The beauty, potential and impediments of Markov chains

Dr Joshua Ross

THE UNIVERSITY OF ADELAIDE AUSTRALIA

Operations Research & Statistics Group
School of Mathematical Sciences
The University of Adelaide

September 2010

Peter Whittle

“Any stochastic process can be formulated as a Markov process by an appropriate choice of variables, and for most physical models the Markov formulation is a natural one, so that no generality and little convenience is lost if we restrict our attention to processes of the Markov type." [1]

[1] P. Whittle (1957) On the use of the normal approximation in the treatment of stochastic processes, *JRSS-B*, 19, 268–281.

Andrei Andreyevich Markov

[2]

14 June 1856 – 20 July 1922

[2] Copyright unknown. Downloaded on 4/09/10 from http://www-history.mcs.st-and.ac.uk/Biographies/Markov.html.

The Markov property

For all (finite) sequences of times $0 < t_1 < t_2 < \cdots < t_n < t_{n+1}$ and states $j_1, j_2, \ldots, j_n, j_{n+1} \in S$, we have

$$\Pr\left(X(t_{n+1}) = j_{n+1} | X(t_n) = j_n, \ldots, X(t_1) = j_1\right) = \Pr\left(X(t_{n+1}) = j_{n+1} | X(t_n) = j_n\right).$$

Given the past and the present, only the present is of use in predicting the future.

Markov chain theory – very quickly!

Let $p_{ij}(t) = \Pr\left(X(t+s) = j | X(s) = i\right)$ (note elapsed time).

$P(t) = (p_{ij}(t),\ i,j \in S)$ is the Transition Function.

Provided $P(t)$ is standard (satisfies a simple property), then

(i) $q_i := \lim_{t \downarrow 0} \frac{1 - p_{ii}(t)}{t}$ exists and $0 \le q_i \le \infty$;

(ii) $q_{ij} := \lim_{t \downarrow 0} \frac{p_{ij}(t)}{t}$ exists $\forall j \neq i$ and $0 \le q_{ij} < \infty$.

Note: q_{ij} is the *transition rate* from state i to state j ($j \neq i$);
q_i is the transition rate out of state i.

Think of q_{ij} and q_i as

$$p_{ij}(\epsilon) = q_{ij}\epsilon + o(\epsilon) \quad (j \neq i)$$

$$p_{ii}(\epsilon) = 1 - q_i\epsilon + o(\epsilon).$$

Markov chain theory – very quickly!

$Q = (q_{ij},\ i, j \in S)$ is the q-matrix.

A q-matrix Q is said to be bounded if $q := \sup_i\{q_i\} < \infty$.

If Q is bounded, then

$$p'_{ij}(t) = \sum_{k \in S} p_{ik}(s) q_{kj} \quad \forall i, j \in S$$

or

$$P'(t) = P(t)Q$$

and

$$P(t) = e^{tQ} := \sum_{n=0}^{\infty} Q^n \frac{t^n}{n!}$$

is the unique transition function for the q-matrix Q.

Markov chain theory – very quickly!

Sample behaviour of general Markov chains.

Let T_{jn} be the time spent in state j on the n-th visit.

(1) $T_{jn} \sim \text{Exp}(q_j)$.

(2) Once this time has elapsed then next state is k with probability $p_{jk} = \frac{q_{jk}}{q_j}$.

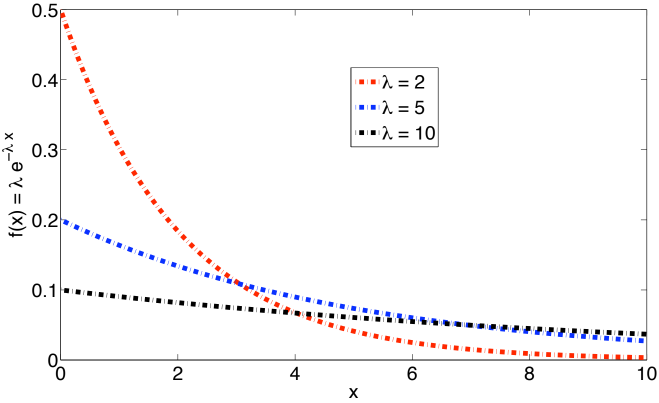
(3) Given the sequence of states visited, successive holding times are independent.

