1 An Introduction To Markov Chains

We are dealing with a discrete state space and time is discrete- called a "discrete parameter" process. The process is Markovian, so that:

$$prob\,(X_{n+1} = k \mid X_0 = h, \ldots\; X_n = j)$$
$$= prob\,(X_{n+1} = k \mid X_n = j)$$

is the transition probability, where time is indexed by "n" and the random variable X takes on the discrete values $\{i \in S\}$, $S \subset Z$, i.e. S is a subset of the integers.

Thus: $X_n = k \Rightarrow$ the random variable X at time period n is in the state designated by k.

1.1 A Worked Example To Illustrate The Basic Ideas- 2 States

Working example- Rainfall in Tel Aviv- see Cox & Miller, p.78.

We have 2 states $\{0, 1\}$; "failure, success"; or "no rain, rain."

If at the nth trial there is "failure", then the probabilities at the $(n+1)$ trial are:

$$prob(\text{failure at } n+1 \mid \text{ failure at } n) = 1 - \alpha$$
$$prob(\text{success at } n+1 \mid \text{ failure at } n) = \alpha$$

And if success at trial n, we have:

$$prob(\text{failure at } n+1 \mid \text{ success at } n) = \beta$$

$$prob(\text{success at } n+1 \mid \text{ success at } n) = 1-\beta$$

These transition probabilities can be written:

$$p = \begin{array}{c|cc} & 0 & 1 \\ \hline 0 & 1-\alpha & \alpha \\ 1 & \beta & 1-\beta \end{array}$$

p_{jk} is the probability of transiting from state j at n to state k on the $(n+1)$st. trial.

The estimated probability of dry/wet days in Tel Aviv are:

$$p = \begin{array}{ccc} & \text{Dry} & \text{Wet} \\ \text{Dry} & 0.75 & 0.25 \\ \text{Wet} & 0.34 & 0.66 \end{array}$$

The <u>rows</u> sum to 1.

Let $p^{(n)} = \left(p_0^{(n)}, p_1^{(n)}\right)$ denote the **probability** of finding the system in the state 0, or 1 at time "n", given the initial probabilities

$$p^{(0)} = \left(p_0^{(0)}, p_1^{(0)}\right).$$

Consider the move to state 0 at time n. This can occur in 2 mutually exclusive ways:

a) Were in state 0 at $(n-1)$ and <u>no</u> <u>transition</u> out.

b) Were in state 1 at $(n-1)$ and transition in.

Probability of a):

$$p_0^{(n-1)}(1-\alpha)$$

Probability of b):

$$p_1^{(n-1)}\beta$$

and as events are mutually exclusive we can add the probabilities.

We have therefore:

$$p_0^{(n)} \; = \; p_0^{(n-1)}(1-\alpha) \, + \, p_1^{(n-1)}\beta$$

$$p_1^{(n)} \; = \; p_0^{(n-1)}\alpha \, + \, p_1^{(n-1)}(1-\beta)$$

or

$$p^{(n)} = p^{(n-1)}P$$

therefore

$$p^{(n)} = p^{(0)}P^n$$

Note that the sequence $\left\{p_i^{(n)}\right\}_{i\,=\,0,\,1}$ is a sequence of marginal probabilities.

Let the $(j,\,k)$ element of P^n be: $p_{jk}^{(n)}$

If $p^{(0)} = (1, 0)\,, \qquad p^{(n)} = \left(p_{00}^{(n)}, p_{01}^{(n)}\right)$

If $p^{(0)} = (0, 1)\,, \qquad p^{(n)} = \left(p_{10}^{(n)}, p_{11}^{(n)}\right)$

$\uparrow$

these represent Dirac delta functions:
the probabilities are 0, or 1 only.

$p_{jk}^{(n)} = prob($state k at time $n \mid$ state j at time $0)$

<u>Note</u>: These are "n- step" transition probabilities.

<u>Example</u>: The Tel Aviv data:

$$p^5 = \begin{bmatrix} 0.58 & 0.42 \\ 0.57 & 0.43 \end{bmatrix} \qquad p^{10} = \begin{bmatrix} 0.575 & 0.425 \\ 0.575 & 0.425 \end{bmatrix}$$

Above leads to the idea, is there an equilibrium distribution to which the process converges <u>**independent**</u> of initial conditions?

