LMS/EPSRC Short Course Stability, Coupling Methods and Rare Events
Heriot-Watt University, Edinburgh, 4-9 September 2006
Lecture Notes on:

Coupling and Harris Processes

Serguei Foss and Takis Konstantopoulos

1 A simple example

Consider Markov chain X_n in a countable S with transition probabilities $p_{i,j}$ such that: There exists a state $a \in S$ and $\varepsilon > 0$, with the property:

$$p_{i,a} \geq \varepsilon > 0, \quad \text{for all } i \in S.$$

We will show that there is a unique stationary distribution π

$$\sum_{j \in S} |P(X_n \in j) - \pi(j)| \leq 2(1-\varepsilon)^n,$$

regardless of the initial state X_0. We give two proofs.

Proof 1 ("analytic") Think in terms of of the one-step transition matrix $P = [p_{i,j}]$ as a linear operator acting on $\mathbb{R}^S$. Equip $\mathbb{R}^S$ with the norm $||x|| := \sum_{j \in S} |x_j|$. Stroock (2000) [pg. 28-29] proves that, for any $\rho \in \mathbb{R}^S$, such that $\sum_{i \in S} \rho_i = 0$, we have

$$||\rho P|| \leq (1-\varepsilon)||\rho||.$$

He then claims that this implies that

$$||\rho P^n|| \leq (1-\varepsilon)^n ||\rho||, \quad n \in \mathbb{N},$$

and uses this to show that, for any $\mu \in \mathbb{R}^S$ with, $\mu_i \geq 0$ for all $i \in S$, and $\sum_{i \in S} \mu_i = 1$, it holds that $||\mu P^n - \mu P^m|| \leq 2(1-\varepsilon)^m$, for $m \leq n$.

Proof 2 (probabilistic) Consider the following experiment. Suppose the current state is i. Toss a coin with $P(heads) = \varepsilon$. If heads show up then move to state a. If tails show up, then move to state j with probability

$$\widetilde{p}_{i,j} = \frac{p_{i,j} - \varepsilon \delta_{a,j}}{1-\varepsilon}.$$

(That this IS a valid probability is a consequence of the assumption!) In this manner, we obtain a process that has precisely transition probabilities $p_{i,j}$. Note that state a will be visited either because of heads in a coin toss or because it was chosen so by the alternative transition probability. So state a will be visited at least as many times as the number of heads in a coin toss. This means that state a is positive recurrent. And so a stationary probability π exists. We will show that this π is unique and that the distibution of the chain converges to it. To do this, consider two chains X, $\overline{X}$, both with transition probabilities

$p_{i,j}$, and realise them as follows. The first one starts with X_0 distributed according to an arbitrary μ. The second one starts with $\overline{X}_0$ distributed according to π. Now do this: Use the same coin for both. So, if heads show up then move both chains to a. If tails show up then realise each one according to $\widetilde{p}$, independently. Repeat this at the next step, by tossing a new coin, independently of the past. Thus, as long as heads have not come up yet, the chains are moving independently. Of course, sooner or later, heads will show up and the chains will be the same thereafter. Let T be the first time at which heads show up. We have:

$$\begin{aligned} P(X_n \in B) = P(X_n \in B, T > n) + P(X_n \in B, T \le n) \\ = P(X_n \in B, T > n) + P(\overline{X}_n \in B, T \le n) \\ \le P(T > n) + P(\overline{X}_n \in B) = P(T > n) + \pi(B). \end{aligned}$$

Similarly,

$$\begin{aligned} \pi(B) = P(\overline{X}_n \in B) = P(\overline{X}_n \in B, T > n) + P(\overline{X}_n \in B, T \le n) \\ = P(X_n \in B, T > n) + P(\overline{X}_n \in B, T \le n) \\ \le P(T > n) + P(X_n \in B). \end{aligned}$$

Hence

$$|P(X_n \in B) - \pi(B)| \le P(T > n) = (1 - \varepsilon)^n.$$

Finally, check that $\sup_{B \subseteq S} P(X_n \in B) - \pi(B)| = \frac{1}{2} \sum_{i \in S} |P(X_n = i) - \pi(i)|$. □

2 Coupling

Loosely speaking (more information is being provided in Hermann Thorisson's lectures), coupling means putting random objects on a common probability space. By judicioulsy choosing the common probability space, one can often achieve small miracles. What we did above was an example of a coupling: We put two Markov chains on the same probability space.

