APTS Applied Stochastic Processes

Stephen Connor[1] & Matt Roberts[2]
stephen.connor@york.ac.uk and mattiroberts@gmail.com

(Some slides originally produced by Wilfrid Kendall, Christina Goldschmidt and Amanda Turner)

[1]Department of Mathematics, University of York

[2]Probability Laboratory, University of Bath

APTS Southampton, 8th-12th April 2019

Two notions in probability

> *"…you never learn anything unless you are willing to take a risk and tolerate a little randomness in your life."*
> *– Heinz Pagels,*
> *The Dreams of Reason, 1988.*

This module is intended to introduce students to two important notions in stochastic processes — reversibility and martingales — identifying the basic ideas, outlining the main results and giving a flavour of some significant ways in which these notions are used in statistics.

These notes outline the content of the module; they represent work-in-progress and will grow, be corrected, and be modified as time passes. Comments and suggestions are most welcome! Please feel free to e-mail us.

What you should be able to do after working through this module

After successfully completing this module an APTS student will be able to:

- describe and calculate with the notion of a reversible Markov chain, both in discrete and continuous time;
- describe the basic properties of discrete-parameter martingales and check whether the martingale property holds;
- recall and apply some significant concepts from martingale theory;
- explain how to use Foster-Lyapunov criteria to establish recurrence and speed of convergence to equilibrium for Markov chains.

First of all, read the preliminary notes . . .

They provide notes and examples concerning a basic framework covering:

- Probability and conditional probability;
- Expectation and conditional expectation;
- Discrete-time countable-state-space Markov chains;
- Continuous-time countable-state-space Markov chains;
- Poisson processes.

Some useful texts (I)

> *"There is no such thing as a moral or an immoral book. Books are well written or badly written."*
> *– Oscar Wilde (1854–1900),*
> *The Picture of Dorian Gray, 1891, preface*

The next three slides list various useful textbooks.
At increasing levels of mathematical sophistication:

1. Häggström (2002) "Finite Markov chains and algorithmic applications".
2. Grimmett and Stirzaker (2001) "Probability and random processes".
3. Breiman (1992) "Probability".
4. Norris (1998) "Markov chains".
5. Ross (1996) "Stochastic processes".
6. Williams (1991) "Probability with martingales".

Some useful texts (II): free on the web

1. Doyle and Snell (1984) “Random walks and electric networks” available on web at
   `www.arxiv.org/abs/math/0001057`.
2. Kelly (1979) “Reversibility and stochastic networks” available on web at
   `http://www.statslab.cam.ac.uk/~frank/BOOKS/kelly_book.html`.
3. Kindermann and Snell (1980) “Markov random fields and their applications” available on web at
   `www.ams.org/online_bks/conm1/`.
4. Meyn and Tweedie (1993) “Markov chains and stochastic stability” available on web at
   `www.probability.ca/MT/`.
5. Aldous and Fill (2001) “Reversible Markov Chains and Random Walks on Graphs” *only* available on web at
   `www.stat.berkeley.edu/~aldous/RWG/book.html`.

Markov chains and reversibility

"People assume that time is a strict progression of cause to effect, but actually from a non-linear, non-subjective viewpoint, it's more like a big ball of wibbly-wobbly, timey-wimey . . . stuff."
The Tenth Doctor,
Doctor Who, in the episode "Blink", 2007

Reminder: convergence to equilibrium

Recall from the preliminary notes that if a Markov chain X on a countable state space (in discrete time) is

- irreducible
- aperiodic (only an issue in discrete time)
- positive recurrent (only an issue for infinite state spaces)

then

$$\mathbb{P}\left[X_n = i | X_0 = j\right] \to \pi_i$$

as $n \to \infty$ for all states i.

π is the unique solution to $\pi P = \pi$ such that $\sum_i \pi_i = 1$.

A simple example

Consider simple symmetric random walk X on $\{0, 1, \ldots, k\}$, with "prohibition" boundary conditions: moves $0 \to -1$, $k \to k+1$ are replaced by $0 \to 0$, $k \to k$.

1. X is **irreducible** and **aperiodic**, so there is a unique equilibrium distribution $\pi = (\pi_0, \pi_1, \ldots, \pi_k)$.
2. The **equilibrium equations** $\pi P = \pi$ are solved by $\pi_i = \frac{1}{k+1}$ for all i.
3. Consider X in equilibrium:

$$\begin{aligned}\mathbb{P}\left[X_{n-1} = x, X_n = y\right] &= \mathbb{P}\left[X_{n-1} = x\right] \mathbb{P}\left[X_n = y | X_{n-1} = x\right] \\ &= \pi_x p_{x,y}\end{aligned}$$

   and

$$\mathbb{P}\left[X_{n-1} = y, X_n = x\right] = \pi_y p_{y,x} = \pi_x p_{x,y}.$$

4. In equilibrium, the chain looks the same forwards and backwards. We say that the chain is **reversible**. ▸ ANIMATION

Reversibility

Definition

Suppose that $(X_{n-k})_{0 \leq k \leq n}$ and $(X_k)_{0 \leq k \leq n}$ have the same distribution for every n. Then we say that X is **reversible**.

Detailed balance

1. Generalising the calculation we did for the random walk shows that a discrete-time Markov chain is **reversible** if it starts from equilibrium and the **detailed balance equations** hold: $\pi_x p_{x,y} = \pi_y p_{y,x}$. ‣ ANIMATION
2. **If** one can solve for π in $\pi_x p_{x,y} = \pi_y p_{y,x}$, then it is easy to show that $\pi P = \pi$.
3. So, if one can solve the detailed balance equations, and if the solution can be normalized to have unit total probability, then the result also solves the equilibrium equations.
4. In continuous time we instead require $\pi_x q_{x,y} = \pi_y q_{y,x}$, and if we can solve this system of equations then $\pi Q = 0$.
5. From a computational point of view, it is usually worth trying to solve the (easier) detailed balance equations first; if these are insoluble then revert to the more complicated $\pi P = \pi$ or $\pi Q = 0$.

Detailed balance and reversibility

Definition

The Markov chain X satisfies **detailed balance** if

Discrete time: there is a non-trivial solution of $\pi_x p_{x,y} = \pi_y p_{y,x}$;

Continuous time: there is a non-trivial solution of $\pi_x q_{x,y} = \pi_y q_{y,x}$.

Theorem

The irreducible Markov chain X satisfies **detailed balance** *and the solution $\{\pi_x\}$ can be normalized by $\sum_x \pi_x = 1$* **if and only if** *$\{\pi_x\}$ is an equilibrium distribution for X and X started in equilibrium is statistically the same whether run forwards or backwards in time.*

We will now consider progressively more and more complicated Markov chains:

- the $M/M/1$ queue;
- a discrete-time chain on a 8×8 state space;
- Gibbs samplers;
- and Metropolis-Hastings samplers (briefly).

$M/M/1$ queue

Here is a continuous-time example, the $M/M/1$ queue. We have

- Arrivals: $x \to x+1$ at rate λ;
- Departures: $x \to x-1$ at rate μ **if** $x > 0$.

Detailed balance gives $\mu\pi_x = \lambda\pi_{x-1}$ and therefore when $\lambda < \mu$ (stability) the equilibrium distribution is $\pi_x = \rho^x(1-\rho)$ for $x = 0, 1, \ldots$, where $\rho = \frac{\lambda}{\mu}$ (the traffic intensity). ▸ ANIMATION

Reversibility is more than a computational device: it tells us that if a stable $M/M/1$ queue is in equilibrium then people *leave* according to a Poisson process of rate λ. (This is known as Burke's theorem.)

Hence, if a stable $M/M/1$ queue feeds into another stable $\cdot/M/1$ queue then in equilibrium the second queue on its own behaves as an $M/M/1$ queue in equilibrium.

Random chess (Aldous and Fill 2001, Ch1, Ch3§2)

Example (A mean knight's tour)

Place a chess knight at the corner of a standard 8×8 chessboard. Move it randomly, at each move choosing uniformly from the available legal chess moves independently of the past.

1. Is the resulting Markov chain periodic?

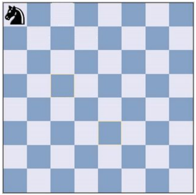
   (What if you sub-sample at even times?)

2. What is the equilibrium distribution?

   (Use detailed balance)

3. What is the mean time till the knight returns to its starting point?

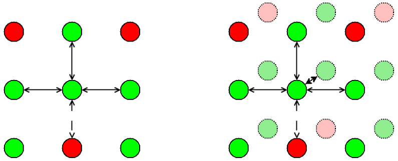
   (Inverse of equilibrium probability)



▸ ANIMATION

The Ising model

Pattern of spins $S_i = \pm 1$ on (finite fragment of) lattice (so i is a vertex of the lattice).



Probability mass function:

$$\mathbb{P}\left[S_i = s_i \text{ all } i\right] \propto \exp\left(J \sum_{i,j:i\sim j} s_i s_j\right)$$

or, if there is an external field $\{\widetilde{s}_i\}$,

$$\mathbb{P}\left[S_i = s_i \text{ all } i\right] \propto \exp\left(J \sum_{i,j:i\sim j} s_i s_j + H \sum_i s_i \widetilde{s}_i\right).$$

(Here $i \sim j$ means that i is a neighbour of j in the lattice.)

Gibbs sampler (or heat-bath) for the Ising model

For a configuration $\mathbf{s}$, let $\mathbf{s}^{(i)}$ be the configuration obtained from $\mathbf{s}$ by flipping spin i. Let $\mathbf{S}$ be a configuration distributed according to the Ising measure.

Consider a Markov chain with states which are Ising configurations on an $n \times n$ lattice, moving as follows.

- Suppose the current configuration is $\mathbf{s}$.
- Choose a site i in the lattice uniformly at random.
- Flip the spin at i with probability $\mathbb{P}\left[\mathbf{S} = \mathbf{s}^{(i)} \middle| \mathbf{S} \in \{\mathbf{s}, \mathbf{s}^{(i)}\}\right]$; otherwise, leave it unchanged.

