Martingales and Evolving Sets

Jan M. Swart (Czech Academy of Sciences)

Thursday, January 7th, 2020

Optional stopping

Optional stopping, version 3 Let $(M_t)_{t \geq 0}$ be a martingale with respect to some filtration $(\mathcal{F}_t)_{t \geq 0}$ and let τ be a stopping time. Assume that:

(i) There exists a constant $C < \infty$ such that
$|M_t - M_{t-1}| \leq C$ a.s. for all $t \geq 1$.

(ii) $\mathbb{E}[\tau] < \infty$.

Then $\mathbb{E}[M_\tau] = \mathbb{E}[M_0]$.

Proof

$$|M_{\tau \wedge t}| \leq |M_0| + \sum_{s=1}^{\tau \wedge t} |M_s - M_{s-1}| \leq |M_0| + C\tau.$$

Since $\mathbb{E}[|M_0| + C\tau] < \infty$, by optional stopping and dominated convergence

$$\mathbb{E}[M_0] = \mathbb{E}[M_{\tau \wedge t}] \underset{t \to \infty}{\longrightarrow} \mathbb{E}[M_\tau].$$

An application

Let $(B_t)_{t \geq 1}$ be i.i.d. and uniformly distributed on $\{0, 1\}$ and let
$\tau_{010} := \inf\{t \geq 0 : (B_{t-2}, B_{t-1}, B_t) = (0, 1, 0)\}$.
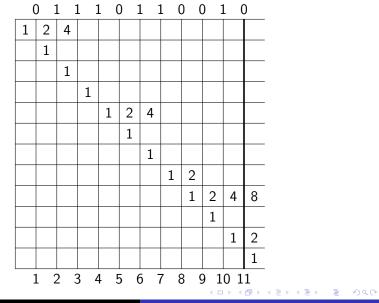
Claim $\mathbb{E}[\tau_{010}] = 10$.

Proof At each time $t = 0, 1, \ldots$ a gambler enters with one dollar.
The gambler bets all his money on the next outcome being 0.
If he wins, he bets all he has on the next outcome being 1.
If he wins again, he bets all he has on the next outcome being 0.

Let $S_0 := 1$ and $S_t :=$ the sum of money owned by all gamblers
after the t-th outcome.
Then $S_t - t$ is a martingale and hence $\mathbb{E}[\tau] = \mathbb{E}[S_\tau] - 1$.
We claim $S_\tau = 1 + 2 + 8$ a.s. ∎

	0	1	1	1	0	1	1	0	0	1	0	
1	2	4										
	1											
		1										
			1									
				1	2	4						
					1							
						1						
							1	2				
								1	2	4	8	
									1			
										1	2	
											1	

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |

Doob transform

$X = (X_t)_{t \geq 0}$ Markov chain with finite state space $\mathcal{X}$ and transition kernel P. Assume $z \in \mathcal{X}$ is a trap. Let

$$h(x) := \mathbb{P}^x\big[X_t = z \text{ for some } t \geq 0\big] \qquad (x \in \mathcal{X})$$

Then h is a harmonic function, i.e., $Ph = h$.
Let $\mathcal{X}_+ := \{x \in \mathcal{X} : h(x) > 0\}$. Then for each $x \in \mathcal{X}_+$,

$$\mathbb{P}^x\big[(X_t)_{t \geq 0} \in \, \cdot \, \big| X_t = z \text{ for some } t \geq 0\big]$$

is the law of the Markov chain with state space $\mathcal{X}_+$, initial state x, and *Doob transformed* transition kernel

$$P^h(x, y) := h(x)^{-1} P(x, y) h(y) \qquad (x, y \in \mathcal{X}_+).$$

Doob transform

Proof Let $A := \{X_t = z \text{ for some } t \geq 0\}$.
For each $(x_0, \ldots, x_n)$ with $x_0 = x$ and $x_n = z$

$$\mathbb{P}^x\big[(X_0, \ldots, X_n) = (x_0, \ldots, x_n)\big] = P(x_0, x_1) \cdots P(x_{n-1}, x_n)$$
$$= h(x_0) P^h(x_0, x_1) \cdots P^h(x_{n-1}, x_n) h(x_n)^{-1}.$$

Here $h(x_n) = h(z) = 1$ and $h(x_0) = h(x) = \mathbb{P}^x(A)$, so

$$\mathbb{P}^x\big[(X_0, \ldots, X_n) = (x_0, \ldots, x_n) \,\big|\, A\big] = P^h(x_0, x_1) \cdots P^h(x_{n-1}, x_n).$$

∎

Let $X = (X_t)_{t \geq 0}$ be the Markov chain with finite state space $\mathcal{X}$, irreducible transition kernel P, and unique invariant law π.