If so, we would have the following relationship:

$\pi = \pi P$, π is a row vector.

$$\Rightarrow \pi\,(I - P) = 0$$

in our example:

$$\pi_0\alpha - \pi_1\beta = 0 \qquad -\pi_0\alpha + \pi_1 p = 0$$

and a <u>**scale**</u> condition:

$$\pi_0 + \pi_1 = 1$$

Solving gives:

$$\pi_0 = \frac{\beta}{\alpha + \beta} \qquad \pi_1 = \frac{\alpha}{\alpha + \beta}$$

and

$$p^{(0)} = \pi,$$

then the probability sequence $p^{(n)}$ is stationary, since

$$p^{(2)} = p^{(1)}P = \pi P = \pi$$

and

$$p^{(1)} = \pi P = \pi$$

therefore

$$p^{(n)} = \pi, \quad \forall n.$$

If we begin with

$$p = \begin{bmatrix} 0.75 & 0,25 \\ 0.338 & 0.662 \end{bmatrix}$$

and we solve for π using

$$\pi \left(I - P \right) = 0$$

we get

$$\pi_0 = 0.575 \qquad \pi_1 = 0.425$$

c.f. P^{10} and you see that within 10 days we have converged already.

Two ways to determine P^n

a) Iterate as above

b) Solve algebraically

If P has distinct eigenvalues, then it can be decomposed into the product:

$$P = Q \Lambda Q^{-1},$$

where Λ is a diagonal matrix with the eigenvalues on the diagonal. The columns of Q contain the eigenvectors corresponding to each eigenvalue, i.e:

$$Pq_i = \lambda_i q_i;$$

having found λ_i, you solve for the vector, q_i.

$i = 1$, 2,... n.

It is easy to show that:

$$P^n = Q\Lambda^n Q,$$

$$\Lambda^n = \begin{pmatrix} \lambda_1^n & & & \oslash \\ & \lambda_2^n & & \\ & & \ddots & \\ \oslash & & & \lambda_n^n \end{pmatrix}$$

Using our example:

$$P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$$

$$|P - \lambda I| = 0$$

$$\Rightarrow (1-\alpha-\lambda)(1-\beta-\lambda) - \alpha\beta = 0$$

<u>Solution</u>:

$$\lambda_1 = 1, \qquad \lambda_2 = 1-\alpha-\beta,$$

unique provided $\alpha + \beta \neq 0$.

Q is obtained by solving:

$$Pq_i = \lambda_i q_i, \quad \lambda_i = 1, \quad 1-\alpha-\beta$$

$$Q = \begin{pmatrix} 1 & \alpha \\ 1 & -\beta \end{pmatrix} \qquad Q^{-1} = (\alpha+\beta)^{-1}\begin{bmatrix} \beta & \alpha \\ 1 & -1 \end{bmatrix}$$

So

$$P^n = (\alpha+\beta)^{-1} \begin{pmatrix} 1 & \alpha \\ 1 & -\beta \end{pmatrix} \begin{bmatrix} 1 & 0 \\ 0 & (1-\alpha-\beta)^n \end{bmatrix} \begin{bmatrix} \beta & \alpha \\ 1 & -1 \end{bmatrix}$$

can be rewritten as:

$$(\alpha+\beta)^{-1} \begin{bmatrix} \beta & \alpha \\ \beta & \alpha \end{bmatrix} + \frac{(1-\alpha-\beta)^n}{\alpha+\beta} \begin{bmatrix} \alpha & -\alpha \\ -\beta & \beta \end{bmatrix}$$

The first term is just $\begin{bmatrix} \pi_0 & \pi_1 \\ \pi_0 & \pi_1 \end{bmatrix}$ as derived above and the second term converges to $\oslash$.

There are other questions of interest:

a) Proportion of time spent in each state

b) Probability of 1st return to a given state

c) Time, and average time, to first return

d) Is the sequence of probability distributions periodic?

e) Is the process ergodic?

f) Is the chain irreducable (to a sequence of closed sub-chains).

We will address some of these issues. But first, we need to extend the analysis to general n states and even to denumerable states.

1.2 Some Definitions

A Markov chain is "homogeneous" when the conditional probability depends only on the state and the

number of transition steps; i.e.
$prob\,(X_{m+n} = k \mid X_m = j) = prob\,(X_n = k \mid X_0 = j)$.