Gibbs sampler for the Ising model

Noting that $s_i^{(i)} = -s_i$, careful calculation yields

$$\mathbb{P}\left[\mathbf{S} = \mathbf{s}^{(i)} \middle| \mathbf{S} \in \{\mathbf{s}, \mathbf{s}^{(i)}\}\right] = \frac{\exp\left(-J \sum_{j:j\sim i} s_i s_j\right)}{\exp\left(J \sum_{j:j\sim i} s_i s_j\right) + \exp\left(-J \sum_{j:j\sim i} s_i s_j\right)} \,.$$

We have transition probabilities

$$p(\mathbf{s}, \mathbf{s}^{(i)}) = \frac{1}{n^2}\,\mathbb{P}\left[\mathbf{S} = \mathbf{s}^{(i)} \middle| \mathbf{S} \in \{\mathbf{s}, \mathbf{s}^{(i)}\}\right], \quad p(\mathbf{s}, \mathbf{s}) = 1 - \sum_i p(\mathbf{s}, \mathbf{s}^{(i)})$$

and simple calculations then show that

$$\sum_i \mathbb{P}\left[\mathbf{S} = \mathbf{s}^{(i)}\right] p(\mathbf{s}^{(i)}, \mathbf{s}) + \mathbb{P}\left[\mathbf{S} = \mathbf{s}\right] p(\mathbf{s}, \mathbf{s}) = \mathbb{P}\left[\mathbf{S} = \mathbf{s}\right],$$

so the chain has the Ising model as its equilibrium distribution.

Detailed balance for the Gibbs sampler

Detailed balance calculations provide a much easier justification: merely check that

$$\mathbb{P}\left[\mathbf{S} = \mathbf{s}\right] p(\mathbf{s}, \mathbf{s}^{(i)}) = \mathbb{P}\left[\mathbf{S} = \mathbf{s}^{(i)}\right] p(\mathbf{s}^{(i)}, \mathbf{s})$$

for all $\mathbf{s}$.

Image reconstruction using the Gibbs sampler

Suppose that we have a black and white image that has been corrupted by some noise. Let $\widetilde{\mathbf{s}}$ represent the noisy image (e.g. $\widetilde{s}_i = 1$ if pixel i is black, and 0 if white), and use it as an external field, with $J, H > 0$. H here measures the "noisiness".

Bayesian interpretation: we observe the noisy signal $\widetilde{\mathbf{S}}$ and want to make inference about the true signal. We obtain posterior distribution $\mathbb{P}\left[\mathbf{S} = \mathbf{s} \middle| \widetilde{\mathbf{S}} = \widetilde{\mathbf{s}}\right] \propto \exp\left(J \sum_{i \sim j} s_i s_j + H \sum_i s_i \widetilde{s}_i\right)$ from which we would like to sample. In order to do this, we run the Gibbs sampler to equilibrium (with $\widetilde{\mathbf{s}}$ fixed), starting from the noisy image.

Image reconstruction using the Gibbs sampler

Here is an animation of a Gibbs sampler producing an Ising model conditioned by a noisy image, produced by systematic scans: 128×128, with 8 neighbours. The noisy image is to the left, a draw from the Ising model is to the right.

▸ ANIMATION

Metropolis-Hastings

An important alternative to the Gibbs sampler, even more closely connected to detailed balance, is Metropolis-Hastings:

- Suppose that $X_n = x$.
- Pick y using a transition probability kernel $q(x, y)$ (the **proposal kernel**).
- Accept the proposed transition $x \to y$ with probability

$$\alpha(x, y) = \min\left\{1, \frac{\pi(y)q(y, x)}{\pi(x)q(x, y)}\right\} .$$

- If the transition is accepted, set $X_{n+1} = y$; otherwise set $X_{n+1} = x$.

Since π satisfies detailed balance, π is an equilibrium distribution (if the chain converges to a unique equilibrium!).

Renewal processes and stationarity

Q: How many statisticians does it take to change a lightbulb?
A: This should be determined using a nonparametric procedure,
since statisticians are not normal.

Stopping times

Let $(X_n)_{n\geq 0}$ be a stochastic process and let us write $\mathcal{F}_n$ for the collection of events "which can be determined from $X_0, X_1, \ldots, X_n$." For example,

$$\left\{\min_{0\leq k\leq n} X_k = 5\right\} \in \mathcal{F}_n$$

but

$$\left\{\min_{0\leq k\leq n+1} X_k = 5\right\} \notin \mathcal{F}_n.$$

Definition

A random variable T taking values in $\{0,1,2,\ldots\}\cup\{\infty\}$ is said to be a **stopping time** (for the process X) if, for all n, $\{T\leq n\}$ is determined by the information available at time n i.e. $\{T\leq n\}\in\mathcal{F}_n$.

Random walk example

Let X be a random walk begun at 0.

- The random time $T = \inf\{n > 0 : X_n \geq 10\}$ is a stopping time.
- Indeed $\{T \leq n\}$ is clearly determined by the information available at time n:

$$\{T \leq n\} = \{X_1 \geq 10\} \cup \ldots \cup \{X_n \geq 10\}\,.$$

- On the other hand, the random time $S = \sup\{0 \leq n \leq 100 : X_n \geq 10\}$ is not a stopping time.

Note that the minimum of two stopping times is a stopping time!

Strong Markov property

Suppose that T is a stopping time for the Markov chain $(X_n)_{n\geq 0}$.

Theorem

Conditionally on $\{T < \infty\}$ and $X_T = i$, $(X_{T+n})_{n\geq 0}$ has the same distribution as $(X_n)_{n\geq 0}$ started from $X_0 = i$. Moreover, given $\{T < \infty\}$, $(X_{T+n})_{n\geq 0}$ and $(X_n)_{0\leq n<T}$ are ***conditionally independent*** *given X_T.*

This is called the **strong** Markov property.

Hitting times and the Strong Markov property

Consider an irreducible recurrent Markov chain on a discrete state-space S. Fix $i \in S$ and let

$$H_0^{(i)} = \inf\{n \geq 0 : X_n = i\}.$$

For $m \geq 0$, recursively let

$$H_{m+1}^{(i)} = \inf\{n > H_m^{(i)} : X_n = i\}.$$

It follows from the strong Markov property that the random variables

$$H_{m+1}^{(i)} - H_m^{(i)}, m \geq 0$$

are **independent and identically distributed** and also independent of $H_0^{(i)}$.

Suppose we start our Markov chain from $X_0 = i$. Then $H_0^{(i)} = 0$.

Consider the number of visits to state i which have occurred by time n (not including the starting point!) i.e.

$$N^{(i)}(n) = \#\left\{k \geq 1 : H_k^{(i)} \leq n\right\}.$$

This is an example of a **renewal process**.

Renewal processes

Definition

Let $Z_1, Z_2, \ldots$ be i.i.d. integer-valued random variables such that $\mathbb{P}[Z_1 > 0] = 1$. Let $T_0 = 0$ and, for $k \geq 1$, let

$$T_k = \sum_{j=1}^{k} Z_j$$

and, for $n \geq 0$,

$$N(n) = \#\{k \geq 1 : T_k \leq n\}.$$

Then $(N(n))_{n \geq 0}$ is a (discrete) **renewal process**.

Example

Suppose that $Z_1, Z_2, \ldots$ are i.i.d. Geom(p) i.e.

$$\mathbb{P}\left[Z_1 = k\right] = (1-p)^{k-1}p, \quad k \geq 1.$$

Then we can think of Z_1 as the number of independent coin tosses required to first see a head, if heads has probability p.

So $N(n)$ has the same distribution as the number of heads in n independent coin tosses i.e. $N(n) \sim \text{Bin}(n, p)$ and, moreover,

$$\begin{aligned}&\mathbb{P}\left[N(k+1) = n_k + 1 | N(0) = n_0, N(1) = n_1, \ldots, N(k) = n_k\right] \\ &= \mathbb{P}\left[N(k+1) = n_k + 1 | N(k) = n_k\right] = p\end{aligned}$$

and

$$\begin{aligned}&\mathbb{P}\left[N(k+1) = n_k | N(0) = n_0, N(1) = n_1, \ldots, N(k) = n_k\right] \\ &= \mathbb{P}\left[N(k+1) = n_k | N(k) = n_k\right] = 1 - p.\end{aligned}$$

So, in this case, $(N(n))_{n \geq 0}$ is a Markov chain.

Renewal processes are not normally Markov...

The example on the previous slide is essentially the only example of a discrete renewal process which is Markov.

Why? Because the geometric distribution has the memoryless property:

$$\mathbb{P}\left[Z_1 - r = k | Z_1 > r\right] = (1-p)^{k-1} p, \quad k \geq 1.$$

So, regardless of what I know about the process up until the present time, the distribution of the remaining time until the next renewal is again geometric. The geometric is the only discrete distribution with this property.

Delayed renewal processes

Definition

Let Z_0 be a non-negative integer-valued random variable and, independently, let $Z_1, Z_2, \ldots$ be independent strictly positive and identically distributed integer-valued random variables.
For $k \geq 0$, let

$$T_k = \sum_{j=0}^{k} Z_j$$

and, for $n \geq 0$,

$$N(n) = \#\{k \geq 0 : T_k \leq n\}.$$

Then $(N(n))_{n\geq 0}$ is a (discrete) **delayed renewal process**, with delay Z_0.

Strong law of large numbers

Suppose that $\mu := \mathbb{E}[Z_1] < \infty$. Then the SLLN tells us that

$$\frac{T_k}{k} = \frac{1}{k}\sum_{j=0}^{k} Z_j \to \mu \text{ a.s. as } k \to \infty.$$

One can use this to show that

$$\frac{N(n)}{n} \to \frac{1}{\mu} \text{ a.s. as } n \to \infty$$

which tells us that we see renewals at a long-run average rate of $1/\mu$.

Probability of a renewal

Think back to our motivating example of hitting times of state i for a Markov chain. Suppose we want to think in terms of convergence to equilibrium: we would like to know what is the probability that at some large time n there is a renewal (i.e. a visit to i). We have $N(n) \approx n/\mu$ for large n (where μ is the expected return time to i), so as long as renewals are evenly spread out, the probability of a renewal at a particular large time should look like $1/\mu$.