Assume X is adapted to a filtration $(\mathcal{F}_t)_{t \geq 0}$ and that τ is an $\mathcal{F}_t$-stopping time.

By definition, τ is a *strong stationary time* if

$$\mathbb{P}\big[X_\tau = x \,\big|\, \tau\big] = \pi(x) \qquad (x \in \mathcal{X}).$$

Equivalently, X_τ is independent of τ and has law π.

Diaconis and Fill (1990) have shown that any irreducible Markov chain has a strong stationary time.

The following treatment is influenced by Morris and Peres (2005).

Evolving set process

Let R be defined by

$$\pi(x)P(x,y) = R(x,y)\pi(y) \qquad (x,y \in \mathcal{X}),$$

i.e., $R'(y,x) := R(x,y)$ is the transition kernel of the reversed Markov chain.

We also write

$$P(x,A) := \sum_{y \in A} P(x,A) \quad \text{and} \quad R(A,y) := \sum_{x \in A} R(x,y).$$

Let $2^{\mathcal{X}}$ denote the space of all subsets of $\mathcal{X}$.

Let $A^{c} := \mathcal{X} \backslash A$ denote the complement of a set $A \subset \mathcal{X}$.

We inductively construct:
$(X_t, S_t)_{t \geq 0}$ with values in $\mathcal{X} \times 2^{\mathcal{X}}$,
$(U_t)_{t \geq 1}$ with values in $[0, 1]$.

Conditional on $(X_s, S_s)_{0 \leq s \leq t-1}$ and $(U_s)_{1 \leq s \leq t-1}$:

(i) We choose X_t according to the probability law $P(X_{t-1}, \cdot)$.

(ii) If $X_{t-1} \in S_{t-1}$, we choose U_t uniformly on $[0, R(S_{t-1}, X_t)]$.
If $X_{t-1} \notin S_{t-1}$, we choose U_t uniformly on $[R(S_{t-1}, X_t), 1]$.

(iii) We set $S_t := \{y \in \mathcal{X} : R(S_{t-1}, y) \geq U_t\}$.

- $(X_t)_{t \geq 0}$ is autonomous with transition kernel P.
- Either $X_t \in S_t$ for all $t \geq 0$, or $X_t \in S_t^c$ for all $t \geq 0$.

Lemma Assume that X_0 has law π and is independent of S_0. Then:

(a) At each time $t \geq 1$, the random variable U_t is uniformly distributed on $[0,1]$ and independent of X_t and $(S_s)_{0 \leq s \leq t-1}$.

(b) At each time $t \geq 0$, the random variable X_t has law π and is independent of $(S_s)_{0 \leq s \leq t}$.

Proof $\mu_{[a,b]} :=$ uniform distribution on the interval $[a,b]$. Part (b) is true by assumption at time $t = 0$. We now prove parts (a) and (b) by induction, showing that if (b) holds at time $t-1$, then (a) and (b) hold at time t.

Evolving set process

We start by proving (a). By our induction hypothesis, conditional on $(S_s)_{0 \le s \le t-1}$, the random variable X_{t-1} has law π and hence

$$\mathbb{P}\big[(X_{t-1}, X_t) = (x, y) \,\big|\, (S_s)_{0 \le s \le t-1}\big] = \pi(x)P(x, y) \qquad (\star)$$