Let
$$p_{jk} = p_{jk}^{(1)} = prob\,(X_{n+1} = k \mid X_n = j)\,, \quad \forall m.$$

Since the system must move to some state from any initial state (including “no move at all”)

$$\sum_{k=0}^{\infty} p_{jk} = 1$$

P, the matrix of transition probabilities (1 step):

$$P = (p_{jk}) = \begin{bmatrix} p_{00} & p_{01} & p_{02} & \cdots \\ p_{10} & p_{11} & p_{12} & \cdots \\ \vdots & & \ddots & \end{bmatrix}$$

Example: Simple random walk on a lattice with reflecting barrier at origin.

State of the system: $\{0$, 1, 2,...$\}$

Let
$$p_{j,\,j+1} = p, \qquad j = 0,\ 1,\ 2,...$$
$$p_{jj} = 1 - p - q, \qquad j = 1,\ 2,...$$

$$p_{00} = 1 - p$$
$$p_{j,\,j-1} = q$$
$p_{jk} = 0$, otherwise

$$
P = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ \vdots \end{array}
\begin{array}{c}
\begin{array}{cccccc} 0 & 1 & 2 & 3 & 4 & \end{array} \\
\left[\begin{array}{cccccc}
1-p & p & 0 & 0 & 0 & \ldots \\
q & 1-p-q & p & 0 & 0 & \ldots \\
0 & q & 1-p-q & p & 0 & 0\ldots \\
0 & 0 & q & 1-p-q & p & 0\ldots \\
0 & 0 & 0 & q & & \\
\vdots & \vdots & \vdots & \ddots & &
\end{array}\right]
\end{array}
$$

Using the notation:

$$\left(p_{jk}^{(n)}\right) = p^n,$$

where p^n can be defined, we have that:

$$p^{m+n} = p^m p^n$$

or

$$p_{jk}^{(m+n)} = \sum_p p_{jl}^{(m)} p_{lk}^{(n)}$$

and we recall this is the discrete version of the Chapman-Kolmogorov equation for a homogeneous Markov chain.

Since

$$pi = i, \qquad p^n i = i$$

and so all p^n are stochastic matrices.

1.3 The Classification Of States

Markov chains can be classified into a set of different states depending on limiting behavior. We begin with the system in state j.

We define:

$f_{jj}^{(n)}$ = probability that next occurrence of j is at step n.

$$f_{jj}^{(1)} = p_{jj}$$

$$f_{jj}^{(n)} = prob\,(X_r \neq j, r = 1, 2, \ldots\ n-1; X_n = j \mid X_0 = j)\,, \quad n = 2, 3, \ldots$$

$f_{jj}^{(n)}$ is the first return probability for state j in n steps.

Define:

$f_{jk}^{(n)}$ = first passage (j to k) probability in step n.

$$f_{jk}^{(n)} = prob\,(X_r \neq k, r = 1, 2, \ldots\ n-1; X_n = k \mid X_0 = j)$$

Starting in state j, the probability that state j is **eventually** re-entered is given by:

$$f_j = \sum_{n=1}^{\infty} f_{jj}^{(n)}$$

Definitions:

If $f_j = 1$, i.e. eventual return is certain, the state is said to be "recurrent."

If $f_j < 1$, the state j is said to be "transient."

If state j is transient, the probability of **never** returning to state j is $1 - f_j > 0$.

Definition:

For a recurrent state, j, $f_{jj}^{(n)}$ $n = 1, 2,...$ is a probability distribution. We define its mean value, the "mean recurrent time" by:

$$\mu_j = \sum_{n=1}^{\infty} n f_{jj}^{(n)}$$

If μ_j is ∞, state j is null-recurrent.

Definition:

Since $f_{jk} = \sum_{n=1}^{\infty} f_{jk}^{(n)} =$ probability of **ever** entering state k; if $f_{jk} = 1$, then the first passage from j to k is **certain**.

$\mu_{jk} = \sum_{n=1}^{\infty} n f_{jk}^{(n)} =$ mean first passage time j to k.

Starting in state j, if subsequent returns to state j have non-zero probability for times kl, l some integer$>$ j, $k = 0, 1, 2,...$ then state j is said to be periodic; note this implies

$$p_{jj}^{(n)} \equiv 0 \; \forall n,$$

except for $n = kl$, $k = 0, 1, 2,...$

Definition:

A state which is not periodic, is said to be "**aperiodic**."

An aperiodic state which is positive recurrent is called "**ergodic**." The system is ergodic if all states are ergodic.