This intuition turns out to be correct as long as every sufficiently large integer time is a possible renewal time. In particular, let

$$d = \gcd\{n : \mathbb{P}[Z_1 = n] > 0\}.$$

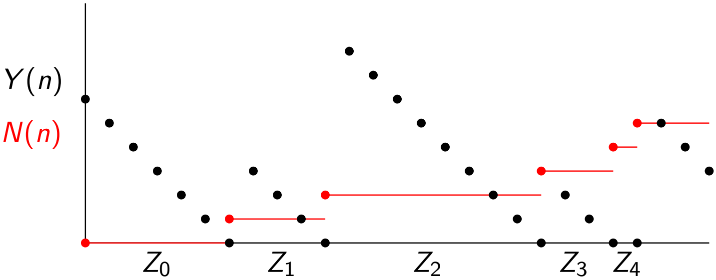
If $d = 1$ then this is fine; if we are interpreting renewals as returns to i for our Markov chain, this says that the chain is **aperiodic**.

An auxiliary Markov chain

We saw that a delayed renewal process $(N(n))_{n\geq 0}$ is not normally itself Markov. But we can find an auxiliary process which is. For $n \geq 0$, let

$$Y(n) := T_{N(n-1)} - n.$$

This is the time until the next renewal.



For $n \geq 0$,

$$Y(n) := T_{N(n-1)} - n.$$

$(Y(n))_{n \geq 0}$ has very simple transition probabilities: if $k \geq 1$ then

$$\mathbb{P}\left[Y(n+1) = k - 1 | Y(n) = k\right] = 1$$

and

$$\mathbb{P}\left[Y(n+1) = i | Y(n) = 0\right] = \mathbb{P}\left[Z_1 = i + 1\right] \text{ for } i \geq 0.$$

A stationary version

Recall that $\mu = \mathbb{E}[Z_1]$. Then the stationary distribution for this auxiliary Markov chain is

$$\nu_i = \frac{1}{\mu} \mathbb{P}[Z_1 \geq i+1], \quad i \geq 0.$$

If we start a delayed renewal process $(N(n))_{n\geq 0}$ with $Z_0 \sim \nu$ then the time until the next renewal is always distributed as ν. We call such a delayed renewal process **stationary**.

Notice that the stationary probability of being at a renewal time is $\nu_0 = 1/\mu$.

Size-biasing and inter-renewal intervals

The stationary distribution

$$\nu_i = \frac{1}{\mu} \mathbb{P}\left[Z_1 \geq i+1\right], \quad i \geq 0$$

has an interesting interpretation.

Let Z^* be a random variable with probability mass function

$$\mathbb{P}\left[Z^* = i\right] = \frac{i \, \mathbb{P}\left[Z_1 = i\right]}{\mu}, \qquad i \geq 1.$$

We say that Z^* has the **size-biased distribution** associated with the distribution of Z_1.

Now, conditionally on $Z^* = k$, let $L \sim \mathrm{U}\{0, 1, \ldots, k-1\}$. Then (unconditionally), $L \sim \nu$.

Interpretation

We are looking at a large time n and want to know how much time there is until the next renewal. Intuitively, n has more chance to fall in a longer interval. Indeed, it is i times more likely to fall in an interval of length i than an interval of length 1. So the inter-renewal time that n falls into is **size-biased**. Again intuitively, it is equally likely to be at any position inside that renewal interval, and so the time until the next renewal should be uniform on $\{0, 1, \ldots, Z^* - 1\}$ i.e. it should have the same distribution as L.

Convergence to stationarity

Theorem (Blackwell's renewal theorem)

Suppose that the distribution of Z_1 in a delayed renewal process is such that $\operatorname{gcd}\{n : \mathbb{P}[Z_1 = n] > 0\} = 1$ *and* $\mu := \mathbb{E}[Z_1] < \infty$*. Then*

$$\mathbb{P}[\textit{renewal at time } n] = \mathbb{P}[Y(n) = 0] \to \frac{1}{\mu}$$

as $n \to \infty$.

The coupling approach to the proof

Let Z_0 have a general delay distribution and let $\tilde{Z}_0 \sim \nu$ independently. Let N and $\tilde{N}$ be independent delayed renewal processes with these delay distributions and inter-renewal times $Z_1, Z_2, \ldots$ and $\tilde{Z}_1, \tilde{Z}_2, \ldots$ respectively, all i.i.d. random variables. Let

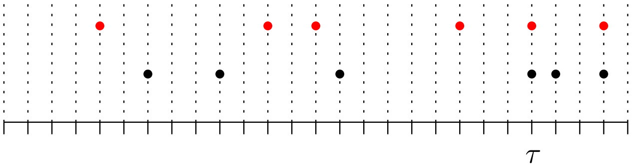
$$I(n) = \mathbb{1}_{\{N \text{ has a renewal at } n\}},$$
$$\tilde{I}(n) = \mathbb{1}_{\left\{\tilde{N} \text{ has a renewal at } n\right\}}.$$

Finally, let

$$\tau = \inf\{n \geq 0 : I(n) = \tilde{I}(n) = 1\}.$$

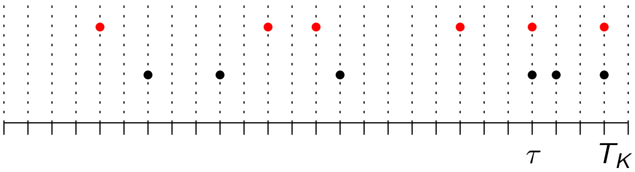
We have

$$\tau = \inf\{n \geq 0 : I(n) = \tilde{I}(n) = 1\}.$$



We argue that $\tau < \infty$ almost surely in the case where $\{n : \mathbb{P}[Z_1 = n] > 0\} \not\subseteq a + m\mathbb{Z}$ for any integers $a \geq 0$, $m \geq 2$. (In the general case, it is necessary to adapt the definition of $I(n)$ appropriately).

The coupling approach



τ is certainly smaller than T_K, where

$$K = \inf\{k \geq 0 : T_k = \tilde{T}_k\} = \inf\{k \geq 0 : T_k - \tilde{T}_k = 0\}.$$

But $T_k - \tilde{T}_k = Z_0 - \tilde{Z}_0 + \sum_{i=1}^{k}(Z_i - \tilde{Z}_i)$ and so $(T_k - \tilde{T}_k)_{k\geq 0}$ is a random walk with zero-mean step-sizes (such that, for all $m \in \mathbb{Z}$, $\mathbb{P}\left[T_k - \tilde{T}_k = m\right] > 0$ for large enough k) started from $Z_0 - \tilde{Z}_0 < \infty$. In particular, it is recurrent and so $K < \infty$, which implies that $T_K < \infty$.

The coupling approach

Now let

$$I^*(n) = \begin{cases} I(n) & \text{for } n \leq \tau \\ \tilde{I}(n) & \text{for } n > \tau. \end{cases}$$

Then $(I^*(n))_{n\geq 0}$ has the same distribution as $(I(n))_{n\geq 0}$. Moreover, $\mathbb{P}\left[I^*(n) = 1 | \tau < n\right] = \mathbb{P}\left[\tilde{I}(n) = 1\right] = \frac{1}{\mu}$ and so

$$\begin{aligned}
&\left|\mathbb{P}\left[I(n) = 1\right] - \frac{1}{\mu}\right| = \left|\mathbb{P}\left[I^*(n) = 1\right] - \frac{1}{\mu}\right| \\
&= \left|\mathbb{P}\left[I^*(n) = 1 | \tau < n\right] \mathbb{P}\left[\tau < n\right] + \mathbb{P}\left[I^*(n) = 1 | \tau \geq n\right] \mathbb{P}\left[\tau \geq n\right] - \frac{1}{\mu}\right| \\
&= \left|\mathbb{P}\left[I^*(n) = 1 | \tau \geq n\right] - \frac{1}{\mu}\right| \mathbb{P}\left[\tau \geq n\right] \\
&\leq \mathbb{P}\left[\tau \geq n\right] \to 0 \text{ as } n \to \infty.
\end{aligned}$$

Convergence to stationarity

We have proved:

Theorem (Blackwell's renewal theorem)

Suppose that the distribution of Z_1 in a delayed renewal process is such that $\gcd\{n : \mathbb{P}[Z_1 = n] > 0\} = 1$ *and* $\mu := \mathbb{E}[Z_1] < \infty$*. Then*

$$\mathbb{P}[\textit{renewal at time } n] \to \frac{1}{\mu}$$

as $n \to \infty$*.*

Convergence to stationarity

We can straightforwardly deduce the usual convergence to stationarity for a Markov chain.

Theorem

Let X be an irreducible, aperiodic, positive recurrent Markov chain (i.e. $\mu_i = \mathbb{E}\left[H_1^{(i)} - H_0^{(i)}\right] < \infty$). Then, whatever the distribution of X_0,

$$\mathbb{P}[X_n = i] \to \frac{1}{\mu_i}$$

as $n \to \infty$.

Note the interpretation of the stationary probability of being in state i as the inverse of the mean return time to i.

Decomposing a Markov chain

Consider an irreducible, aperiodic, positive recurrent Markov chain X, fix a reference state α and let $H_m = H_m^{(\alpha)}$ for all $m \geq 0$.

Recall that $(H_{m+1} - H_m, m \geq 0)$ is a collection of i.i.d. random variables, by the Strong Markov property.

More generally, it follows that the collection of **pairs**

$$\left(H_{m+1} - H_m, (X_{H_m+n})_{0 \leq n \leq H_{m+1}-H_m}\right), m \geq 0,$$

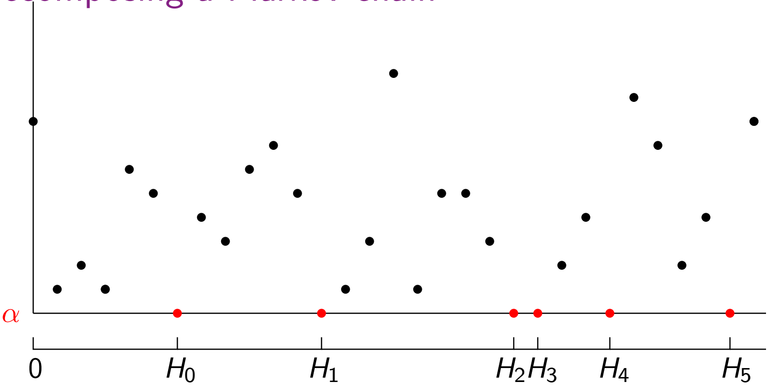
(where the first element of the pair is the time between the mth and $(m+1)$st visits to α, and the second element is a path which starts and ends at α and doesn't touch α in between) are **independent and identically distributed**.