Therefore, if we condition both on $(S_s)_{0 \le s \le t-1}$ and X_t, then the conditional law of X_{t-1} is $R(\,\cdot\,, X_t)$ and hence

$$
\begin{aligned}
&\mathbb{P}\big[U_t \in \,\cdot\, \big|\, X_t, \ (S_s)_{0 \le s \le t-1}\big] \\
&\quad = \mathbb{P}\big[X_{t-1} \in S_{t-1} \big|\, X_t, \ (S_s)_{0 \le s \le t-1}\big] \cdot \mu_{[0, R(S_{t-1}, X_t)]} \\
&\qquad + \mathbb{P}\big[X_{t-1} \in S_{t-1}^{c} \big|\, X_t, \ (S_s)_{0 \le s \le t-1}\big] \cdot \mu_{[R(S_{t-1}, X_t), 1]} \\
&\quad = R(S_{t-1}, X_t)\mu_{[0, R(S_{t-1}, X_t)]} + R(S_{t-1}^{c}, X_t)\mu_{[R(S_{t-1}, X_t), 1]} = \mu_{[0,1]},
\end{aligned}
$$

which proves that (a) holds at time t.

Formula ($\star$) shows that X_t is independent of $(S_s)_{0 \leq s \leq t-1}$, so

$$\mathbb{P}\big[(S_s)_{0 \leq s \leq t-1} \in \cdot \,\big|\, X_t\big] = \mathbb{P}\big[(S_s)_{0 \leq s \leq t-1} \in \cdot\,\big] \quad \text{a.s.}$$

Since (a) holds at time t, U_t is independent of X_t and $(S_s)_{0 \leq s \leq t-1}$, so

$$\mathbb{P}\big[U_t \in \cdot \,\big|\, X_t, \ (S_s)_{0 \leq s \leq t-1}\big] = \mathbb{P}\big[U_t \in \cdot\,\big] \quad \text{a.s.}$$

Together, these formulas show that $\big(U_t, (S_s)_{0 \leq s \leq t-1}\big)$ is independent of X_t. Since S_t is a function of U_t and S_{t-1}, it follows that also $(S_s)_{0 \leq s \leq t}$ is independent of X_t, which proves that (b) holds at time t. ■

Corollary If X_0 has law π and is independent of S_0, then $(S_t)_{t \geq 0}$, taken on its own, is a Markov chain with state space $2^{\mathcal{X}}$ and transition kernel

$$K(A, B) := \mathbb{P}\big[\{y \in \mathcal{X} : R(A, y) \geq U\} = B\big] \qquad (A, B \in 2^{\mathcal{X}}),$$

where U is uniformly distributed on $[0, 1]$.

Remark $(S_t)_{t \geq 0}$ is clearly not autonomous.

We call the Markov chain with kernel K the *evolving set process*.

- $(S_t^{\mathrm{c}})_{t \geq 0}$ is also an evolving set process.
- $\emptyset, \mathcal{X}$ are traps.
- $\tau := \inf\big\{t \geq 0 : S_t \in \{\emptyset, \mathcal{X}\}\big\}$ is a.s. finite.

Conditioned evolving set process

Lemma Assume $S_0 \neq \emptyset$ is deterministic and X_0 has law π. Then

$$\mathbb{P}\big[(S_t)_{t \geq 0} \in \cdot \,\big|\, X_0 \in S_0\big]$$

is the law of the Markov chain with state space
$2_+^{\mathcal{X}} := \{A \subset \mathcal{X} : A \neq \emptyset\}$ and Doob-transformed transition kernel

$$K^\pi(A, B) := \pi(A)^{-1} K(A, B) \pi(B) \qquad (A, B \in 2_+^{\mathcal{X}}).$$

Proof Since either $X_t \in S_t$ for all $t \geq 0$ or $X_t \in S_t^c$ for all $t \geq 0$, the events $\{S_\tau = \mathcal{X}\}$ and $\{X_0 \in S_0\}$ are a.s. equal. In particular

$$\mathbb{P}\big[S_\tau = \mathcal{X} \,|\, S_0 = A\big] = \mathbb{P}[X_0 \in A] = \pi(A).$$

∎

Construction of the strong stationary time

Corollary Assume $X_0 = x$ and $S_0 = \{x\}$.
Then $X = (X_t)_{t \geq 0}$ is the Markov chain with initial state x and transition kernel P and τ is a strong stationary time for X.

