E-mail: yms@statslab.cam.ac.uk

MARKOV CHAINS

The strong Markov property

Stopping times and statement of the strong Markov property.

The strong Markov property asserts that the process begins afresh not only after any given time n but also after a randomly chosen time. An example of such a time is $H^{\{j\}}$, the time the chain hits a given state $i \in I$. More generally,

Definition 4.1. *A random variable T depending on X_0, X_1, ... and taking values 0, 1, 2, ... , ∞ is called a stopping time if the event $\{T = n\}$ is described in terms of random variables X_0, ..., X_n only, without involving X_{n+1}, X_{n+2}, ... In other words, the indicator $\mathbf{1}(T = n) := \begin{cases} 1, & \text{if } T = n, \\ 0, & \text{if } T \neq n, \end{cases}$ is a function of $X_0, \ldots, X_n$:*

$$\mathbf{1}(T = n) = g(X_0, \ldots, X_n). \tag{4.1}$$

Pictorially, by watching the chain, you know when you should stop without anticipating future states. The hitting time H^A is an example of a stopping time as for $n \geq 1$: $\{H^A = 0\} = \{X_0 \in A\}$, and for $n \geq 1$:

$$\{H^A = n\} = \{X_0 \notin A, \ldots, X_{n-1} \notin A, X_n \in A\}.$$

When A is reduced to a single state i, the hitting time is often called the first passage time:

$$H^j = \inf\, [n \geq 0 : \ X_n = j\}.$$

On the other hand, the last exit time

$$L^A = \sup\, [n : \ X_n \in A]$$

is in general not a stopping time as the event $\{L^A = n\}$ requires knowledge of X_{n+1}, X_{n+2}, ...

Theorem 4.1. *Let $(X_n, n \geq 0)$ be Markov (λ, P) and assume that T is a stopping time. Then, conditional on $T < \infty$ and $X_T = i$, $(X_{n+T}, n \geq 0)$ is Markov (δ_i, P). In particular, conditional on $T < \infty$ and $X_T = i$, random variables X_{T+1}, X_{T+2}, ... are independent of X_0, ..., X_{T-1}.*

Proof. (Non-examinable) Let A be an event determined by the chain before time T, i.e., by X_0, ..., X_{T-1}, and B by the chain after time T, i.e., by X_{T+1}, ... X_{T+n} for some n. We want to check that $\forall\ m \geq 1$ and $i \in I$: (i)

$$\mathbb{P}(A \cap B | T < \infty, X_T = i) = \mathbb{P}(A | T < \infty, X_T = i)\mathbb{P}(B | T < \infty, X_T = i)$$

and (ii) the conditional probability $\mathbb{P}(B | T < \infty, X_T = i)$ is calculated as in the Markov chain (δ_i, P):

$$\mathbb{P}(B | T < \infty, X_T = i) = \sum_{(j_1, \dots j_n) \in B} p_{ij_1} \cdots p_{j_{n-1}j_n}.$$

As in the proof of the Markov property, we first assume that A is of the form $\{X_0 = i_0, \dots, X_{m-1} = i_{m-1}\}$ and B of the form $\{X_{T+1} = j_1, \dots, X_{T+n} = j_n\}$ for some $i_0, \dots, i_{m-1}, j_1, \dots, j_n \in I$. Given m, the event

$$A \cap \{T = m\} \cap \{X_T = i\} = A \cap \{T = m, X_m = i\}$$

is simply

$$\{X_0 = i_0, \dots, X_{m-1} = i_{m-1}, X_m = i_m\}$$

if $T(i_0, \dots, i_{m-1}, i) = m$ and empty if $T(i_0, \dots, i_{m-1}, i) \neq m$. Then the event $A \cap B\{T = m, X_T = i\} = A \cap \{T = m, X_m = i\} \cap B$ has probability

$$\lambda_{i_0} p_{i_0 i_1} \cdots p_{i_{m-1} i} p_{i j_1} \cdots p_{j_{n-1} j_n} \mathbf{1}\big(T(i_0, \dots, i_{m-1}, i) = m\big).$$

For a general B we have to sum over $(j_1, \dots, j_n) \in B$:

$$\lambda_{i_0} p_{i_0 i_1} \cdots p_{i_{m-1} i} \mathbf{1}\big(T(i_0, \dots, i_{m-1}, i) = m\big) \sum_{(j_1, \dots, j_n) \in B} p_{i j_1} \cdots p_{j_{n-1} j_n}.$$

The sum $\sum\limits_{(j_1, \dots, j_n) \in B}$ does not depend on m; it gives the conditional probability $\mathbb{P}(B | T < \infty, X_T = i)$ and is calculated as in the Markov chain (δ_i, P).