Decomposing a Markov chain

Conditionally on $H_{m+1} - H_m = k$, $(X_{H_m+n})_{0\leq n\leq k}$ has the same distribution as the Markov chain X started from α and conditioned to first return to α at time k.

So we can split the path of a recurrent Markov chain into independent chunks ("excursions"), between successive visits to α. The renewal process of times when we visit α becomes stationary. To get back the whole Markov chain, we just need to "paste in" pieces of conditioned path.

Decomposing a Markov chain



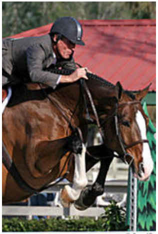
Essentially the same picture will hold true when we come to consider general state-space Markov chains in the last three lectures.

Martingales

"One of these days . . . a guy is going to come up to you and show you a nice brand-new deck of cards on which the seal is not yet broken, and this guy is going to offer to bet you that he can make the Jack of Spades jump out of the deck and squirt cider in your ear. But, son, do not bet this man, for as sure as you are standing there, you are going to end up with an earful of cider."

Frank Loesser,

Guys and Dolls musical, 1950, script



Martingales pervade modern probability

1. We say the random process $X = (X_n : n \geq 0)$ is a **martingale** if it satisfies the **martingale property**:

$$\mathbb{E}\left[X_{n+1} | X_n, X_{n-1}, \ldots\right] = \\ \mathbb{E}\left[X_n \text{ plus jump at time } n+1 | X_n, X_{n-1}, \ldots\right] = X_n\,.$$

2. Simplest possible example: simple symmetric random walk $X_0 = 0$, X_1, X_2, ... The martingale property follows from independence and distributional symmetry of jumps.
3. For convenience and brevity, we often replace $\mathbb{E}\left[\ldots | X_n, X_{n-1}, \ldots\right]$ by $\mathbb{E}\left[\ldots | \mathcal{F}_n\right]$ and think of "conditioning on $\mathcal{F}_n$" as "conditioning on all events which can be determined to have happened by time n".

Thackeray's martingale

1. MARTINGALE:
   - spar under the bowsprit of a sailboat;
   - a harness strap that connects the nose piece to the girth; prevents the horse from throwing back its head.
2. MARTINGALE in gambling:

   The original sense is given in the OED: "a system in gambling which consists in doubling the stake when losing in the hope of eventually recouping oneself." The oldest quotation is from 1815 but the nicest is from 1854: Thackeray in *The Newcomes* I. 266 "You have not played as yet? Do not do so; above all avoid a martingale if you do."
3. Result of playing Thackeray's martingale system and stopping on first win:
   ▸ ANIMATION
   set fortune at time n to be M_n.
   If $X_1 = -1, \ldots, X_n = -n$ then
   $M_n = -1 - 2 - \ldots - 2^{n-1} = 1 - 2^n$, otherwise $M_n = 1$.

Martingales and populations

1. Consider a branching process Y: population at time n is Y_n, where $Y_0 = 1$ (say) and Y_{n+1} is the sum $Z_{n+1,1} + \ldots + Z_{n+1,Y_n}$ of Y_n independent copies of a non-negative integer-valued **family-size r.v.** Z.
2. Suppose $\mathbb{E}[Z] = \mu < \infty$. Then $X_n = Y_n/\mu^n$ defines a martingale.
3. Suppose $\mathbb{E}\left[s^Z\right] = G(s)$. Let $H_n = Y_0 + \ldots + Y_n$ be total of all populations up to time n. Then $s^{H_n}/(G(s)^{H_{n-1}})$ defines a martingale.
4. If ζ is the smallest non-negative root of the equation $G(s) = s$, then ζ^{Y_n} defines a martingale.
5. In all these examples we can use $\mathbb{E}[\ldots|\mathcal{F}_n]$, representing conditioning by all $Z_{m,i}$ for $m \leq n$.

Definition of a martingale

Formally:

Definition

X is a **martingale** if $\mathbb{E}\left[|X_n|\right] < \infty$ (for all n) and

$$X_n = \mathbb{E}\left[X_{n+1} | \mathcal{F}_n\right] .$$

Supermartingales and submartingales

Two associated definitions.

Definition

$(X_n : n \geq 0)$ is a **supermartingale** if $\mathbb{E}[|X_n|] < \infty$ for all n and

$$X_n \geq \mathbb{E}[X_{n+1}|\mathcal{F}_n]$$

(and X_n forms part of conditioning expressed by $\mathcal{F}_n$).

Definition

$(X_n : n \geq 0)$ is a **submartingale** if $\mathbb{E}[|X_n|] < \infty$ for all n and

$$X_n \leq \mathbb{E}[X_{n+1}|\mathcal{F}_n]$$

(and X_n forms part of conditioning expressed by $\mathcal{F}_n$).

Examples of supermartingales and submartingales

1. Consider asymmetric simple random walk: supermartingale if jumps have negative expectation, submartingale if jumps have positive expectation.
2. This holds even if the walk is stopped on its first return to 0.
3. Consider Thackeray's martingale based on asymmetric random walk. This is a supermartingale or a submartingale depending on whether jumps have negative or positive expectation.
4. Consider the branching process (Y_n) and think about Y_n on its own instead of Y_n/μ^n. This is a supermartingale if $\mu < 1$ (sub-critical case), a submartingale if $\mu > 1$ (super-critical case), and a martingale if $\mu = 1$ (critical case).
5. By the conditional form of Jensen's inequality, if X is a martingale then $|X|$ is a submartingale.

More martingale examples

1. Repeatedly toss a coin, with probability of heads equal to p: each Head earns £1 and each Tail loses £1. Let X_n denote your fortune at time n, with $X_0 = 0$. Then
$$\left(\frac{1-p}{p}\right)^{X_n} \text{ defines a martingale.}$$
2. A shuffled pack of cards contains b black and r red cards. The pack is placed face down, and cards are turned over one at a time. Let B_n denote the number of black cards left *just before* the n^{th} card is turned over:
$$\frac{B_n}{r+b-(n-1)},$$
the proportion of black cards left just before the n^{th} card is revealed, defines a martingale.

An example of importance in finance

1. Suppose N_1, N_2, ... are independent identically distributed normal random variables of mean 0 and variance σ^2, and put $S_n = N_1 + \ldots + N_n$.
2. Then the following is a martingale:

$$Y_n = \exp\left(S_n - \frac{n}{2}\sigma^2\right) .$$

▸ ANIMATION

3. A modification exists for which the N_i have non-zero mean μ. **Hint:** $S_n \to S_n - n\mu$.

Martingales and likelihood

1. Suppose that a random variable X has a distribution which depends on a parameter θ. Independent copies X_1, X_2, ... of X are observed at times 1, 2, ... The likelihood of θ at time n is
$$L(\theta; X_1, \ldots, X_n) = p(X_1, \ldots, X_n | \theta)\,.$$
2. If θ_0 is the "true" value then (computing expectation with $\theta = \theta_0$)

$$\mathbb{E}\left[\left. \frac{L(\theta_1; X_1, \ldots, X_{n+1})}{L(\theta_0; X_1, \ldots, X_{n+1})} \right| \mathcal{F}_n \right] = \frac{L(\theta_1; X_1, \ldots, X_n)}{L(\theta_0; X_1, \ldots, X_n)}.$$

Martingales for Markov chains

To connect to the first theme of the course, Markov chains provide us with a large class of examples of martingales.

1. Let X be a Markov chain with countable state-space S and transition probabilities $p_{x,y}$. Let $f : S \to \mathbb{R}$ be any bounded function.
2. Take $\mathcal{F}_n$ to contain all the information about $X_0, X_1, \ldots, X_n$.
3. Then

$$M_n^f = f(X_n) - f(X_0) - \sum_{i=0}^{n-1} \left[\sum_{y \in S} (f(y) - f(X_i)) p_{X_i,y} \right]$$

   defines a martingale.
4. In fact, if M^f is a martingale for all bounded functions f then X is a Markov chain with transition probabilities $p_{x,y}$.

Martingales for Markov chains: harmonic functions

Call a function $f : S \to \mathbb{R}$ **harmonic** if

$$f(x) = \sum_{y \in S} f(y) p_{x,y} \text{ for all } x \in S.$$

We defined

$$M_n^f = f(X_n) - f(X_0) - \sum_{i=0}^{n-1} \left[\sum_{y \in S} (f(y) - f(X_i)) p_{X_i,y} \right]$$

and so we see that if f is harmonic then $f(X_n)$ is itself a martingale.

Martingale convergence

"Hurry please it's time."
T. S. Eliot,
The Waste Land, 1922

The martingale property at random times

The big idea

Martingales M stopped at "nice" times are still martingales. In particular, for a "nice" random T,

$$\mathbb{E}[M_T] = \mathbb{E}[M_0] .$$

For a random time T to be "nice", two things are required:

1. T must not "look ahead";
2. T must not be "too big". ▸ ANIMATION
3. Note that random times T turning up in practice often have positive chance of being infinite.

Stopping times

We have already seen what we mean by a random time "not looking ahead": such a time T is more properly called a *stopping time*.

Example

Let Y be a branching process of mean-family-size μ (recall that $X_n = Y_n/\mu^n$ determines a martingale), with $Y_0 = 1$.

- The random time $T = \inf\{n : Y_n = 0\} = \inf\{n : X_n = 0\}$ is a stopping time.
- Indeed $\{T \leq n\}$ is clearly determined by the information available at time n:

$$\{T \leq n\} = \{Y_n = 0\},$$

  since $Y_{n-1} = 0$ implies $Y_n = 0$ etc.

Stopping times aren't enough

However, even if T is a stopping time, we clearly need a stronger condition in order to say that $\mathbb{E}\left[M_T|\mathcal{F}_0\right] = M_0$.

e.g. let X be a random walk on $\mathbb{Z}$, started at 0.