Proof Equivalently, let X_0 have law π and condition on $X_0 \in S_0$ or equivalently on $X_t \in S_t$. By property (b)

$$\mathbb{P}\big[X_t = y \,\big|\, (S_s)_{0 \leq s \leq t}\big] = \frac{\pi(y)}{\pi(S_t)} 1_{\{y \in S_t\}} \qquad (y \in \mathcal{X}, \ t \geq 0).$$

and hence

$$\mathbb{P}\big[X_t = y \,\big|\, (S_s)_{0 \leq s \leq t}, \ \tau = t\big] = \pi(y) \qquad (y \in \mathcal{X}, \ t \geq 0).$$

∎

Application

$$Q(x, y) := \pi(x) P(x, y), \quad Q(A, B) := \sum_{x \in A, \; y \in B} Q(x, y).$$

Bottleneck ratio:

$$\Phi_* := \min_{A: \; \pi(A) \leq \frac{1}{2}} \frac{Q(A, A^{\mathrm{c}})}{\pi(A)}.$$

Small value of Φ_* signifies strong bottleneck.

$\lambda_\star$ leading eigenvalue, $\gamma_\star := 1 - |\lambda_\star|$ spectral gap, $\tau_{\mathrm{rel}} := 1/\gamma_\star$ relaxation time, $\pi_{\min} := \min_{x \in \mathcal{X}} \pi(x)$.

Thm 12.4 π reversible, then $\tau_{\mathrm{mix}}(\varepsilon) \leq \tau_{\mathrm{rel}} \log \left(1/(\varepsilon \pi_{\min}) \right)$.

Thm 13.10 π reversible, then $\tau_{\mathrm{rel}} \leq 2/\Phi_*^2$.

Thm 17.10 π lazy, then $\tau_{\mathrm{mix}}(\varepsilon) \leq (2/\Phi_*^2) \cdot \log \left(1/(\varepsilon \pi_{\min}) \right)$.

Application

Proof of Thm 17.10 Since $d(t) \leq d^{(\infty)}(t)$, it suffices to show
that
$$d^{(\infty)}(t) \leq \frac{1}{\pi_{\min}} \left(1 - \frac{\Phi_\star^2}{2} \right)^t,$$

where
$$d^{(\infty)}(t) := \max_{x,y \in \mathcal{X}} \left| \frac{P^t(x,y)}{\pi(y)} - 1 \right|.$$

Lemma 1 $P^t(x,y) = \dfrac{\pi(y)}{\pi(x)} \mathbb{P}^{\{x\}} \left[y \in S_t \right].$

Consequence $d^{(\infty)}(t) = \max\limits_{x,y \in \mathcal{X}} \dfrac{\left| \mathbb{P}^{\{x\}} \left[y \in S_t \right] - \pi(x) \right|}{\pi(x)}.$

Application

Since $\pi(x) = \mathbb{P}^{\{x\}}[S_\tau = \mathcal{X}]$,

$$d^{(\infty)}(t) = \max_{x,y \in \mathcal{X}} \frac{1}{\pi(x)} \left| \mathbb{P}^{\{x\}}[y \in S_t] - \mathbb{P}^{\{x\}}[S_\tau = \mathcal{X}] \right|$$

$$= \max_{x,y \in \mathcal{X}} \frac{1}{\pi(x)} \left| \mathbb{P}^{\{x\}}[y \in S_t, \ S_\tau = \emptyset] - \mathbb{P}^{\{x\}}[y \notin S_t, \ S_\tau = \mathcal{X}] \right|$$

$$\leq \max_{x \in \mathcal{X}} \frac{1}{\pi(x)} \mathbb{P}^{\{x\}}[\tau > t].$$

Lemma 2 $\mathbb{P}^{\{x\}}[\tau > t] \leq \sqrt{\dfrac{\pi(x)}{\pi_{\min}}} \left(1 - \dfrac{\Phi_\star^2}{2}\right)^t.$

Consequence

$$d^{(\infty)}(t) \leq \max_{x \in \mathcal{X}} \frac{1}{\sqrt{\pi(x)\pi_{\min}}} \left(1 - \frac{\Phi_\star^2}{2}\right)^t = \frac{1}{\pi_{\min}} \left(1 - \frac{\Phi_\star^2}{2}\right)^t. \qquad \blacksquare$$