Some other examples of coupling are :

Ex. 0: Skorokhod embedding Let X be a zero-mean r.v. with distribution F. Then we can realise X on a probability space Ω supporting at least a standard Brownian motion (W_t) by means of the Skorokhod embedding. This is another instance of coupling. For example, if $F = p\delta_a + q\delta_b$, where $p + q = 1$, $pa + qb = 0$, we let $T = \inf\{t \ge 0 : \; W_t \notin (a, b)\}$ and let $X = W_T$. Then $X \sim F$.

Ex. 1: Dynkin's trick You shuffle well a deck of n cards. Let their values, after the shuffle, be: $x(1), x(2), \dots, x(n)$. (If we are talking about playing cards with J, Q, K included, assign some specific numbers to them, e.g. J=1, Q=2, K=3.)
Experiment (performed by you, in secrecy): Starting from one of the first 10 cards at

random, say card N ($P(N = i) = 1/10, i = 1, \ldots, 10$), you notice the value $x(N)$ and then move forward to the card with index $N + x(N)$, notice its value $x(N + x(N))$, move forward to the card with index $N + x(N + x(N)))$, etc. So, at the i-th step, $i \geq 1$, if you are at position N_i, at a card with value $x(N_i)$, you move to position $N_{i+1} = N_i + x(N_i)$ and notice the value $x(N_{i+1})$. (We take $N_0 = N$). The process stops at the step I for which $N_I + x(N_I) > n$. You notice $x(N_I)$.
The value $x(N_I)$ can be guessed with probability $\approx 1 - c^n$, where $c < 1$.

The coupling here is this: One process starts from onw of the first 10 cards at random. Another process starts from a specific card, say the first one. Then, if n is large, the two processes will meet.

Ex. 2: Common component for 2 distributions Let F, G be two distributions on $\mathbb{R}$. I pick i.i.d. numbers X_n according to F and you pick i.i.d. Y_n according to G. We want to do this in a way that, *for sure* $X_n = Y_n$ for some finite n. The caveat here is that it is not a requirement that my sequence be independent of yours. Obviously, if the supports of F and G are disjoint, then there is no way to achieve equality in finite time. So let us say that the supports of F and G contain a set, say a "small" interval I. Let then μ be some measure supported on I, e.g., the uniform distribution on I. The important thing is that *there is* some nontrivial measure finite μ such that

$$F \geq \mu, \quad G \geq \mu,$$

in the sense that $F(B) \geq \mu(B)$, $G(B) \geq \mu(B)$ for all Borel sets B. Then, letting $||\mu|| := \mu(\mathbb{R})$ and assuming $0 < ||\mu|| < 1$ (this is not a problem), we have

$$F = ||\mu|| \, \frac{\mu}{||\mu||} + ||F - \mu|| \, \frac{F - \mu}{||F - \mu||}$$
$$G = ||\mu|| \, \frac{\mu}{||\mu||} + ||G - \mu|| \, \frac{G - \mu}{||G - \mu||}$$

Now here is how we do a coupling. Let Ω be a probability space supporting 4 independent sequences of i.i.d. random variables: the sequence (δ_n) such that $P(\delta_n = 1) = ||\mu||$, $P(\delta_n = 0) = 1 - ||\mu||$, the sequence (ξ_n) such that $P(\xi_n \in \cdot) = \mu(\cdot)/||\mu||$, the sequence (X'_n) such that $P(X'_n \in \cdot) = (F(\cdot) - \mu(\cdot))/||F - \mu||$, and the sequence (Y'_n) such that $P(Y'_n \in \cdot) = (G(\cdot) - \mu(\cdot))/||G - \mu||$. We can take $\Omega = \{0, 1\} \times \mathbb{R}^3$–the canonical space of $(\delta_n, \xi_n, X'_n, Y'_n)_{n \in \mathbb{N}}$. We now define

$$X_n = \delta_n \xi_n + (1 - \delta_n) X'_n$$
$$Y_n = \delta_n \xi_n + (1 - \delta_n) Y'_n.$$

Clearly, $X_n \sim F$, $Y_n \sim G$, as needed, and if $T = \inf\{n : \delta_n = 1\}$, we have $X_T = Y_T$. But T is a geometric random variable with $P(T > n) = (1 - ||\mu||)^n$ and so a.s. finite.