- $T = \inf\{n > 0 : X_n \geq 10\}$ is a stopping time
- T is typically "too big": so long as it is almost surely finite, $X_T \geq 10$ and we deduce that $0 = \mathbb{E}\left[X_0\right] < \mathbb{E}\left[X_T\right]$.

Optional stopping theorem

Theorem

Suppose M is a martingale and T is a **bounded** *stopping time. Then*

$$\mathbb{E}\left[M_T|\mathcal{F}_0\right] = M_0\,.$$

We can generalize to general stopping times either if M is bounded or (more generally) if M is "uniformly integrable".

Gambling: you shouldn't expect to win

Suppose your fortune in a gambling game is X, a martingale begun at 0 (for example, a simple symmetric random walk). If N is the maximum time you can spend playing the game, and if $T \leq N$ is a **bounded** stopping time, then

$$\mathbb{E}[X_T] = 0\,.$$

▸ ANIMATION

Contrast Fleming (1953):

> *"Then the Englishman, Mister Bond, increased his winnings to exactly three million over the two days. He was playing a progressive system on red at table five. . . . It seems that he is persevering and plays in maximums. He has luck."*

Exit from an interval

Here's an elegant application of the optional stopping theorem.

- Suppose that X is a simple symmetric random walk started from 0. Then X is a martingale.
- Let $T = \inf\{n : X_n = a \text{ or } X_n = -b\}$. ($T$ is almost surely finite.) Suppose we want to find $\mathbb{P}[X \text{ hits } a \text{ before } -b] = \mathbb{P}[X_T = a]$.
- On the (random) time interval $[0, T]$, X is bounded, and so we can apply the optional stopping theorem to see that

$$\mathbb{E}[X_T] = \mathbb{E}[X_0] = 0.$$

- But then

$$\begin{aligned} 0 = \mathbb{E}[X_T] &= a\,\mathbb{P}[X_T = a] - b\,\mathbb{P}[X_T = -b] \\ &= a\,\mathbb{P}[X_T = a] - b(1 - \mathbb{P}[X_T = a]). \end{aligned}$$

Solving gives $\mathbb{P}[X \text{ hits } a \text{ before } -b] = \frac{b}{a+b}$.

Martingales and hitting times

Suppose that X_1, X_2, ... are i.i.d. $N(-\mu, 1)$ random variables, where $\mu > 0$. Let $S_n = X_1 + \ldots + X_n$ and let T be the time when S first exceeds level $\ell > 0$.

Then $\exp\left(\alpha(S_n + \mu n) - \frac{\alpha^2}{2} n\right)$ determines a martingale (for any $\alpha \geq 0$), and the optional stopping theorem can be applied to show

$$\mathbb{E}\left[\exp\left(-pT\right)\right] \sim e^{-(\mu+\sqrt{\mu^2+2p})\ell}, \quad p > 0.$$

This can be improved to an equality, at the expense of using more advanced theory, if we replace the Gaussian random walk S by Brownian motion.

Martingale convergence

Theorem

Suppose X is a non-negative supermartingale. Then there exists a random variable Z such that $X_n \to Z$ a.s. and, moreover, $\mathbb{E}\left[Z|\mathcal{F}_n\right] \leq X_n$.

▸ ANIMATION

Theorem

Suppose X is a bounded martingale (or, more generally, uniformly integrable). Then $Z = \lim_{n\to\infty} X_n$ exists a.s. and, moreover, $\mathbb{E}\left[Z|\mathcal{F}_n\right] = X_n$.

Theorem

Suppose X is a martingale and $\mathbb{E}\left[X_n^2\right] \leq K$ for some fixed constant K. Then one can prove directly that $Z = \lim_{n\to\infty} X_n$ exists a.s. and, moreover, $\mathbb{E}\left[Z|\mathcal{F}_n\right] = X_n$.

Birth-death process

Suppose Y is a discrete-time birth-and-death process started at $y > 0$ and *absorbed at zero*:

$$p_{k,k+1} = \frac{\lambda}{\lambda+\mu}, \quad p_{k,k-1} = \frac{\mu}{\lambda+\mu}, \quad \text{for } k > 0, \text{ with } 0 < \lambda < \mu.$$

Y is a non-negative supermartingale and so $\lim_{n\to\infty} Y_n$ exists.
Y is a biased random walk with a single absorbing state at 0. Let $T = \inf\{n : Y_n = 0\}$; then $T < \infty$ a.s. and so the only possible limit for Y is 0.

Birth-death process

Now let

$$X_n = Y_{n\wedge T} + \left(\frac{\mu-\lambda}{\mu+\lambda}\right)(n \wedge T).$$

This is a non-negative martingale converging to $Z = \frac{\mu-\lambda}{\mu+\lambda} T$.

Thus, recalling that $Y_0 = X_0 = y$ and using the martingale convergence theorem,

$$\mathbb{E}\left[T\right] \leq \left(\frac{\mu+\lambda}{\mu-\lambda}\right) y\,.$$

Likelihood revisited

Suppose i.i.d. random variables X_1, X_2, ... are observed at times 1, 2, ..., and suppose the common density is $f(\theta; x)$. Suppose also that $\mathbb{E}\left[|\log(f(\theta; X_1))|\right] < \infty$. Recall that, if the "true" value of θ is θ_0, then

$$M_n = \frac{L(\theta_1; X_1, \ldots, X_n)}{L(\theta_0; X_1, \ldots, X_n)}$$

is a martingale, with $\mathbb{E}\left[M_n\right] = 1$ for all $n \geq 1$.

The SLLN and Jensen's inequality show that

$$\frac{1}{n} \log M_n \to -c \text{ as } n \to \infty \,,$$

moreover if $f(\theta_0; \cdot)$ and $f(\theta_1; \cdot)$ differ as densities then $c > 0$, and so $M_n \to 0$.

Sequential hypothesis testing

In the setting above, suppose that we want to satisfy

$$\mathbb{P}\left[\text{reject } H_0 | H_0\right] \leq \alpha \quad \text{and} \quad \mathbb{P}\left[\text{reject } H_1 | H_1\right] \leq \beta\,.$$

How large a sample size do we need?
Let

$$T = \inf\{n : M_n \geq \alpha^{-1} \text{ or } M_n \leq \beta\}$$

and consider observing $X_1, \ldots, X_T$ and then rejecting H_0 iff $M_T \geq \alpha^{-1}$.

Sequential hypothesis testing continued

On the (random) time interval $[0, T]$, M is a bounded martingale, and so

$$\mathbb{E}\left[M_T\right] = \mathbb{E}\left[M_0\right] = 1$$

(where we are computing the expectation using $\theta = \theta_0$). So

$$1 = \mathbb{E}\left[M_T\right] \geq \alpha^{-1}\,\mathbb{P}\left[M_T \geq \alpha^{-1} \,|\, \theta_0\right] = \alpha^{-1}\,\mathbb{P}\left[\text{reject } H_0 | H_0\right] .$$

Interchanging the roles of H_0 and H_1 we also obtain $\mathbb{P}\left[\text{reject } H_1 | H_1\right] \leq \beta$.

The attraction here is that on average, fewer observations are needed than for a fixed-sample-size test.

Recurrence

"A bad penny always turns up"
Old English proverb.

Motivation from MCMC

Given a probability density $p(x)$ of interest, for example a Bayesian posterior, we could address the question of drawing from $p(x)$ by using, for example, Gaussian random-walk Metropolis-Hastings:

- Proposals are normal, with mean given by the current location x, and fixed variance-covariance matrix.
- We use the Hastings ratio to accept/reject proposals.
- We end up with a Markov chain X which has a transition mechanism which mixes a density with staying at the starting point.

Evidently, the chain almost surely *never* visits specified points other than its starting point. Thus, it can never be irreducible in the classical sense, and the discrete state-space theory cannot apply.

Recurrence

We already know that if X is a Markov chain on a discrete state-space then its transition probabilities converge to a unique limiting equilibrium distribution if:

1. X is irreducible;
2. X is aperiodic;
3. X is positive-recurrent.

In this case, we call the chain **ergodic**.

What can we say quantitatively, in general, about the speed at which convergence to equilibrium occurs? And what if the state-space is not discrete?

Measuring speed of convergence to equilibrium (I)

- The speed of convergence of a Markov chain X to equilibrium can be measured as discrepancy between two probability measures: $\mathcal{L}(X_n|X_0 = x)$ (the distribution of X_n) and π (the equilibrium distribution).
- Simple possibility: **total variation distance**. Let $\mathcal{X}$ be the state-space. For $A \subseteq \mathcal{X}$, find the maximum discrepancy between $\mathcal{L}(X_n|X_0 = x)(A) = \mathbb{P}[X_n \in A|X_0 = x]$ and $\pi(A)$:

$$\text{dist}_{\mathrm{TV}}(\mathcal{L}(X_n|X_0 = x), \pi) = \sup_{A \subseteq \mathcal{X}} \{\mathbb{P}[X_n \in A|X_0 = x] - \pi(A)\}.$$

- Alternative expression in the case of a discrete state-space:

$$\text{dist}_{\mathrm{TV}}(\mathcal{L}(X_n|X_0 = x), \pi) = \tfrac{1}{2} \sum_{y \in \mathcal{X}} |\mathbb{P}[X_n = y|X_0 = x] - \pi_y|.$$

(There are *many* other possible measures of distance ...)

Measuring speed of convergence to equilibrium (II)

Definition

The Markov chain X is **uniformly ergodic** if its distribution converges to equilibrium in total variation *uniformly in the starting point* $X_0 = x$: for some fixed $C > 0$ and for fixed $\gamma \in (0, 1)$,

$$\sup_{x \in \mathcal{X}} \operatorname{dist}_{\mathrm{TV}}(\mathcal{L}\left(X_n | X_0 = x\right), \pi) \leq C\gamma^n \,.$$

In theoretical terms, for example when carrying out MCMC, this is a very satisfactory property. No account need be taken of the starting point, and accuracy improves in proportion to the length of the simulation.

Measuring speed of convergence to equilibrium (III)

Definition

The Markov chain X is **geometrically ergodic** if its distribution converges to equilibrium in total variation for some $C(x) > 0$ *depending on the starting point* x and for fixed $\gamma \in (0, 1)$,

$$\text{dist}_{\mathrm{TV}}(\mathcal{L}\left(X_n | X_0 = x\right), \pi) \leq C(x)\gamma^n \,.$$

Here, account does need to be taken of the starting point, but still accuracy improves in proportion to the length of the simulation.