For a general A we now sum over $(i_0, \dots, i_{m-1}) \in A$:

$$\mathbb{P}(A \cap B \cap \{T = m, X_T = i\})$$

$$= \sum_{(i_0,\dots,i_{m-1})\in A} \lambda_{i_0} p_{i_0 i_1} \cdots p_{i_{m-1} i} \mathbf{1}\big(T(i_0, \dots, i_{m-1}, i) = m\big) \mathbb{P}(B | T < \infty, X_T = i)$$

$$= \mathbb{P}(A \cap \{T = m, X_T = i\}) \mathbb{P}(B | T < \infty, X_T = i).$$

Summing over m then gives

$$\mathbb{P}(A \cap B \cap \{T < \infty, X_T = i\}) = \mathbb{P}(A \cap \{T < \infty, X_T = i\}) \mathbb{P}(B | T < \infty, X_T = i).$$

Finally, dividing by $\mathbb{P}(T < \infty, X_T = i)$ yields that the conditional probability $\mathbb{P}(A \cap B | T < \infty, X_T = i)$ equals

$$\frac{\mathbb{P}(A \cap \{T < \infty, X_T = i\})}{\mathbb{P}(T < \infty, X_T = i)} \mathbb{P}(B | T < \infty, X_T = i)$$

$$= \mathbb{P}(A | T < \infty, X_T = i) \mathbb{P}(B | T < \infty, X_T = i)$$

as required.

The conditional probability $\mathbb{P}(A \cap \{T = m, X_T = i\} \cap B | X_m = i)$, given that $X_m = 1$, is obtained after division by $\mathbb{P}(X + m = i) = (\lambda P^m)_i$: the ratio is determined by X_0, ..., X_m, and the conditional probability

$$\mathbb{P}\big((A \cap \{T = m\}) \cap \{X_{T+1} = j_1, \dots, X_{T+n} = j_n\} | X_m = i\big)$$

$$= \mathbb{P}\big((A \cap \{T = m\}) \cap \{X_{m+1} = j_1, \dots, X_{m+n} = j_n\} | X_m = i\big).$$

By the Markov property we have the decomposition:

$$\mathbb{P}\big((A \cap \{T = m\}) \cap \{X_{m+1} = j_1, \dots, X_{m+n} = j_n\} | X_m = i\big)$$

$$= \mathbb{P}\big(A \cap \{T = m\} | X_m = i) \mathbb{P}(X_{m+1} = j_1, \dots, X_{m+n} = j_n | X_m = i\big)$$

$$= \mathbb{P}\big(A \cap \{T = m\} | X_m = i) p_{i j_1} \cdots p_{j_{n-1} j_n}.$$

Hence, the unconditional probability

$$\mathbb{P}\big((A \cap \{T = m\}) \cap \{X_{m+1} = j_1, \dots, X_{m+n} = j_n\} \cap \{X_m = i\}\big)$$

$$= \mathbb{P}\big((A \cap \{T = m, X_m = i\}) \cap \{X_{m+1} = j_1, \dots, X_{m+n} = j_n\}\big)$$

equals

$$\begin{aligned}&\mathbb{P}\big(A\cap\{T=m\}|X_m=i)\mathbb{P}(X_m=i)p_{ij_1}\cdots p_{j_{n-1}j_n}\\&\quad=\mathbb{P}\big(A\cap\{T=m,X_m=i\}\big)p_{ij_1}\cdots p_{j_{n-1}j_n}.\end{aligned}$$