Ex. 3: Common component for arbitrarily many distributions We can generalise Example 1 to a family $\{F_\alpha\}$ of distributions instead of just two of them. We say that they have a common component if we can find a measure μ such that $\mu \leq \inf_\alpha F_\alpha$. Then we obtain a coupling for all of them simultaneously, by mimicking the construction above.

3 Harris processes

Doeblin (1940) was the first to prove stability results for Markov chains in a general state space; he assumed that $P^\ell(x,\cdot) \geq \varepsilon Q(\cdot)$ for all $x \in S$ (see below for explanation of notation). Harris (1956) generalised Doeblin's method. Total variation convergence was first considered by Orey (1959).

Definition of Harris processes We say that the Markov process (X_n) is Harris[1] if there is a *recurrent set* $R \subseteq S$, i.e.

$$\boxed{P_x(\tau_R < \infty) = 1 \quad \text{for all } x \in S}$$

where $\tau_R := \inf\{n \geq 1 : \ X_n \in R\}$, and an $\ell \in \mathbb{N}$ such that the family of probability measures $\{P^\ell(x,\cdot),\ x \in R\}$ have a common component; in other words,

$$\boxed{\text{there is } p \in (0,1) \text{ and a prob. measure } Q \text{ s.t. } P^\ell(x,\cdot) \geq pQ(\cdot), \quad \forall x \in R}$$

This recurrent set R is often called a *regeneration set*.[2] The discussion that follows justifies the terminology and shows that a Harris process always possesses an invariant measure (which may possibly have infinite mass).

Coupling Suppose $X_0 = x \in R$. Write X_ℓ^x for the Markov process at time ℓ. As in Ex. 1, we may realise the family of random variables $\{X_\ell^x, x \in R\}$ in a way that $P(X_\ell^x = X_\ell^y \text{ for all } x, y \in R) > 0$. This is done by generating a *single* random variable, say Y, with law Q, and by tossing a coin with probability of success p. If successful, we let $X_\ell^x = Y$, for all $x \in R$. If not, we distribute according to the remaining probability.

Existence of invariant measure We now show that each Harris process has an invariant measure. For simplicity, we shall let $\ell = 1$. This is basically done by an inverse Palm construction. Start with X_1 distributed according to the law Q. Let T_k, $k \in \mathbb{N}$ be the times at which the process hits R. For each such T_k consider a 0/1 r.v. ζ_k with $P(\zeta_k = 1) = p$ and a r.v. Y_k with $P(Y_k \in \cdot) = Q$. Let $K = \inf\{k : \ \zeta_k = 1\}$. Consider the path $\{X_n, 1 \leq n \leq T_K + 1\}$. Forcefully set $X_{T_K+1} = Y_K$. The path $\mathcal{C}_0 := \{X_n, 1 \leq n \leq T_K + 1\}$ is the first cycle of the process. Considering the iterates of T_K, namely, $T_K^{(m+1)} = T_K^{(m)} + T_K \circ \theta^{T_K^{(m)}}$, $m \geq 0$, $T_K^{(0)} \equiv 0$, we obtain the successive cycles $\mathcal{C}_{m+1} = \mathcal{C}_m \circ \theta^{T_K^{(m)}}$. It is clear that X_{T_K+1} has law Q and that the sequence of cycles is stationary. (This is referred to as Palm stationarity.) Moreover, we have a regenerative structure: $\{T_K^{(m)}, 0 \leq m \leq n\}$ is independent of $\{\mathcal{C}_m, m \geq n+1\}$, for all n. The only problem is that the cycle durations may be random variables with infinite mean.

[1] Or that it has the Harris property; or that it is Harris-recurrent

[2] In view of the recent resurgence of interest in Lévy processes, one is warned not to confuse the notion of a "regeneration set" with the notion of a "*random* regenerative set" which is the range of a Lévy process.

Now let P_Q be the law of the process when X_0 is chosen according to Q. Define the measure

$$\mu(\cdot) = E_Q \sum_{n=1}^{T_K+1} \mathbf{1}(X_n \in \cdot).$$

Strong Markov property ensures that μ is stationary, i.e.,

$$\mu(\cdot) = \int_S P(x, \cdot)\mu(dx).$$

Positive Harris recurrence If, in addition, to the Harris property, we also have

$$E_Q(T_K) < \infty, \tag{1}$$

we then say that the process is *positive Harris recurrent.* In such a case, $\pi(\cdot) = \mu(\cdot)/E_Q(T_K)$ defines a stationary probability measure. Moreover, the assumption that R is recurrent ensures that there is no other stationary probability measure.