ϕ-irreducibility (I)

We make two observations about Markov chain irreducibility:

1. The discrete theory fails to apply directly even to well-behaved chains on non-discrete state-spaces.
2. Suppose ϕ is a measure on the state-space: then we could ask for the chain to be irreducible *on sets of positive ϕ-measure*.

Definition

The Markov chain X is ϕ-**irreducible** if for any state x and for any subset B of the state-space which is such that $\phi(B) > 0$, we find that X has positive chance of reaching B if begun at x.

(That is, if $T_B = \inf\{n \geq 1 : X_n \in B\}$ then if $\phi(B) > 0$ we have $\mathbb{P}\left[T_B < \infty | X_0 = x\right] > 0$ for all x.)

ϕ-irreducibility (II)

1. We call ϕ an **irreducibility measure**. It is possible to modify ϕ to construct a **maximal irreducibility measure** ψ; one such that any set B of positive measure under some irreducibility measure for X is of positive measure for ψ.
2. Irreducible chains on countable state-space are c-irreducible where c is counting measure ($c(A) = |A|$).
3. If a chain has unique equilibrium measure π then π will serve as a maximal irreducibility measure.

Regeneration and small sets (I)

The discrete-state-space theory works because (a) the Markov chain regenerates each time it visits individual states, and (b) it has a positive chance of visiting specified individual states.

In effect, this reduces the theory of convergence to a question about renewal processes, with renewals occurring each time the chain visits a specified state.

We want to extend this idea by thinking in terms of renewals when visiting sets instead.

Regeneration and small sets (II)

Definition

A set E of positive ϕ-measure is a **small set of lag** k for X if there is $\alpha \in (0,1)$ and a probability measure ν such that **for all** $x \in E$ the following **minorisation condition** is satisfied

$$\mathbb{P}\left[X_k \in A | X_0 = x\right] \geq \alpha \nu(A) \qquad \text{for all } A\,.$$

Regeneration and small sets (III)

Why is this useful? Consider a small set E of lag 1, so that for $x \in E$,

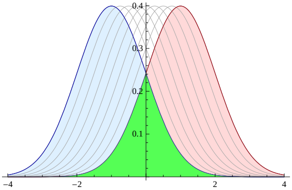
$$p(x, A) = \mathbb{P}\left[X_1 \in A | X_0 = x\right] \geq \alpha\nu(A) \quad \text{for all } A.$$

This means that, given $X_0 = x$, we can think of sampling X_1 as a two-step procedure. With probability α, sample X_1 from ν. With probability $1 - \alpha$, sample X_1 from the probability distribution $\frac{p(x,\cdot)-\alpha\nu(\cdot)}{1-\alpha}$.

For a small set of lag k, we can interpret this as follows: if we sub-sample X every k time-steps then, every time it visits E, there is probability α that X forgets its entire past and starts again, using probability measure ν.

Regeneration and small sets (IV)

Consider the Gaussian random walk described above. **Any** bounded set is small of lag 1. For example, consider the set $E = [-2, 2]$.

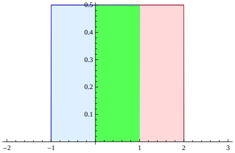


The green region represents the overlap of all the Gaussian densities centred at all points in E. Let α be the area of the green region and let f be its upper boundary. Then $f(x)/\alpha$ is a probability density and, for any $x \in E$,

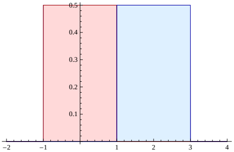
$$\mathbb{P}\left[X_1 \in A | X_0 = x\right] \geq \alpha \int_A \frac{f(x)}{\alpha} dx = \alpha\nu(A).$$

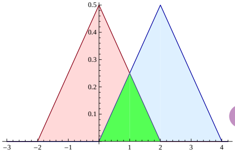
Regeneration and small sets (V)

Let X be a RW with transition density $p(x, \mathrm{d}\, y) = \frac{1}{2}\, \mathbb{1}_{\{|x-y|<1\}}$. Consider the set $[0,1]$: this is small of lag 1, with $\alpha = 1/2$ and ν the uniform distribution on $[0,1]$.



The set $[0,2]$ is *not* small of lag 1, but *is* small of lag 2.





ANIMATION

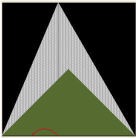
Regeneration and small sets (VI)

Small sets would not be very interesting except that:

1. All ϕ-irreducible Markov chains X possess small sets;
2. Consider chains X with continuous transition density kernels. They possess *many* small sets of lag 1;
3. Consider chains X with measurable transition density kernels. They need possess *no* small sets of lag 1, but will possess many sets of lag 2;
4. Given just one small set, X can be represented using a chain which has a single recurrent atom.

In a word, small sets **discretize** Markov chains.

Animated example: a random walk on $[0, 1]$



▸ ANIMATION

Transition density $p(x, y) = 2 \min\{\frac{y}{x}, \frac{1-y}{1-x}\}$.
Detailed balance equations (in terms of densities):

$$\pi(x)p(x, y) = \pi(y)p(y, x)$$

Spot an invariant probability density: $\pi(x) = 6x(1 - x)$.
For any $A \subset [0, 1]$ and all $x \in [0, 1]$,

$$\mathbb{P}\left[X_1 \in A | X_0 = x\right] \geq \frac{1}{2}\nu(A),$$

where $\nu(A) = 2 \int_A \min\{x, 1 - x\} \, \mathrm{d}\, x$. Hence, the whole state-space is small.

Regeneration and small sets (VII)

Here is an indication of how we can use the discretization provided by small sets.

Theorem

Suppose that π is a stationary distribution for X. Suppose that the whole state-space $\mathcal{X}$ is a small set of lag 1 i.e. there exists a probability measure ν and $\alpha \in (0,1)$ such that

$$\mathbb{P}\left[X_1 \in A | X_0 = x\right] \geq \alpha\nu(A) \textit{ for all } x \in \mathcal{X}.$$

Then

$$\sup_{x\in\mathcal{X}} \mathit{dist}_{\mathrm{TV}}(\mathcal{L}(X_n | X_0 = x), \pi) \leq (1-\alpha)^n$$

and so X is uniformly ergodic.

ANIMATION

Harris-recurrence

This motivates what we should mean by recurrence for non-discrete state spaces. Suppose X is ϕ-irreducible and ϕ is a maximal irreducibility measure.

Definition

X is **(ϕ-)recurrent** if, for ϕ-almost all starting points x and any subset B with $\phi(B) > 0$, when started at x the chain X hits B eventually with probability 1.

Definition

X is **Harris-recurrent** if we can drop "ϕ-almost" in the above.

Small sets and ϕ-recurrence

Small sets help us to identify when a chain is ϕ-recurrent:

Theorem

Suppose that X is ϕ-irreducible (and aperiodic). If there exists a small set C such that for all $x \in C$

$$\mathbb{P}\left[T_C < \infty | X_0 = x\right] = 1\,,$$

then X is ϕ-recurrent.

Example

- Random walk on $[0, \infty)$ given by $X_{n+1} = \max\{X_n + Z_{n+1}, 0\}$, where increments Z have negative mean.
- The Metropolis-Hastings algorithm on $\mathbb{R}$ with $N(0, \sigma^2)$ proposals.

Foster-Lyapunov criteria

"Even for the physicist the description in plain language will be the criterion of the degree of understanding that has been reached."
Werner Heisenberg,
Physics and philosophy:
The revolution in modern science, 1958

From this morning

Let X be a Markov chain and let $T_B = \inf\{n \geq 1 : X_n \in B\}$. Let ϕ be a measure on the state-space.

- X is ϕ-irreducible if $\mathbb{P}[T_B < \infty | X_0 = x] > 0$ for all x whenever $\phi(B) > 0$.
- A set E of positive ϕ-measure is a small set of lag k for X if there is $\alpha \in (0, 1)$ and a probability measure ν such that for all $x \in E$,

$$\mathbb{P}[X_k \in A | X_0 = x] \geq \alpha\nu(A) \quad \text{for all } A\,.$$

- All ϕ-irreducible Markov chains possess small sets.
- X is ϕ-recurrent if, for ϕ-almost all starting points x, $\mathbb{P}[T_B < \infty | X_0 = x] = 1$ whenever $\phi(B) > 0$.

Renewal and regeneration

Suppose C is a small set for ϕ-recurrent X, with lag 1: for $x \in C$,

$$\mathbb{P}\left[X_1 \in A | X_0 = x\right] \geq \alpha\nu(A)\,.$$

Identify **regeneration events**: X regenerates at $x \in C$ with probability α and then makes a transition with distribution ν; otherwise it makes a transition with distribution $\frac{p(x,\cdot)-\alpha\nu(\cdot)}{1-\alpha}$.

The regeneration events occur as a **renewal sequence**. Set

$$p_k \;=\; \mathbb{P}\left[\text{next regeneration at time } k \;\middle|\; \text{regeneration at time } 0\right].$$

If the renewal sequence is **non-defective** (i.e. $\sum_k p_k = 1$) and **positive-recurrent** (i.e. $\sum_k k p_k < \infty$) then there exists a stationary version. This is the key to equilibrium theory whether for discrete or continuous state-space.

▸ ANIMATION

Positive recurrence

Here is the **Foster-Lyapunov criterion for positive recurrence** of a ϕ-irreducible Markov chain X on a state-space $\mathcal{X}$.

Theorem

Suppose that there exist a function $\Lambda : \mathcal{X} \to [0, \infty)$, positive constants a, b, c, and a small set $C = \{x : \Lambda(x) \leq c\} \subseteq \mathcal{X}$ such that

$$\mathbb{E}\left[\Lambda(X_{n+1}) | \mathcal{F}_n\right] \leq \Lambda(X_n) - a + b\, \mathbb{1}_{\{X_n \in C\}}\,.$$

Then $\mathbb{E}\left[T_A | X_0 = x\right] < \infty$ for any A such that $\phi(A) > 0$ and, moreover, X has an equilibrium distribution.