Exponential holding times

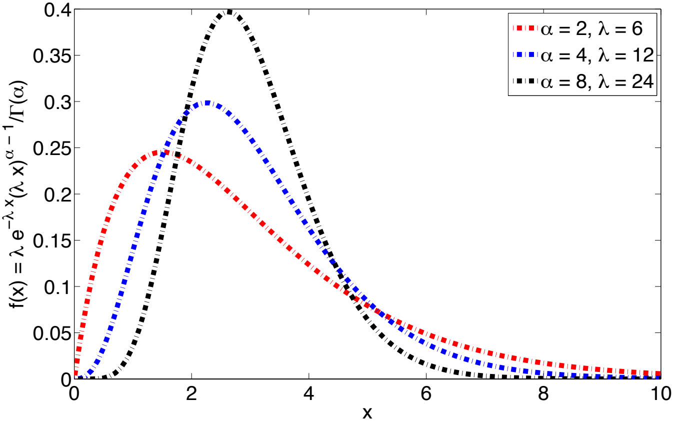
The holding time in each state is Exponentially distributed.

The lifetime of a sea otter (*Enhydra lutris*) is not Exponentially distributed, or the infectious period of a person with the ‘flu is not Exponentially distributed, so Markov chains are useless.

Exponential holding times



Phase-type distributions: Gamma



Phase-type distributions: Gamma

A) $$\longrightarrow \boxed{i} \xrightarrow{q_i(=\lambda)}$$

B) $$\longrightarrow \boxed{i_1} \xrightarrow{\alpha\lambda} \boxed{i_2} \xrightarrow{\alpha\lambda} \cdots \xrightarrow{\alpha\lambda} \boxed{i_\alpha} \xrightarrow{\alpha\lambda}$$

Phase-type distributions

Phase-type distributions describe the random time taken for a Markov chain to reach an absorbing state.

The set of phase-type distributions is dense in the field of all positive-valued distributions. Therefore it can be used to approximate any positive-valued distribution.

Dependence upon the past

The Markov property

$$\Pr\left(X(t_{n+1}) = j_{n+1} | X(t_n) = j_n, \ldots, X(t_1) = j_1\right) = \Pr\left(X(t_{n+1}) = j_{n+1} | X(t_n) = j_n\right),$$

doesn't hold for my process, as the jumps possible (and their respective rates) depend upon where I was two (in general *m*) jumps ago, so Markov chains are useless.

Dependence upon the past

Example: $X(t)$ is a model for habitat suitability based upon rainfall; the habitat is classified as being *dry*, *standard*, or *wet*.

Suitability transitioning from dry to wet results in different transition rates to if suitability had changed from standard to wet – a large downpour is more likely to be followed by some rain, for example.

Dependence upon the past

It is the previous *state* which the future *state* must only depend upon.

Define

$$Y(t_n) = (X(t_n), X(t_{n-1}), \ldots, X(t_{n-m}))\,,$$

then the Markov property holds!

Hitting times

The random variable *first hitting time* (FHT) corresponds to the first time a CTMC enters a specified state j.

It is of great importance in population modelling, where we are interested in the extinction time of a species, corresponding to the first time the chain enters state 0.

Definition

Let $T^j = \inf\{t : X(t) = j\}$, the first time the chain enters state j.

We will work with the Laplace-Stieljtes transform (LST) of T^j, conditioned on the process starting in state i:

$$\tau_i^j(\theta) = \mathbb{E}[e^{-\theta T^j} | X(0) = i]$$

(note $0 \leq \tau_i^j(\theta) \leq 1$) where $X(0)$ is the initial state of our Markov chain.

Calculation

By conditioning on the time of the first jump, and the state visited at that time, we have:

$$\tau_i^j(\theta) = \int_0^\infty \sum_{k \neq i} e^{-\theta t} \tau_k^j(\theta) q_i e^{-q_i t} \frac{q_{ik}}{q_i} dt, \text{ for } i \neq j$$

(and $\tau_j^j(\theta) = 1$), which evaluates to:

$$\tau_i^j(\theta) = \frac{1}{q_i + \theta} \sum_{k \neq i} \tau_k^j(\theta) q_{ik}, \text{ for } i \neq j.$$

That is,

$$Q\tau^j(\theta) = \theta \tau^j(\theta),$$

where $\tau^j(\theta) = (\tau_i^j(\theta), i \in S)$.

Discussion

Hence we have a system of linear equations for which the LST of the distribution of FHT satisfies. It can be easily seen that if $|S| < \infty$ then there is a unique solution. In the case of an infinite state space, the solution we seek is the *maximal* solution [3].

This result combined with numerical Laplace inversion provides a method for evaluating the distribution of FHT [4].

Also, note, in case Q bounded we can also evaluate via $P(t) = e^{tQ}$, and finally note that we are evaluating a phase-type distribution!

[3] J. Norris (1997) *Markov chains*, Cambridge University Press

[4] J. Abate and W. Whitt (1995) Numerical inversion of Laplace transforms of probability distributions, *ORSA Journal on Computing*, 7, pp. 36–43.

Time until sustained progression

Many diseases, for example Multiple Sclerosis (MS), result in patients having periods (typically days or weeks) of *relapse* during which new symptoms appear, followed by periods (typically months or years) of *remission* during which patients fully or partially recover.