A sufficient condition for positive Harris recurrence is that

$$\sup_{x \in R} E_x \tau_R < \infty. \tag{2}$$

This is a condition that does not depend on the (usually unknown) measure Q. To see the sufficiency, just use the fact that T_K may be represented as a geometric sum of r.v.'s with uniformly bounded means, so that (2) implies (1). To check (2) the Lyapunov function methods are very useful. Let us also offer a further remark on the relation between (2) and (1): it can be shown that if (1) holds, then there is a $R' \subseteq R$ such that

$$\sup_{x \in R'} E_x \tau_{R'} < \infty.$$

We next give a brief description of the stability by means of coupling, achieved by a positive Harris recurrent process. To avoid periodicity phenomena, we assume that the discrete random variable T_K has aperiodic distribution under P_Q. (A sufficient condition for this is: $P_Q(T_K = 1) > 0$.) Then we can construct a successful coupling between the process starting from an arbitrary $X_0 = x_0$ and its stationary version. Assume that $E_{x_0}\tau_R < \infty$. (This x_0 may or may not be an element of R.) Let $\{X_n\}$ be the resulting process. Then one can show, by means of backwards coupling construction, that the process $\{X_n\}$ *strongly couples* (in the sense of the definition of the previous subsection). with the stationary version. The paper by Foss et al. (1998) contains the proof of this assertion for the special case where R is a singleton; however, the construction can be extended to the general case.

φ-irreducibility and φ-recurrence Recall the concepts for Markov chains in countable state space.

The chain is *irreducible* if for all $i, j \in S$ there is n such that $P_i(X_n = j) > 0$. This is equivalent to: For all $i, j \in S$, $P_i(\tau_j < \infty) > 0$. where τ_j be the first hitting time of j.

The chain is *recurrent* if for all $i, j \in S$, $P_i(\tau_j < \infty) = 1$. Equivalently, for all $i, j \in S$, $P_i(X_n = 1 \text{ i.o.}) = 1$.

For general state space S, these definitions are too restrictive: one might call them 'point irreducibility' and 'point recurrence'. Irreducibility and recurrence ought to be defined in terms of sets.

We fix some σ-finite measure φ on S.

We say that the process is *φ-irreducible* if for all $x \in S$ and all measurable $B \subseteq S$ with $\varphi(B) > 0$, there is n such that $P_x(X_n \in B) > 0$. This is equivalent to: for all $x \in S$ and all measurable $B \subseteq S$ with $\varphi(B) > 0$, $P_x(\tau_B < \infty) > 0$.

We say that the process is *φ-recurrent* if for all $x \in S$ and all measurable $B \subseteq S$ with $\varphi(B) > 0$, $P_x(\tau_B < \infty) = 1$. Equivalently, for all $x \in S$ and all measurable $B \subseteq S$ with $\varphi(B) > 0$, $P_x(X_n \in B \text{ i.o.}) = 1$.

Theorem 1. *Every Harris process is Q-recurrent.*

Orey (1971) has proved that

Theorem 2 (Orey's C-set theorem)**.** *If X is φ-irreducible and B is a φ-positive set then there is a smaller set $C \subseteq B$, an integer ℓ, and $p \in (0, 1)$ such that $\varphi(C) > 0$ and*

$$P^\ell(x, \cdot) \geq p\varphi_C(\cdot), \quad \forall x \in C,$$

where φ_C is the restriction of φ on C.

In other words, Orey has proved that φ-irreducibility implies precisely that the process is Harris. And so the notions of a Harris process and of a φ-recurrent process are equivalent.

4 Perfect simulation: a simple case study

The problem is as follows: given a Harris process possessing a unique invariant probability measure, find a method to draw samples from that invariant measure. The classical method of forming Cesàro sums is not an acceptable solution because the sums are not distributed according to the invariant probability measure. In the limit they are, but one cannot simulate an infinite vector. On the other hand, it is assumed that this invariant probability measure is unknown, so we cannot consider a stationary process to start with. We describe a method below, in the case where S is countable. For more general spaces, see Foss et al. (1998).