Sketch of proof

1. Suppose $X_0 \notin C$. Then $Y_n = \Lambda(X_n) + an$ is non-negative supermartingale up to time $T_C = \inf\{m \geq 1 : X_m \in C\}$: if $T_C > n$ then
$$\mathbb{E}\left[Y_{n+1} | \mathcal{F}_n\right] \leq (\Lambda(X_n) - a) + a(n+1) = Y_n\,.$$
Hence, $Y_{\min\{n, T_C\}}$ converges.
2. So $\mathbb{P}\left[T_C < \infty\right] = 1$ (otherwise $\Lambda(X_n) > c$, $Y_n > c + an$ and so $Y_n \to \infty$). Moreover, $\mathbb{E}\left[Y_{T_C} | X_0\right] \leq \Lambda(X_0)$ (martingale convergence theorem) so $a\,\mathbb{E}\left[T_C | X_0\right] \leq \Lambda(X_0)$.
3. Now use the finiteness of b to show that $\mathbb{E}\left[T^* | X_0\right] < \infty$, where T^* is the time of the first regeneration in C.
4. ϕ-irreducibility: X has a positive chance of hitting A between regenerations in C. Hence, $\mathbb{E}\left[T_A | X_0\right] < \infty$.

A converse

Suppose, on the other hand, that $\mathbb{E}\left[T_C | X_0 = x\right] < \infty$ for all starting points x, where C is some small set. The Foster-Lyapunov criterion for positive recurrence follows for $\Lambda(x) = \mathbb{E}\left[T_C | X_0 = x\right]$ as long as $\mathbb{E}\left[T_C | X_0 = x\right]$ is bounded for $x \in C$.

Example: general reflected random walk

Let

$$X_{n+1} = \max\{X_n + Z_{n+1}, 0\},$$

for $Z_1, Z_2, \ldots$ i.i.d. with continuous density $f(z)$, $\mathbb{E}[Z_1] < 0$ and $\mathbb{P}[Z_1 > 0] > 0$. Then

(a) X is Lebesgue-irreducible on $[0, \infty)$;

(b) Foster-Lyapunov criterion for positive recurrence applies.

Similar considerations often apply to Metropolis-Hastings Markov chains based on random walks.

Geometric ergodicity

Here is the **Foster-Lyapunov criterion for geometric ergodicity** of a ϕ-irreducible Markov chain X on a state-space $\mathcal{X}$.

Theorem

Suppose that there exist a function $\Lambda : \mathcal{X} \to [1, \infty)$, *positive constants* $\gamma \in (0, 1)$, b, $c \geq 1$, *and a small set* $C = \{x : \Lambda(x) \leq c\} \subseteq \mathcal{X}$ *with*

$$\mathbb{E}\left[\Lambda(X_{n+1}) | \mathcal{F}_n\right] \leq \gamma \Lambda(X_n) + b\, \mathbb{1}_{\{X_n \in C\}} \,.$$

Then $\mathbb{E}\left[\gamma^{-T_A} | X_0 = x\right] < \infty$ *for any* A *such that* $\phi(A) > 0$ *and, moreover (under suitable periodicity conditions),* X *is geometrically ergodic.*

Sketch of proof

1. Suppose $X_0 \notin C$. Then $Y_n = \Lambda(X_n)/\gamma^n$ defines non-negative supermartingale up to time T_C: if $T_C > n$ then

$$\mathbb{E}\left[Y_{n+1}|\mathcal{F}_n\right] \leq \gamma \times \Lambda(X_n)/\gamma^{n+1} = Y_n\,.$$

   Hence, $Y_{\min\{n,T_C\}}$ converges.
2. So $\mathbb{P}\left[T_C < \infty\right] = 1$ (otherwise $\Lambda(X) > c$ and so $Y_n > c/\gamma^n$ does not converge). Moreover, $\mathbb{E}\left[\gamma^{-T_C}|X_0\right] \leq \Lambda(X_0)$.
3. Finiteness of b shows that $\mathbb{E}\left[\gamma^{-T^*}|X_0\right] < \infty$, where T^* is the time of the first regeneration in C.
4. From ϕ-irreducibility there is a positive chance of hitting A between regenerations in C. Hence, $\mathbb{E}\left[\gamma^{-T_A}|X_0\right] < \infty$.

Two converses

Suppose, on the other hand, that $\mathbb{E}\left[\gamma^{-T_C} | X_0\right] < \infty$ for all starting points X_0 (and fixed $\gamma \in (0,1)$), where C is some small set and T_C is the first time for X to return to C. The Foster-Lyapunov criterion for geometric ergodicity then follows for $\Lambda(x) = \mathbb{E}\left[\gamma^{-T_C} | X_0 = x\right]$ as long as $\mathbb{E}\left[\gamma^{-T_C} | X_0 = x\right]$ is bounded for $x \in C$.

But more is true! Strikingly, for Harris-recurrent Markov chains the existence of a geometric Foster-Lyapunov condition is equivalent to the property of geometric ergodicity.

Uniform ergodicity follows if the function Λ is bounded above.

Example: reflected simple asymmetric random walk

Let

$$X_{n+1} = \max\{X_n + Z_{n+1}, 0\}\,,$$

for $Z_1, Z_2, \ldots$ i.i.d. such that
$\mathbb{P}\left[Z_1 = -1\right] = q = 1 - p = 1 - \mathbb{P}\left[Z_1 = +1\right] > \frac{1}{2}$.

(a) X is (counting-measure-) irreducible on non-negative integers;

(b) Foster-Lyapunov criterion for positive recurrence applies, using $\Lambda(x) = x$ and $C = \{0\}$:

$$\mathbb{E}\left[\Lambda(X_1) | X_0 = x_0\right] = \begin{cases} \Lambda(x_0) - (q-p) & \text{if } x_0 \notin C\,, \\ 0 + p & \text{if } x_0 \in C\,. \end{cases}$$

(c) Foster-Lyapunov criterion for geometric ergodicity applies, using $\Lambda(x) = e^{ax}$ and $C = \{0\} = \Lambda^{-1}(\{1\})$.

Cutoff

"I have this theory of convergence, that good things always happen with bad things."
Cameron Crowe, Say Anything film, 1989

Convergence: cutoff or geometric decay?

What we have so far said about convergence to equilibrium will have left the misleading impression that the distance from equilibrium for a Markov chain is characterized by a gentle and rather geometric decay.

It is true that this is typically the case after an extremely long time, and it can be the case over all time. However, it is entirely possible for "most" of the convergence to happen quite suddenly at a specific threshold.

The theory for this is developing fast, but many questions remain open. In this section we describe a a few interesting results, and look in detail at a specific easy example.

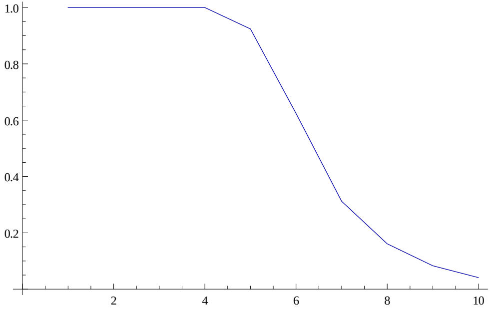
Cutoff: first example

Consider repeatedly shuffling a pack of n cards using a **riffle shuffle**.

Write P_n^t for the distribution of the cards at time t.
This shuffle can be viewed as a random walk on S_n with uniform equilibrium distribution π_n.

Cutoff: first example

With $n = 52$, the total variation distance $\mathrm{dist}_{\mathrm{TV}}(P_n^t, \pi_n)$ of P_n^t from equilibrium decreases like this:



Riffle shuffle: sharp result (Bayer and Diaconis 1992)

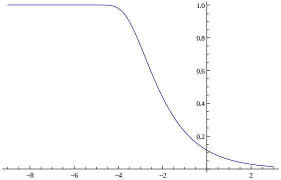
Let

$$\tau_n(\theta) = \frac{3}{2}\log_2 n + \theta\,.$$

Then

$$\mathrm{dist}_{\mathrm{TV}}(P_n^{\tau_n(\theta)}, \pi_n) = 1 - 2\Phi\left(\frac{-2^{-\theta}}{4\sqrt{3}}\right) + O(n^{-1/4})\,.$$

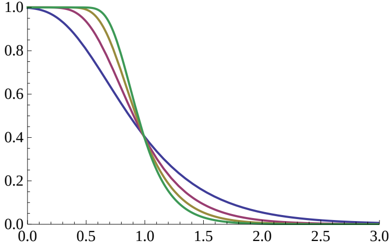
As a function of θ this looks something like:



So as n gets large, convergence to uniform happens quickly after about $(3/2)\log_2 n$ shuffles (≈ 7 when $n = 52$).

Cutoff: the general picture

Scaling the x-axis by the cutoff time, we see that the total variation distance drops more and more rapidly towards zero as n becomes larger: the curves in the graph below tend to a step function as $n \to \infty$.



Moral: effective convergence can be much faster than one realizes, and occur over a fairly well-defined period of time.

Cutoff: more examples

There are *many* examples of this type of behaviour:

$\mathcal{X}_n$	Chain	τ_n
S_n	Riffle shuffle	$\frac{3}{2}\log_2 n$
S_n	Top-to random	??
S_n	Random transpositions	??
$\mathbb{Z}_2^n$	Symmetric random walk	$\frac{1}{4} n \log n$

- Methods of proving cutoff include coupling theory, eigenvalue-analysis and group representation theory . . .

An example in more detail: the top-to-random shuffle

Let us show how to prove cutoff in a very simple case: the **top-to-random shuffle**. This is another random walk X on the symmetric group S_n: each 'shuffle' consists of removing the top card and replacing it into the pack uniformly at random.

Hopefully it's not too hard to believe that the equilibrium distribution of X is again the *uniform distribution* π_n on S_n (*i.e.* $\pi_n(\sigma) = 1/n!$ for all permutations $\sigma \in S_n$).