Definition:

A state j is said to "communicate" with a state k, if there **exists** a **finite** $\underline{\mathbf{r}}$ such that $p_{jk}^{(r)} > 0$.

If j communicates with k and k with j, they are inter-communicating.

Karlin & Taylor say that k is "accessible" for j if $p_{jk}^{(r)} > 0$, for some finite r and that (j, k) communicate, if j is accessible for k and k for j.

If two states do not communicate (KT), either

$$p_{jk}^{(n)} = 0 \; \forall n \geq 0$$

or

$$p_{kj}^{(n)} = 0 \; \forall n \geq 0$$

or both.

The concept of communication (KT) is an equivalence relation

i) $j \longleftrightarrow j$, reflexive.

ii) if $i \longleftrightarrow j$, then $j \longleftrightarrow i$, symmetry.

iii) if $j \longleftrightarrow k$, $k \longleftrightarrow i$, then $j \longleftrightarrow i$ transitivity (???).

Recognizing that communication creates equivalence classes is important.

Definition:

A **set** of states, C, is called "closed" if each state in C communicates only with other states in C.

A **single** state that is **closed** is an "**absorbing**

state." Once a closed set is entered, it is **never** left.

1.4 Decomposition Theorem

(a) The states of an arbitrary Markov chain may be divided into 2 sets (one of which may be empty); one set is the set of all recurrent states; the other is the set of all transient states.

(b) The recurrent states can be decomposed into closed sets uniquely.

Within each closed set all states communicate (KT), they are all of the same type and period.

Between any two closed sets, there is no communication.

1.4.1 Implication For Finite Chains

A finite chain cannot consist only of transient states, for if so, then all $p_{jk}^{(n)} \to 0$ as $n \to \infty$; but we require that

$$\sum_k p_{jk}^{(n)} = 1$$

A finite chain cannot have any null-recurrent states.

Definition:

An **irreducible** chain is one in which all states communicate (KT), recall that communication is an equivalence relation that induces an equivalence class.

- an irreducible chain forms a single closed set.

By the decomposition theorem, the properties of a general Markov chain depend on those of the irreducible chains and on the transient states.

An irreducible **finite** chain must be positive recurrent, if further it is aperiodic, then it is ergodic and $\exists$ a unique equilibrium distribution π.

1.5 Limiting Properties Of Irreducible (infinite) Chains

Assume all chains are aperiodic.

The first major question **infinite** state space chains is whether the chain is

ergodic,	null-recurrent,	or transient
$\uparrow$	$\uparrow$	$\uparrow$
all states revisited with finite recurrence times	recurrence, but only at ∞ recurrence times	positive probability some states never revisited

1.5.1 Theorem

For an irreducible ergodic chain the limiting probabilities (π_k) form a probability distribution satisfying the equilibrium conditions and any solution of the equations:

$$x_k = \sum_{j=0}^{\infty} x_j p_{jk}, \quad k = 0, 1, 2, ...$$

i.e. $x = xp$

$\ni |x_k| < \infty$, and is a scalar multiple of (π_k).

The equilibrium conditions are:

$$\pi_k = \sum_{j=0}^{\infty} p_{jk}\pi_j; \quad k = 0, 1, 2, ...$$

The converse theorem is also true; i.e. any chain satisfying the above conditions is ergodic, that is, if the equations

$$x_k = \sum_{j=0}^{\infty} x_j p_{jk}, \quad k = 0, 1, 2, ...$$

have an absolutely convergent solution (that is, if $\ddot{x}_j$ is a solution, $\sum_{j=0}^{\infty} |\ddot{x}_j| < \infty$), then the process is ergodic, provided only that $\ddot{x}_j$ is everywhere non-negative.

1.5.2 Transient Systems

An infinite irreducible aperiodic chain is **transient** if and only if the system of equations below has a non-zero, bounded solution

$$y_j = \sum_{k=1}^{\infty} p_{jk} y_k$$

Note the order

The idea is that, if, for any state, say 0, we can solve the equations

$$f_{j0} = p_{j0} + \sum_{k=1}^{\infty} p_{jk} \underbrace{f_{k0}}$$

$\swarrow$ probability of ever reaching 0 from j

$\downarrow$ probability of ever reaching 0 from state k

obtain a solution with $f_{j0} < 1$, then we can be assured that the system is transient.