Summing over m yields

$$\begin{aligned}&\mathbb{P}\big((A\cap\{T<\infty,X_T=i\}\cap\{X_{m+1}=j_1,\ldots,X_{m+n}=j_n\}\big)\\&\quad=\mathbb{P}\big(A\cap\{T<\infty,X_T=i\}\big)p_{ij_1}\cdots p_{j_{n-1}j_n}\end{aligned}$$

and dividing by $\mathbb{P}(T<\infty,X_m=i)$:

$$\begin{aligned}&\mathbb{P}\big(A\cap\{X_{m+1}=j_1,\ldots,X_{m+n}=j_n\}|T<\infty,X_T=i\big)\\&\quad=\mathbb{P}\big(A\cap\{T<\infty,X_T=i\}\big)p_{ij_1}\cdots p_{j_{n-1}j_n}.\end{aligned}$$

Now for a general event B determined by X_{T+1}, ..., X_{T+n} we sum over $(j_1,\ldots,j_n)\in B$. $\square$

Examples and remarks. 4.1. In the homogeneous birth and death process (see Example 3.3), what is the distribution of the hitting time $H^{(0)}=\inf\ \{n\geq 0 : X_n=0\}$ (the time to extinction)? In other words, what the probabilities $\mathbb{P}_i(H^{(0)}=k)$ for given i and k? These can be found by calculating the probability-generating function (PGF)

$$\begin{aligned}\phi_i(s)&=\mathbb{E}_i\left(s^{H^{(0)}}\right)=\sum_{0\leq n<\infty}s^n\mathbb{P}_i(H^{(0)}=n)\\&=s^0\mathbb{P}_i(H^{(0)}=0)+s^1\mathbb{P}_i(H^{(0)}=1)+s^2\mathbb{P}_i(H^{(0)}=2)+\ldots\end{aligned}.\qquad(4.2)$$

As we know from IA Probability, the PGF determines the probabilities $\mathbb{P}_i(H^{(0)}=k)$ uniquely.

By the strong Markov property:

$$\phi_i(s)=\big(\phi(s)\big)^i,\quad i\geq 1,\qquad(4.3)$$

where $\phi(s)=\phi_1(s)$. This becomes apparent after the following argument:

a) Given that $X_0=i$, we can only hit 0 if we first hit state $i-1$. Let $H^{i\to i-1}$ denote the corresponding hitting time. After we hit $i-1$, we have to hit $i-2$; let $H^{i-1\to i-2}$ be the respective hitting time. Continuing in this manner till we introduce the hitting time $H^{1\to 0}$, we can write

$$H^{(0)}=H^{i\to i-1}+H^{i-1\to i-2}+\ldots+H^{1\to 0}.\qquad(4.4)$$

b) Therefore,

$$\phi_i(s) = \mathbb{E}_i\left(s^{H^{(0)}}\right) = \mathbb{E}_i\left(s^{H^{i\to i-1}+H^{i-1\to i-2}+\ldots+H^{1\to 0}}\right);$$

by the strong Markov property, this is

$$\mathbb{E}_i\left(s^{H^{i\to i-1}}\right)\mathbb{E}_{i-1}\left(s^{H^{i-1\to i-2}}\right)\cdots\mathbb{E}_1\left(s^{H^{1\to 0}}\right). \tag{4.5}$$

Next, by the homogeneity of the process and the strong Markov property again, each factor equals $\mathbb{E}_1\left(s^{H^{1\to 0}}\right) = \phi_1)s)$, and so

$$(4.5) = \big(\phi_1(s)\big)^i,$$

as required.

Now we calculate ϕ_1, using the (usual) Markov property:

$$\phi_1(s) = (1-p)s + ps\phi_2(s) = (1-p)s + ps\big(\phi_1(s)\big)^2. \tag{4.6}$$

We see that ϕ_1 satisfies the quadratic equation

$$ps\phi^2 - \phi + (1-p)s = 0,$$

with roots

$$\frac{1}{2ps}\left(1 \pm \sqrt{1-4p(1-p)s^2}\right), \quad 0 < s < 1. \tag{4.7}$$

We choose the $-$ sign, as the limit $\lim\limits_{s\to 0}\psi_1(s)$ gives $\mathbb{P}_1(H^{(0)} = 0)$, the zero-order coefficient in the right-hand-side of (4.2). But $\mathbb{P}_1(H^{(0)} = 0) = 0$: we can't hit 0 from 1 in zero steps!