Theorem (Aldous & Diaconis (1986))

Let $\tau_n(\theta) = n \log n + \theta n$*. Then*

1. $\operatorname{dist}_{\mathrm{TV}}(P_n^{\tau_n(\theta)}, \pi_n) \leq e^{-\theta}$ *for* $\theta \geq 0$ *and* $n \geq 2$*;*
2. $\operatorname{dist}_{\mathrm{TV}}(P_n^{\tau_n(\theta)}, \pi_n) \to 1$ *as* $n \to \infty$*, for* $\theta = \theta(n) \to -\infty$*.*

Strong uniform times

Recall from lecture 2 that a **stopping time** is a non-negative integer-valued random variable T, with $\{T \leq k\} \in \mathcal{F}_k$ for all k. Let X be a random walk on a group G, with uniform equilibrium distribution π.

Definition

A **strong uniform time** T is a stopping time such that for each $k < \infty$ and $\sigma \in G$,

$$\mathbb{P}\left[X_k = \sigma \mid T = k\right] = \pi(\sigma) = 1/|G|\,.$$

Strong uniform times (SUT's) are useful for the following reason...

Lemma (Aldous & Diaconis (1986))

Let X be a random walk on a group G, with uniform stationary distribution π, and let T be a SUT for X. Then for all $k \geq 0$,

$$\operatorname{dist}_{\mathrm{TV}}(P^k, \pi) \leq \mathbb{P}\left[T > k\right] .$$

Proof.

For any set $A \subseteq G$,

$$\begin{aligned}
\mathbb{P}\left[X_k \in A\right] &= \sum_{j \leq k} \mathbb{P}\left[X_k \in A, T = j\right] + \mathbb{P}\left[X_k \in A, T > k\right] \\
&= \sum_{j \leq k} \pi(A)\, \mathbb{P}\left[T = j\right] + \mathbb{P}\left[X_k \in A \mid T > k\right] \mathbb{P}\left[T > k\right] \\
&= \pi(A) + \left(\mathbb{P}\left[X_k \in A \mid T > k\right] - \pi(A)\right) \mathbb{P}\left[T > k\right] .
\end{aligned}$$

So $|P^k(A) - \pi(A)| \leq \mathbb{P}\left[T > k\right]$. $\square$

Back to shuffling: the upper bound

Consider the card originally at the bottom of the deck (suppose for convenience that it's $Q\heartsuit$). Let

- T_1 = time until the 1st card is placed below $Q\heartsuit$;
- T_2 = time until a 2nd card is placed below $Q\heartsuit$;
- ...
- T_{n-1} = time until $Q\heartsuit$ reaches the top of the pack.

Then note that:

- at time T_2, the 2 cards below $Q\heartsuit$ are equally likely to be in either order;
- at time T_3, the 3 cards below $Q\heartsuit$ are equally likely to be in any order;
- ...

... so at time T_{n-1}, the $n-1$ cards below $Q\heartsuit$ are uniformly distributed.

Hence, at time $T = T_{n-1} + 1$, $Q\heartsuit$ is inserted uniformly at random, and now the cards are *all* uniformly distributed!

Since T is a SUT, we can use it in our Lemma to upper bound the total variation distance between π_n and the distribution of the pack at time k.

Note first of all that

$$T = T_1 + (T_2 - T_1) + \cdots + (T_{n-1} - T_{n-2}) + (T - T_{n-1}),$$

and that

$$T_{i+1} - T_i \overset{\text{ind}}{\sim} \text{Geom}\left(\frac{i+1}{n}\right).$$

We can find the distribution of T by turning to the **coupon collector's problem**. Consider a bag with n distinct balls - keep sampling (with replacement) until each ball has been seen at least once.

Let W_i = number of draws needed until i distinct balls have been seen. Then

$$W_n = (W_n - W_{n-1}) + (W_{n-1} - W_{n-2}) + \cdots + (W_2 - W_1) + W_1 \,,$$

where

$$W_{i+1} - W_i \overset{\text{ind}}{\sim} \text{Geom}\left(\frac{n-i}{n}\right) \,.$$

Thus, $T \overset{d}{=} W_n$.

Now let A_d be the event that ball d has not been seen in the first k draws.

$$\mathbb{P}\left[W_n > k\right] = \mathbb{P}\left[\cup_{d=1}^n A_d\right] \leq \sum_{d=1}^n \mathbb{P}\left[A_d\right]$$

$$= n\left(1 - \frac{1}{n}\right)^k \leq n e^{-k/n}.$$

Plugging in $k = \tau_n(\theta) = n \log n + \theta n$, we get

$$\mathbb{P}\left[W_n > \tau_n(\theta)\right] \leq e^{-\theta}.$$

Now use the fact that T and W_n have the same distribution, the important information that T is a SUT for the chain, and the Lemma above to deduce part 1 of our cutoff theorem.

The lower bound

To prove lower bounds of cutoffs, a frequent trick is to find a set B such that $|P_n^{\tau_n(\theta)}(B) - \pi_n(B)|$ is large, where $\tau_n(\theta)$ is now equal to $n \log n + \theta(n)n$, with $\theta(n) \to -\infty$.
So let

$$B_i = \{\sigma : \text{bottom } i \text{ original cards remain in original relative order}\}.$$

This satisfies $\pi_n(B_i) = 1/i!$. Furthermore, we can argue that, for any fixed i, with $\theta = \theta(n) \to -\infty$,

$$P_n^{\tau_n(\theta)}(B_i) \to 1 \quad \text{as } n \to \infty.$$

Therefore,

$$\operatorname{dist}_{\mathrm{TV}}(P_n^{\tau_n(\theta)}, \pi_n) \geq \max_i \left(P_n^{\tau_n(\theta)}(B_i) - \pi_n(B_i)\right) \to 1\,. \quad \square$$

Final comments...

So how does this shuffle compare to others?

$\mathcal{X}_n$	Chain	τ_n
S_n	Top-to random	$n \log n$
S_n	Riffle shuffle	$\frac{3}{2} \log_2 n$
S_n	Random transpositions	
S_n	Overhand shuffle	

Aldous, D. and P. Diaconis (1986).
Shuffling cards and stopping times.
The American Mathematical Monthly 93(5), 333–348.

Aldous, D. J. and J. A. Fill (2001).
Reversible Markov Chains and Random Walks on Graphs.
Unpublished.

Athreya, K. B. and P. Ney (1978).
A new approach to the limit theory of recurrent Markov chains.
Trans. Amer. Math. Soc. 245, 493–501.

Bayer, D. and P. Diaconis (1992).
Trailing the dovetail shuffle to its lair.
Ann. Appl. Probab. 2(2), 294–313.

Breiman, L. (1992).
Probability, Volume 7 of *Classics in Applied Mathematics*.
Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM).
Corrected reprint of the 1968 original.

Doyle, P. G. and J. L. Snell (1984).
Random walks and electric networks, Volume 22 of *Carus Mathematical Monographs*.
Washington, DC: Mathematical Association of America.

Fleming, I. (1953).
Casino Royale.
Jonathan Cape.

Grimmett, G. R. and D. R. Stirzaker (2001).
Probability and random processes (Third ed.).
New York: Oxford University Press.

Häggström, O. (2002).
Finite Markov chains and algorithmic applications, Volume 52 of *London Mathematical Society Student Texts*.
Cambridge: Cambridge University Press.

Jerrum, M. (2003).
Counting, sampling and integrating: algorithms and complexity.
Lectures in Mathematics ETH Zürich. Basel: Birkhäuser Verlag.

Kelly, F. P. (1979).
Reversibility and stochastic networks.
Chichester: John Wiley & Sons Ltd.
Wiley Series in Probability and Mathematical Statistics.

Kendall, W. S. (2004).
Geometric ergodicity and perfect simulation.
Electron. Comm. Probab. 9, 140–151 (electronic).

Kendall, W. S., F. Liang, and J.-S. Wang (Eds.) (2005).
Markov chain Monte Carlo: Innovations and Applications.
Number 7 in IMS Lecture Notes. Singapore: World Scientific.

Kendall, W. S. and G. Montana (2002).
Small sets and Markov transition densities.
Stochastic Process. Appl. 99(2), 177–194.

Kindermann, R. and J. L. Snell (1980).
Markov random fields and their applications, Volume 1 of *Contemporary Mathematics*.
Providence, R.I.: American Mathematical Society.

Levin, D. A., Y. Peres, and E. L. Wilmer (2009).
Markov chains and mixing times.
American Mathematical Soc.

Meyn, S. P. and R. L. Tweedie (1993).
Markov chains and stochastic stability.
Communications and Control Engineering Series. London: Springer-Verlag London Ltd.

Murdoch, D. J. and P. J. Green (1998).
Exact sampling from a continuous state space.
Scand. J. Statist. 25(3), 483–502.

Norris, J. R. (1998).
Markov chains, Volume 2 of *Cambridge Series in Statistical and Probabilistic Mathematics*.
Cambridge: Cambridge University Press.
Reprint of 1997 original.

Nummelin, E. (1978).
A splitting technique for Harris recurrent Markov chains.
Z. Wahrsch. Verw. Gebiete 43(4), 309–318.

Nummelin, E. (1984).
General irreducible Markov chains and nonnegative operators, Volume 83 of *Cambridge Tracts in Mathematics*.
Cambridge: Cambridge University Press.

Ross, S. M. (1996).
Stochastic processes (Second ed.).
Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York.

Williams, D. (1991).
Probability with martingales.
Cambridge Mathematical Textbooks. Cambridge: Cambridge University Press.

Williams, D. (2001).
Weighing the odds: A course in probability and statistics, Volume 548.
Springer.

Photographs used in text

- Police phone box
  en.wikipedia.org/wiki/Image:Earls_Court_Police_Box.jpg
- Lightbulb en.wikipedia.org/wiki/File:Gluehlampe_01_KMJ.jpg
- The standing martingale en.wikipedia.org/wiki/Image:Hunterhorse.jpg
- The cardplayers en.wikipedia.org/wiki/Image:
  Paul_C%C3%A9zanne%2C_Les_joueurs_de_carte_%281892-95%29.jpg
- Chinese abacus en.wikipedia.org/wiki/Image:Boulier1.JPG
- Error function en.wikipedia.org/wiki/Image:Error_Function.svg
- Boomerang en.wikipedia.org/wiki/Image:Boomerang.jpg
- Alexander Lyapunov
  en.wikipedia.org/wiki/Image:Alexander_Ljapunow_jung.jpg
- Riffle shuffle (photo by Johnny Blood)
  en.wikipedia.org/wiki/Image:Riffle_shuffle.jpg