Let S be a finite set and $p_{i,j}$ transition probabilities on S, assumed to be irreducible. Given any probability μ_i on S, we can define a Markov chain $\{X_n, n \geq 0\}$ by requiring that

$$X_0 \sim \mu,$$
$$\text{and for all } n \geq 1, \ \ (X_n \mid X_{n-1} = i) \sim p_{i,\cdot}$$

The random sequence thus constructed is ergodic and satisfies all usual limit theorems. In particular, if π denotes the unique invariant probability measure, we have that $X_n \Rightarrow \pi$, weakly, and $n^{-1}\sum_1^n \mathbf{1}(X_k \in \cdot) \to \pi(\cdot)$, a.s.

We choose a particular way to realise the process. Define a sequence of random maps $\xi_n : S \to S$, $n \in \mathbb{Z}$, such that $\{\xi_n(i), i \in S, n \in \mathbb{Z}\}$ are independent, and

$$P(\xi_n(i) = j) = p_{i,j}, \quad i, j \in S, \quad n \in \mathbb{Z}.$$

We assume that these random maps are defined on a probability space $(\Omega, \mathscr{F}, P)$ supporting a measurable flow $\theta : \Omega \to \Omega$ such that $\xi_n \circ \theta = \xi_{n+1}$ for all $n \in \mathbb{Z}$. (E.g., let $\Omega = S^{\mathbb{Z}}$ and $\theta(\omega)(n) = \omega(n+1)$.) Let Y be an random variable with distribution μ, independent of everything else, and define

$$X_0 = Y, \quad X_n = \xi_{n-1} \cdots \xi_0(Y), \quad n \geq 1.$$

(Notational convention: we use a little circle when we denote composition with respect to the ω variable and nothing when we denote composition with respect to the variable $i \in S$.) Clearly this sequence $(X_0, X_1, \ldots)$ realises our Markov chain. Observe that for any $m \in \mathbb{Z}$, the sequence $(Y, \xi_m(Y), \xi_{m+1}\xi_m(Y), \ldots)$ also realises our Markov chain.

Next define the *forward strong coupling time* by

$$\sigma = \inf\{n \geq 0 : \forall i, j \in S \;\; \xi_n \cdots \xi_0(i) = \xi_n \cdots \xi_0(j)\},$$

and the *backward coupling time* by

$$\tau = \inf\{n \geq 0 : \forall i, j \in S \;\; \xi_0 \cdots \xi_{-n}(i) = \xi_0 \cdots \xi_{-n}(j)\}.$$

Proposition 1.
(i) σ is a.s. finite.
(ii) $\tau \sim \sigma$.

Proof. Consider the Markov chain $\{\xi_n, n \in \mathbb{Z}\}$, with values in S^S. Let $\Delta = \{\eta \in S^S : \eta(1) = \cdots = \eta(d)\}$. Then $\{\sigma < \infty\} = \cup_{n \geq 0}\{\xi_n \in \Delta\}$. Although the chain is not the product chain (its components are not independent), the probability that it hits the set Δ in finite time is larger than or equal to the probability that the product chain hits the set Δ in finite time. But the product chain is irreducible, hence the latter probability is 1. Hence (i) holds. To prove (ii), notice that $\xi_{\sigma+n} \cdots \xi_0 \in \Delta$, for all $n \geq 0$, a.s. Hence, for each $n \geq 0$,

$$\{\sigma \leq n\} = \{\xi_n \cdots \xi_0 \in \Delta\}. \tag{3}$$

The same argument applied to the maps composed in reverse order shows that

$$\{\tau \leq n\} = \{\xi_0 \cdots \xi_{-n} \in \Delta\}. \tag{4}$$

The events on the right sides of (3) and (4) have the same probability due to stationarity. This proves (ii). □

Corollary 1. *$\sigma < \infty$, a.s.*

Warning: While it is true that, for each n, the random maps $\xi_n \cdots \xi_0$ and $\xi_0 \cdots \xi_{-n}$ have the same distribution, it is not true, in general, that the processes $\{\xi_n \cdots \xi_0, n \geq 0\}$ and $\{\xi_0 \cdots \xi_{-n}, n \geq 0\}$ are indistinguishable. They are both Markov, the first one being a realisation of our original Markov process, but the second one not.

Define now the random variable

$$Z(i) = \xi_0 \cdots \xi_{-\tau}(i).$$

By the very definition of τ, $Z(i) = Z(j)$ for all i, j. So we may set $Z(i) = Z$.

Proposition 2. $Z \sim \pi$.