Thus, the answer is

$$\phi_i(s) = \left(\frac{1-\sqrt{1-4p(1-p)s^2}}{2ps}\right)^i, \; i \geq 1. \tag{4.7}$$

A couple of useful properties of PGF $\phi_1(s)$ is recollected below. First, near $s = 0$:

$$\phi_1(s)\Big|_{s\sim 0} \sim \frac{1-(1-2p(1-p)s^2}{2ps} \sim qs,$$

with $\phi_1(0) = 0 (= \mathbb{P}_1(H^{(0)} = 0))$, as was noted above. Second, at $s = 1$:

$$\begin{aligned}\phi_1(s) &= \frac{1-\sqrt{1-4p(1-p)}}{2p} = \frac{1-\sqrt{(1-2p)^2}}{2p}\\ &= \frac{1-|1-2p|}{2p} = \begin{cases} 1, & p \le 1/2,\\ (1-p)/p, & p > 1/2.\end{cases}\end{aligned}$$

On the other hand,

$$\begin{aligned}\phi_1(1) &= 1^0\mathbb{P}_1(H^{(0)} = 0) + 1^1\mathbb{P}_1(H^{(0)} = 1) + 1^2\mathbb{P}_1(H^{(0)} = 2) + \dots\\ &= \mathbb{P}_1(H^{(0)} < +\infty),\end{aligned} \tag{4.8}$$

which agrees with previously established equations (3.11).

Next,

$$\begin{aligned}\phi_1'(1) &= 0 \cdot \mathbb{P}_1(H^{(0)} = 0) + 1 \cdot \mathbb{P}_1(H^{(0)} = 1) + 2 \cdot \mathbb{P}_1(H^{(0)} = 2) + \dots\\ &= \mathbb{E}_1 H^{(0)}.\end{aligned}$$

But from Eqn (3.18) we know that $\mathbb{E}_1 H^{(0)} = \begin{cases} 1/(1-2p), & p < 1/2,\\ +\infty, & p \ge 1/2.\end{cases}$

Therefore,

$$\phi_1'(1) = \lim_{s \to 1-} \phi_1'(s) = \begin{cases} \dfrac{1}{1-2p}, & p < \dfrac{1}{2},\\ +\infty, & p \ge \dfrac{1}{2}.\end{cases} \tag{4.9}$$

5.2. An important application of the strong Markov property is when you observe the chain only at certain times, for example, when it jumps, i.e., changes its states (when $X_{n+1} \neq X_n$) or enters a subset $J \subset I$ (i.e., $X_n \in J$). The new chain is formally described by introducing the sequence of observation times, viz.

$$\widehat{T}_0^{\mathrm{j}} = \inf\{n > 0 : \ X_n \neq X_{n-1}\}, \text{ or } T_0^J = \inf\{n \ge 0 : \ X_n \in J\}, \tag{4.10}$$

and

$$\widehat{T}_{m+1}^{\mathrm{j}} = \inf\{n > \widehat{T}_m^{\mathrm{j}} : \ X_n \neq X_{n-1}\}, \text{ or } T_{m+1}^J = \inf\{n > T_m^J : \ X_n \in J\}. \tag{4.11}$$

Then the chain $(Y_n, n \geq 0)$ is defined by $Y_n = X_{\widehat{T}^{\mathrm{j}}_m}$ and the chain $(X^J_n, n \geq 0)$ by $X^J_n = X_{T^J_m}$.

In both examples, each $\widehat{T}^{\mathrm{j}}_m$ and T^J_m is a stopping time. The strong Markov property then gurantees that both (Y_n) and (X^J_n) are indeed Markov chains: (Y_n) is called the *jump chain* and (X^J_n) a partially observed chain generated by (X_n).

The transistion probabilities for the new chains are straightforward. Let $P = (p_{ij})$ be the transition matrix of the original chain (X_n). Then, in the jump chain (Y_n):

$$\widehat{p}_{ij} = \begin{cases} \dfrac{p_{ij}}{1 - p_{ii}}, & i \neq j, \\ 0, & i = j, \end{cases} \qquad i, j \in I, \tag{4.12}$$

and in the partially observed chain (X^J_n):

$$p^J_{ij} = p_{ij} + \sum_{k \geq 1} \sum_{j_1, \ldots, j_k \in I \setminus J} p_{ij_1} \cdots p_{j_k j}, \quad \text{for} \quad i, j \in J \tag{4.13}$$

(index k in (4.13) indicates the number of times chain (X_n) was outside J before returning to J (more precisely, to state $j \in J$).