and many results are known.

Fastest Mixing Markov Chain

•Fastest Mixing Markov Chain on a Path, S. Boyd, P. Diaconis, J. Sun, and L. Xiao, The American Mathematical Monthly, 113, 70-74, 2006

We consider a random walk on a path with n nodes, with symmetric transition probabilities, i.e., the probability of making a transition between node i and node i+1 is the same as making a transition from node i+1 to node i. For such a Markov chain the uniform distribution is an equilibrium distribution, and the rate of convergence of the distribution to uniform is determined by the smallest second-largest eigenvalue magnitude of the associated transition matrix. We address the question: What choice of transition probabilities results in the fastest mixing Markov chain on the path? This question can be posed in terms of matrices as: Among all symmetric, stochastic, tridiagonal matrices, what is the minimum value the second-largest eigenvalue magnitude can attain? In this note we prove that fastest mixing is obtained when we assign at each node a probability 1/2 of moving to the left, 1/2 of moving to the right, and a probability 1/2 of remaining at the two boundary nodes. The optimal second-largest eigenvalue magnitude is given by $\cos(\pi/n)$.

•The Fastest Mixing Markov Process on a Graph and a Connection to a Maximum Variance Unfolding Problem J. Sun, S. Boyd, L. Xiao, and P. Diaconis, SIAM Review, 48, 681-699, 2006

We consider a Markov process on a connected graph, with edges labeled with transition rates between the adjacent vertices. The distribution of the Markov process converges to the uniform distribution at a rate determinined by the second smallest eigenvalue of the Laplacian of the weighted graph. In this paper we consider the problem of assigning transition rates to the edges so as to maximize the second smallest eigenvalue, subject to a linear constraint on the rates. This is the problem of finding the fastest mixing Markov process (FMMP) on the graph. We show that the FMMP problem is a convex optimization problem, which can in turn be expressed as a semidefinite program, and therefore effectively solved numerically. We formulate a dual of the FMMP problem, and show that it has a natural geometric interpretation as a maximum variance unfolding (MVU) problem, i.e., the problem of choosing a set of points to be as far apart as possible, measured by their variance, while respecting local distance constraints.

•Minimizing Effective Resistance of a Graph A. Ghosh, S. Boyd and A. Saberi, SIAM Review, 50, 37-66, 2008.

The effective resistance between two nodes of a weighted graph is the electrical resistance seen between the nodes of a resistor network with branch conductances given by the edge weights. The effective resistance comes up in many applications and fields in addition to electrical network analysis, including, for example, Markov chains and continuous-time averaging networks. In this paper we study the problem of allocating edge weights on a given graph in order to minimize the total effective resistance, i.e., the sum of the resistances between all pairs of nodes. We show that this is a convex optimization problem, and can be solved efficiently either numerically, or, in some cases, analytically. We show that optimal allocation of the edge weights can reduce the total effective resistance of the graph (compared to uniform weights) by a factor that grows unboundedly with the size of the graph. We show that among all graphs with n nodes, the path has the largest value of optimal total effective resistance, and the complete graph the least.

EXAM I, Spring 2008

1. (12 pts). A coin, having probability p of landing heads and probability $q = 1 - p$ of landing tails, is continually flipped until at least two head and one tail have been flipped. Find the expected number of flips that lands on heads.

2. (18 pts). A Markov chain $X_0, X_1, X_2, \cdots$ has the transition probability matrix: $\mathbf{P} = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/4 & 1/2 & 1/4 \\ 1/2 & 1/2 & 0 \end{pmatrix}$ and the initial distribution $\mathbb{P}(X_0 = 0) = 0.4, \mathbb{P}(X_0 = 1) = 0.2, \mathbb{P}(X_0 = 2) = 0.4$. Determine the following:

(a). Is the Markov chain time reversible? Why?

(b). $\lim_{n\to\infty} \mathbb{P}(X_{n+4} = 2, X_{n+3} = 1, X_{n+2} = 0 | X_{n+1} = 0, X_n = 1)$.

(c). $\lim_{n\to\infty} \mathbb{P}(X_{n+2} = 0, X_{n+1} = 1, X_n = 2 \mid X_2 = 0)$.

(d). $\mathbb{P}(X_2 = 1 \mid X_0 = 2)$.

(e). $\lim_{n\to\infty} \mathbb{P}(X_n = 2, X_{n+1} = 0, X_{n+2} = 1)$.

(f). $\lim_{n\to\infty} \mathbb{P}(X_n = 0 \mid X_{2n} = 1)$

(g). The mean return time to state 2.

(h). The long run fraction of time the process spends in state 1.

(i). Starting from state 1, what is the expected number of visit to state 2 before returning to state 1?

3. (6 pts). Trials are performed in a sequence. If the last two trials were successes, then the next trial is a success with probability .8; otherwise the next trial is a success with probability .5. Define a Markov chain which will help you to determine the long run proportion of trials are successes. Give the one-step transition matrix but do NOT find the stationary distribution in order to save time.

4. (14 pts). Next semester your professor will teach Math630 and Math850 and possess three textbooks for Math 630 and two textbooks for Math850. He will always keep two at home and three at the office. At the beginning of a workday, he will randomly take one of his two textbooks at home to the office. And at the end of the day, he will randomly take one of his four textbooks in the office to home.

(1). Define a Markov chain with 3 states which will help you to determine the proportion of days that your professor can't use his textbook for Math850 at home.

(2). Is your Markov chain time reversible? Why?

(3). What fraction of workdays your professor can't use his textbook for Math850 at home?