1.5.3 Recurrence

An irreducible aperiodic system is recurrent if there is a **solution** of the following inequalities, such that $y_j \to \infty$ as $j \to \infty$

$$\sum_{k=1}^{\infty} p_{jk} y_k \le y_j, \quad j = 1, 2, ...$$

There are a lot more results on chains, with and without finiteness of the state space.

Some useful books are:

Cox & Miller- Stochastic Processes; a good basic book, good examples.

Karlin & Taylor- A 1st Course in Stochastic Processes; regarded as a standard treatment, better on the algebraic details than Cox & Miller.

Feller- Intro to the Theory of Probability; a classic, a must have book, but not a good text book.

Kemeng & Snoll- Finite Markov Chains; a well-regarded book, but is limited to finite chains.

The big key to the analysis of such problems is the **formulation** of the basic problem, and creation of the transition probability matrix.

1.6 Some Discussion Of Some Examples

A) The Mover-Stayer Model of Blumen, Kogen, & McCarthy (1955)

Industry is in c categories; recorded duration of stay and times of transit between categories for workers

$$\begin{array}{ll} \text{Stayers-} & \text{never move} \\ \text{Movers-} & p_{jk} \neq 0 \ \forall j, k \end{array}$$

States: industrial categories 1, 2,... c.

Each individual provides over time a sample path; statistical problem is to estimate p_{jk}, the transition probabilities.

B) Random Walk

- provides a variety of prototypical examples of use in a variety of disciplines

- Random Walk (simple 1 dim.) with reflective barrier was mentioned above

- 3 dim. random walk

- Random walk that is self avoiding; if at time n particle is at position $(i,\, j)$, then at time $(n+1)$ the particle can move to one of:

$$(i,\, j+1)\,,\;(i+1,\, j)\,,\;(i-1,\, j)\,,\;(i,\, j-1)$$

provided that no previous point visited is revisited. This might be the model for randomized search.

Note this is an example of a **non**-Markovian process that cannot be converted into a Markov process by an extension of the state space.

- Random walk with absorbing barriers; eg. states $\{0,\,1,\,2,...\,a\}$, $\{0,\,a\}$ absorbing

$$p_{aa} = p_{00} = 1$$
$$p_{ak} = p_{0k} = 0$$

$$P = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ \vdots \\ a \end{array} \begin{array}{c} \begin{array}{cccccc} 0 & 1 & 2 & 3\ldots & & a \end{array} \\ \left[\begin{array}{cccccc} 1 & 0 & 0 & 0\ldots & & 0 \\ q & 1-p-q & p & 0\ldots & & 0 \\ 0 & q & 1-p-q & p & 0\ldots & 0 \\ & & q & & & \\ & & & & & \\ 0 & 0 & 0 & \ddots & & 1 \end{array} \right] \end{array}$$

1.7 Ehrenfest Model Of Diffusion

Originally derived to model a gas separated by a wall (permeable) into 2 regions. Question: if at fixed points of Five there is a random migration from one region to the other, how do the probabilities of being in each region change over time?

This model can be given an economic interpretation- suppose we have s computer owners; at time t, s_1 are Mac users and s_2 are PC users. Question: given the properties assumed for the transitions, how does the proportion of Mac and PC users change over time?

We could easily add absorbing boundaries and state dependence.

Mobility of people between **income** levels can be formulated as a Markov process and its properties evaluated. Of key interest here, is the mean time to exit a state, or to enter a higher state; the probability of eventual return to a given state, and so on. One major question is whether an equilibrium distribution exists.

Kemeng & Snoll discuss converting input/output problems into a Markov process- see pg. 200-206.

A final comment on finite chains.

If P can be factored as follows:

$$P = \begin{array}{c|cc} & C & T \\ \hline C & P_1 & \varphi \\ T & R_{21} & Q_2 \end{array} \qquad\qquad R_{21} = \varphi$$

P_1, Q_1 are irreducible square matrices.

Let P_1 be "primitive", that is, the maximum modulus of the eigenvalues is 1 and it is unique.

$P_1 \sim$ transition probabilities within C

$Q_1 \sim$ transition probabilities within T

$R_{21} \sim$ transition from T to C

Transitions out of C are impossible and C is a closed set, states in C are ergodic, those in T are transient.

The equilibrium distribution is over states in C.

(Saved under x:\ user\ramseyj\notes\mkvchain.tex)