Eventually, many patients are found to progress to a more severe form of MS, called *secondary progressive*, characterised by gradual worsening of disease between relapses.

Time until sustained progression

The expanded disability status scale (EDSS) was introduced to quantify the status of an MS patient [5].

It ranges from 0 (normal neurological examination) to 10 (death due to MS), in half point steps.

Due to the remitting nature of MS, many studies focus on the *time to sustained progression*, typically defined as an increase in the EDSS level that lasts for six months or longer [6].

---

[5] http://www.mult-sclerosis.org/expandeddisabilitystatusscale.html

[6] Jacobs *et al.* (1996) Intramuscular interferon beta-1a for disease progression in relapsing multiple sclerosis, *Annals of Neurology*, 39, pp. 285–294

Time until sustained progression

Mandel considered data from a phase III clinical trial designed to evaluate Avonex, a drug for treating MS patients [7].

Patients had to have an EDSS smaller than 3.5 at enrollment.

Mandel assigned each patient a state based upon their EDSS:

$\{1\}$ = no disability (EDSS ≤ 1.5)
$\{2\}$ = minimal disability (EDSS $= 2, 2.5$)
$\{3\}$ = moderate or severe disability (EDSS ≥ 3).

[7] Mandel, M. (2010) Estimating disease progression using panel data, *Biostatistics*, 11, pp. 304–316.

Time until sustained progression

Mandel then fitted a continuous-time Markov chain model with three states, corresponding to the clumped EDSS statuses, based upon each patients EDSS at six-monthly intervals.

He evaluated the distribution of time to *sustained progression*, defined as sojourning in state 3 continuously for at least six months, for the treated and untreated patients, to discern the effectiveness of the drug.

This random variable corresponds to the first time a continuous-time Markov chain enters a state j and upon that visit sojourns there for at least Δ continuous time units, so we call it the *first sustained state time* (FSST).

Time until sustained progression

He calculated the distribution of time to sustained progression (FSST). He used a Monte Carlo (simulation) algorithm.

This was one of two methods he proposed, the other being a recursion formula requiring successive numerical integrations and involving an infinite sum.

Definitions

Let

$$A^j_\Delta(t) = I\{X(s) = j : t \leq s < t + \Delta\} \quad (t, \Delta \geq 0),$$

the indicator of being in state j continuously throughout the period $[t, t + \Delta)$.

FSST is

$$Y^{j,\Delta} = \inf\{t : A^j_\Delta(t) = 1\},$$

being the first time at which the process enters state j, and upon that visit sojourns there for at least Δ continuous time units.

Definitions

We will work with

$$y_i^{j,\Delta}(\theta) = \mathbb{E}[e^{-\theta Y^{j,\Delta}} | X(0) = i] \quad (0 \leq y_i^{j,\Delta}(\theta) \leq 1),$$

being the LST of the FSST $Y^{j,\Delta}$ conditioned on the process starting in state i.

Calculation

Condition on the time of first jump and state first visited:

$$
\begin{aligned}
y_i^{j,\Delta}(\theta) \quad = \quad & \left[\int_0^\infty \sum_{k \neq i} e^{-\theta t} y_k^{j,\Delta}(\theta) q_i e^{-q_i t} \frac{q_{ik}}{q_i} dt \right] I\{i \neq j\} + \\
& \left[\int_\Delta^\infty q_i e^{-q_i t} dt + \int_0^\Delta \sum_{k \neq i} e^{-\theta t} y_k^{j,\Delta}(\theta) q_i e^{-q_i t} \frac{q_{ik}}{q_i} dt \right] I\{i = j\}
\end{aligned}
$$

Calculation

which evaluates to

$$
\begin{aligned}
y_i^{j,\Delta}(\theta) \quad = \quad & \left[\frac{1}{q_i + \theta} \sum_{k \neq i} y_k^{j,\Delta}(\theta) q_{ik} \right] I\{i \neq j\} + \\
& \left[e^{-\Delta q_i} + \frac{1 - e^{-\Delta(\theta + q_i)}}{q_i + \theta} \sum_{k \neq i} y_k^{j,\Delta}(\theta) q_{ik} \right] I\{i = j\}.
\end{aligned}
$$

Discussion

It can be easily seen that if $|S| < \infty$ then there is a unique solution. In the case of an infinite state space, the proof of Proposition 1 of Pollett and Stefanov [8] can be modified slightly to establish that the solution we seek is the maximal solution to the system of equations.

The results established here may once again be combined with numerical Laplace inversion to provide a method for evaluating the distribution of FSST [9].

[8] P.K. Pollett and V.T. Stefanov (2002) Path integrals for continuous-time Markov chains, *Journal of Applied Probability*, 39, pp. 901–904.