Proof of Lemma 1 Start the coupled process $(X_t, S_t)_{t \geq 0}$ such that X_0 has law π and $S_0 = \{x\}$. By property (b) X_t has law π and is independent of S_t, so

$$\pi(y)\mathbb{P}\big[y \in S_t\big] = \mathbb{P}\big[X_t = y, \ y \in S_t\big]$$
$$= \mathbb{P}\big[X_t = y, \ X_0 \in \{x\}\big] = \pi(x)P^t(x, y).$$

∎

Application

Proof of Lemma 2 Define

$$S_t^{\#} := \begin{cases} S_t & \text{if } \pi(S_t) \leq 1/2, \\ S_t^{\mathrm{c}} & \text{otherwise.} \end{cases}$$

Lemma 17.16 $\mathbb{E}\left[\sqrt{\pi(S_{t+1}^{\#})} \,\Big|\, S_t\right] \leq \left(1 - \dfrac{\Phi_\star^2}{2}\right) \sqrt{\pi(S_t^{\#})}.$

Consequence

$$\mathbb{E}\left[\sqrt{\pi(S_{t+1}^{\#})}\right] = \mathbb{E}\left[\mathbb{E}\left[\sqrt{\pi(S_{t+1}^{\#})} \,\Big|\, S_t\right]\right] \leq \left(1 - \frac{\Phi_\star^2}{2}\right) \mathbb{E}\left[\sqrt{\pi(S_t^{\#})}\right],$$

so by induction

$$\mathbb{E}^{\{x\}}\left[\sqrt{\pi(S_t^{\#})}\right] \leq \left(1 - \frac{\Phi_\star^2}{2}\right)^t \sqrt{\pi(x)}.$$

Application

Since
$$\mathbb{E}^{\{x\}}\Big[\sqrt{\pi(S_t^{\#})}\Big] \geq \sqrt{\pi_{\min}}\,\mathbb{P}^{\{x\}}\big[S_t^{\#} \neq \emptyset\big],$$

it follows that
$$\mathbb{P}^{\{x\}}\big[\tau > t\big] = \mathbb{P}^{\{x\}}\big[S_t^{\#} \neq \emptyset\big]$$
$$\leq \frac{1}{\sqrt{\pi_{\min}}}\mathbb{E}^{\{x\}}\Big[\sqrt{\pi(S_t^{\#})}\Big] \leq \frac{\sqrt{\pi(x)}}{\sqrt{\pi_{\min}}}\Big(1 - \frac{\Phi_{\star}^2}{2}\Big)^t.$$

∎

Application

Recall

$$S_t^\# := \begin{cases} S_t & \text{if } \pi(S_t) \le 1/2, \\ S_t^c & \text{otherwise.} \end{cases}$$

Proof of Lemma 17.16 (sketch) Since $\pi(S_t)$ is a martingale, by Jensen's inequality, $\sqrt{\pi(S_t)}$ is a supermartingale and hence so is $\sqrt{\pi(S_t^\#)} = \sqrt{\pi(S_t)} \wedge \sqrt{\pi(S_t^c)}$.

Lemma 17.14 says that if $\pi(S) \le \frac{1}{2}$, then

$$\mathbb{E}\Big[S_{t+1}/S_t \,\Big|\, U_{t+1} \le \tfrac{1}{2}, \ S_t = S\Big] = 1 + 2\frac{Q(S, S^c)}{\pi(S)} \ge 1 + 2\Phi_*,$$

$$\mathbb{E}\Big[S_{t+1}/S_t \,\Big|\, U_{t+1} \ge \tfrac{1}{2}, \ S_t = S\Big] = 1 - 2\frac{Q(S, S^c)}{\pi(S)} \le 1 - 2\Phi_*,$$

which allows to prove that $\pi(S_t^\#)$ fluctuates enough and hence $\sqrt{\pi(S_t^\#)}$ tends to decrease on average, with explicit estimates. ∎

Final remark Laziness is use right at the start of the proof of Lemma 17.14 to conclude that $y \notin S$ implies $R(S, y) \leq \frac{1}{2}$.
This then becomes important in connection with the inequalities $U_{t+1} \leq \frac{1}{2}$ and $U_{t+1} \geq \frac{1}{2}$.