Proof.

$$\begin{aligned} P(Z = k) &= \lim_{n \to \infty} P(Z = k, \tau \leq n) \\ &= \lim_{n \to \infty} P(\xi_0 \cdots \xi_{-n}(i) = k) \\ &= \lim_{n \to \infty} P(\xi_n \cdots \xi_0(i) = k) = \pi(k), \end{aligned}$$

the latter equality following from the convergence $X_n \Rightarrow \pi$. □

We now describe the actual algorithm. We think of an element $\eta \in S^S$ either as a map $\eta : S \to S$ or as a vector $(\eta(1), \eta(2), \ldots, \eta(d))$, where d is the cardinality of S. So, we may think of $\xi_0, \xi_{-1}, \xi_{-2}, \ldots$ as i.i.d. random vectors in S^d. These vectors are needed at each step of the algorithm and we assume we have random number generator that allows picking these vectors. The algorithm can be written as:

1. $\psi_0 =$ identity
2. $n = 1$
3. $\psi_n = \psi_{n-1}\xi_{n-1}$
4. If $\psi_n \notin \Delta$, set $n = n + 1$ and repeat the previous step;
   otherwise, set $Z = \psi_n(1)$ and stop.

For example, let $S = \{1, 2, 3, 4\}$. Suppose that the ξ_{-k}, $k \geq 0$, are drawn as follows:

$$\begin{aligned} \xi_0 &= (2, 3, 2, 1) \\ \xi_{-1} &= (2, 1, 4, 3) \\ \xi_{-2} &= (2, 4, 2, 1) \\ \xi_{-3} &= (2, 1, 4, 3) \\ \xi_{-4} &= (4, 1, 1, 2) \\ \xi_{-5} &= (1, 2, 3, 3) \\ &\cdots\cdots \end{aligned}$$

The successive passes of the algorithm yield the following vectors:

$$\begin{aligned}
\psi_0 &= \text{identity} = (1,2,3,4) \\
\psi_1 &= \psi_0\xi_0 = (2,3,2,1) \\
\psi_2 &= \psi_1\xi_{-1} = (3,2,1,2) \\
\psi_3 &= \psi_2\xi_{-2} = (2,2,2,3) \\
\psi_4 &= \psi_3\xi_{-3} = (2,2,3,2) \\
\psi_5 &= \psi_4\xi_{-4} = (2,2,2,2)
\end{aligned}$$

The algorithm stops at the 5-th step, yielding the sample $(2,2,2,2)$. We know that this sample is drawn from the true invariant distribution.

Some other points now. Let us look again at the Markov chains

$$\begin{aligned}
\varphi_n &:= \xi_{n-1}\cdots\xi_0 \\
\psi_n &:= \xi_0\cdots\xi_{-(n-1)}, \quad n \geq 1.
\end{aligned}$$

They satisfy

$$\begin{aligned}
\varphi_n &:= \xi_{n-1}\varphi_{n-1} \\
\psi_n &:= \psi_{n-1}\xi_{-(n-1)}, \quad n \geq 1.
\end{aligned}$$

These equations prove the Markov property of the sequences. The first (forward) chain has the property that individual components are also Markov. Indeed, the i-th component $\{\varphi_n(i), n = 0,1,\ldots\}$ is our original Markov chain $\{X_n, n = 0,1,\ldots\}$, with transition probabilities $p_{k,\ell}$, started from $\mu = \delta_i$ (i.e. $X_0 = i$, a.s.) The component processes are, however, not independent. We can easily compute the transition probabilities $p(x,y) := P(\varphi_n = y \mid \varphi_{n-1} = x)$ as follows:

$$\begin{aligned}
p(x,y) &= P(\xi_{n-1}\varphi_{n-1} = y \mid \varphi_{n-1} = x) \\
&= P(\xi_{n-1}(x(i)) = y(i), i = 1,\ldots,d) \\
&= P(\xi_0(x(i)) = y(i), i = 1,\ldots,d) \qquad (5)
\end{aligned}$$

If x is a permutation then the above probability is a product of d terms. Otherwise, if two or more components of x are equal, the number of factors reduces. We observe again that the set Δ is a closed set, while the states in $S^S - \Delta$ are transitive.