[9] J. Abate and W. Whitt (1995) Numerical inversion of Laplace transforms of probability distributions, *ORSA Journal on Computing*, 7, pp. 36–43.

Mandel's MS Model

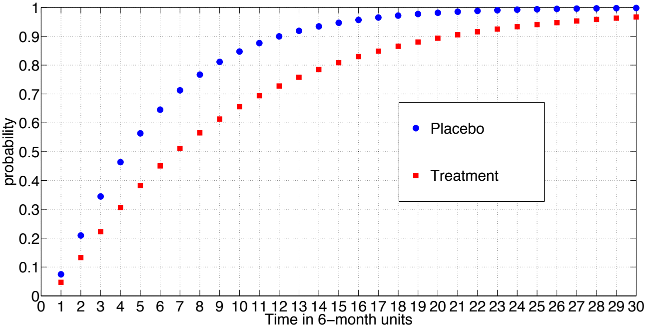


Figure: Distribution functions for time to sustained progression.

Growth in the dimensionality of the state space

Consider the SIS (*susceptible-infectious-susceptible*) model of disease dynamics.

$$q_{(s,i),(s-1,i+1)} = \beta i(N-s)/(N-1),$$
$$q_{(s,i),(s+1,i-1)} = \mu i.$$

Since $s + i = N$ at all times, we have state space

$$S = \{0, 1, \ldots, N\},$$

and hence the size of state space $|S| = N + 1$.

Growth in the dimensionality of the state space

Consider the SIR (*susceptible-infectious-recovered*) model of disease dynamics.

$$q_{(s,i,r),(s-1,i+1,r)} = \beta is/(N-1),$$
$$q_{(s,i),(s,i-1,r+1)} = \gamma i.$$

Now $s+i+r=N$, so we need to keep track of two of these, say (s,i), and the size of state space $|S| = (N+1)(N+2)/2$.

Phase-type distributions: Gamma

$A)$ $$\longrightarrow \boxed{i} \xrightarrow{q_i(=\lambda)}$$

$B)$ $$\longrightarrow \boxed{i_1} \xrightarrow{\alpha\lambda} \boxed{i_2} \xrightarrow{\alpha\lambda} \cdots \xrightarrow{\alpha\lambda} \boxed{i_\alpha} \xrightarrow{\alpha\lambda}$$

Phase-type distributions

We have added α extra dimensions to the state description!

Dependence upon the past

It is the previous *state* which the future *state* must only depend upon.

Define

$$Y(t_n) = (X(t_n), X(t_{n-1}), \ldots, X(t_{n-m}))\,,$$

then the Markov property holds!

Dependence upon the past

We have added *m* extra dimensions to the state description!

Growth in the dimensionality of the state space

Are there computationally-efficient ways of handling these chains? [10] [11] [12]

---

[10] G. Latouche and V. Ramaswami (1999) *Introduction to matrix-analytic methods in stochastic modelling*, SIAM. (G 769.)

[11] R. Sidje (1998) EXPOKIT: A software package for computing matrix exponentials, *ACM Trans. Math. Softw.*, 24, pp. 130–156.

[12] P. Pollett and D. Stewart (1994) An efficient procedure for computing quasi-stationary distributions of Markov chains with sparse transition structure, *Advances in Applied Probability*, 26, pp. 68–79.

Growth in the dimensionality of the state space

In many cases, our Markov chain is going to become impossible to handle practically.

Can we approximate the Markov chain by a process with reduced dimensionality, to sufficient accuracy?[13]

There are many methods available, all with slightly different assumptions and approaches. One question I am currently pursuing with collaborators is:
When do these methods give the same approximation?

[13] Keywords: Diffusion approximation; Moment-closure approximation; Saddlepoint approximation; Lumpability.

IP Method

The IP method involves identifying a subset $\hat{S}$ of S and considering the forward equations for those states in $\hat{S}$.

These forward equations will involve probabilities which are not part of the reduced state space $\hat{S}$, so we approximate these probabilities using the probabilities of the nearest $(d+1)$ states in $\hat{S}$ via a Lagrange interpolating polynomial of degree d.

The size of $\hat{S}$ is chosen to make computations feasible, or efficient.

What is the most efficient way of computing this approximation for high-dimensional systems?

Is there a way of approximating/bounding the error?[14]

[14] ... in the absence of some estimate of the rate of convergence, such a conclusion is rather lame. Ideally, one will have some idea of, or better some good upper bound for, the distance from x to $\xi \ldots$" Sir John Kingman, 1973.

Thanks!

Thanks for your attention!

Slides for this talk are available upon request:

`http://www.maths.adelaide.edu.au/joshua.ross`

Operations Research & Statistics Group

School of Mathematical Sciences

THE UNIVERSITY OF ADELAIDE AUSTRALIA