On the other hand, the (backward) process $\{\psi_n\}$, albeit Markovian, does not have the same transition probabilities as the forward chain, neither are its components Markovian. To see this very explicitly, let us compute the transition probabilities $\bar{p}(x,y) := P(\psi_n = y \mid \psi_{n-1} =$

x):

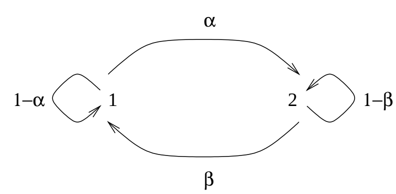
$$
\begin{aligned}
\bar{p}(x,y) &= P(\psi_{-(n-1)}\xi_{-(n-1)} = y \mid \psi_{-(n-1)} = x) \\
&= P(x\xi_{-(n-1)} = y) \\
&= P(x(\xi_0(1)) = y(1), \ldots, x(\xi_0(d)) = y(d)) \\
&= \prod_{i=1}^{d} P(x(\xi_0(i)) = y(i)) \\
&= \prod_{i=1}^{d} P(\xi_0(i) \in x^{-1}(y(i))) \\
&= \prod_{i=1}^{d} \sum_{j \in x^{-1}(y(i))} p_{i,j}
\end{aligned} \tag{6}
$$

For this chain we observe that not only the set Δ is closed, but also that each individual element of Δ is an absorbing state. The elements of $S^S - \Delta$ are transient.

Example. Let us consider $S = \{1,2\}$ and transition probability matrix

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

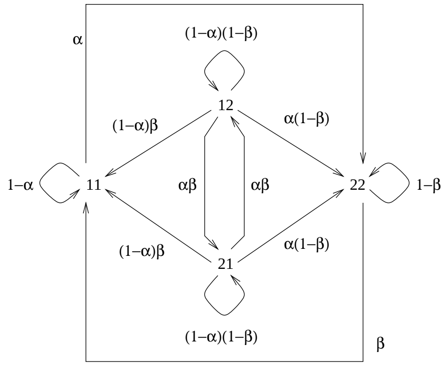
The transition diagram is shown below. The chains $\{\varphi_n\}$ and $\{\psi_n\}$ take values in $S^S =$



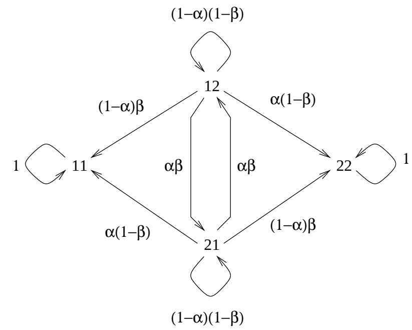
$S^2 = \{11, 12, 21, 22\}$. Their transition diagrams are shown next. Notice the properties mentioned before: The set Δ is closed in both cases. Individual states of Δ are absorbing in the backward case. All other states are transient in both cases. We computed the transition probabilities following equations (5) and (6). For instance, with $x = (21)$ and $y = (11)$, we have

$$
\begin{aligned}
p(x,y) &= P(\xi_0(x(1)) = y(1),\ \xi_0(x(2)) = y(2)) \\
&= P(\xi_0(2) = 1,\ \xi_0(1) = 1) = \beta(1-\alpha), \\
\bar{p}(x,y) &= P(\xi_0(1) \in x^{-1}(y(1)),\ \xi_0(2) \in x^{-1}(y(2)) \\
&= P(\xi_0(1) \in x^{-1}(1),\ \xi_0(2) \in x^{-1}(1)) \\
&= P(\xi_0(1) = 2,\ \xi_0(2) = 2) = \alpha(1-\beta).
\end{aligned}
$$

The invariant probability measure of the original 2-state chain is $\pi = \beta/(\alpha+\beta), \alpha/(\alpha+\beta))$. When the backward chain is absorbed at Δ, each of its components has distribution π. It is easily checked that this is not the case, in general, with the forward chain. For instance, suppose $\alpha = 1$ and $0 < \beta < 1$. Then $\pi = (\beta/(1+\beta), 1/(1+\beta))$. But forward coupling can only occur at state 2. In other words, at time τ, the distribution of the chain takes

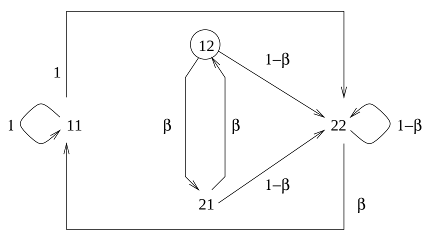


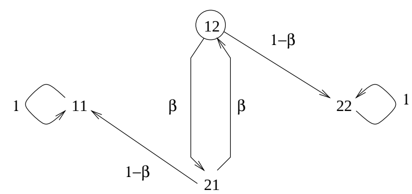
Transition diagram for the forward chain



Transition diagram for the backward chain

value 2 with probability one. This is not the stationary distribution! (Of course, if the forward coupled is allowed to evolve ad infinitum, it certainly converges to π.) On the other hand, it can be easily checked that the backward chain, starting at the state (12) (see next figure), is absorbed at state (11) with probability $\beta/(1+\beta)$ and at state (22) with probability $1/(1+\beta)$). This is precisely the stationary distribution.





5 An example of a non-Harris process

Let A, B, C be three non-collinear points on the plane, forming the vertices of a triangle. Let X_0 be an arbitrary point on the plane. Toss a 3-sided coin and pick one of the vertices at random. Join X_0 with the chosen vertex by a straight line segment and let X_1 be the middle of the segment. Repeat the process independently. We thus obtain a Markov process $(X_n, n \geq 0)$. The process has a unique stationary version constructed as follows. Let $(\xi_n, n \in \mathbb{Z})$ be i.i.d. random elements of the plane, such that ξ_1 is distributed on $\{A, B, C\}$ uniformly. By assigning some Cartesian coordinates, define

$$X_{n+1} = \frac{1}{2}(X_n + \xi_n).$$

Iterate the recursion, starting from m and ending up at $n \geq m$. Fix n and let $m \to -\infty$. We obtain

$$X_n = \sum_{k=1}^{\infty} \frac{1}{2^k} \xi_{n-k}.$$

The so-constructed process is Markovian with the required transitions and is also a stationary process. There is no other stationary solution to the recursion. So the stationary distribution is the distribution of the random variable

$$\sum_{k=1}^{\infty} \frac{1}{2^k} \xi_k.$$

It is not hard to see that the stationary distribution is supported on a set of Lebesgue measure zero (the so-called Sierpiński gasket).

6 Discussion

1. Is Stroock's proof correct? Carry out the steps in detail.
2. Show, in detail, that if X_n is non-periodic positive Harris recurrent then the law of X_n converges in total variation to the unique invariant probability measure.
3. Show that any irreducible, (positive) recurrent Markov chain in a countable state space S is (positive) Harris recurrent.
4. Let $X_{a,b}$ be a random variable with $P(X_{a,b} = a) = p$, $P(X_{a,b} = b) = 1 - p$, $EX_{a,b} = 0$. Let Y be another 0-mean random variable. Show that, by randomising a, b, we can write $Y = X_{A,B}$. Use this to couple any 0-mean random variable and a Brownian motion.
5. Why is the process of the last example non-Harris? Find a formula for the characteristic function of the stationary distribution.
6. In the same example, let X_0 have a law with density. Will (X_n) couple with its stationary version?
7. How are Lyapunov function methods used for checking positive Harris recurrence? (For the countable state-space case check the excellent book of Brémaud (2001).)

References

Asmussen, S.(2003). *Applied Probability and Queues.* Springer.

Brémaud, P. (2001). *Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues.* Springer.

Doeblin, W. (1940). Eléments d'une théorie gńérale des chaînes simples constantes de Markoff. *Ann. Sci. Ecole Normale Sup.* **57**, 61-111.

Foss, S.G., Tweedie, R.L. and Corcoran, J.N. (1998). Simulating the invariant measures of Markov chains using backward coupling at regeneration times. *Prob. Eng. Inf. Sci.* **12**, 303-320.

Foss, S. and Konstantopoulos, T. (2004). An overview of some stochastic stability methods. *J. Oper. Res. Soc. Japan*, 275-303.

Harris, T.E. (1956). The existence of stationary measures for cerain Markov processes. *Proc. 3d Berkeley Symp. Math. Stat. Prob.* **2**, 113-124.

Meyn, S. and Tweedie, R. (1993). *Markov Chains and Stochastic Stability.* Springer, New York.

Orey, S. (1959). Recurrent Markov chains. *Pacific J. Math.* **9**, 805-827.

Orey, S. (1971). *Limit Theorems for Markov Chain Transition Probabilities.* Van Nostrand Reinhold.

Stroock, D. (2005) *An Introduction to Markov Processes.* Springer.