6

Appendix: probability and measure

Section 6.1 contains some reminders about countable sets and the discrete version of measure theory. For much of the book we can do without explicit mention of more general aspects of measure theory, except an elementary understanding of Riemann integration or Lebesgue measure. This is because the state-space is at worst countable. The proofs we have given may be read on two levels, with or without a measure-theoretic background. When interpreted in terms of measure theory, the proofs are intended to be rigorous. The basic framework of measure and probability is reviewed in Sections 6.2 and 6.3. Two important results of measure theory, the monotone convergence theorem and Fubini's theorem, are needed a number of times: these are discussed in Section 6.4. One crucial result which we found impossible to discuss convincingly without measure theory is the strong Markov property for continuous-time chains. This is proved in Section 6.5. Finally, in Section 6.6, we discuss a general technique for determining probability measures and independence in terms of π-systems, which are often more convenient than σ-algebras.

6.1 Countable sets and countable sums

A set I is *countable* if there is a bijection $f : \{1, \dots , n\} \to I$ for some $n \in \mathbb{N}$, or a bijection $f : \mathbb{N} \to I$. In either case we can *enumerate* all the elements of I

$$f_1, f_2, f_3, \dots$$

https://doi.org/10.1017/CBO9780511810633.008 Published online by Cambridge University Press

where in one case the sequence terminates and in the other it does not. There would have been no loss in generality had we insisted that all our Markov chains had state-space $\mathbb{N}$ or $\{1, \dots, n\}$ for some $n \in \mathbb{N}$: this just corresponds to a particular choice of the bijection f.

Any subset of a countable set is countable. Any finite cartesian product of countable sets is countable, for example $\mathbb{Z}^n$ for any n. Any countable union of countable sets is countable. The set of all subsets of $\mathbb{N}$ is uncountable and so is the set of real numbers $\mathbb{R}$.

We need the following basic fact.

Lemma 6.1.1. *Let I be a countably infinite set and let $\lambda_i \geq 0$ for all $i \in I$. Then, for any two enumerations of I*

$$i_1, i_2, i_3, \dots ,$$
$$j_1, j_2, j_3, \dots$$

we have

$$\sum_{n=1}^{\infty} \lambda_{i_n} = \sum_{n=1}^{\infty} \lambda_{j_n}.$$

Proof. Given any $N \in \mathbb{N}$ we can find $M \geq N$ and $N' \geq M$ such that

$$\{i_1, \dots , i_N\} \subseteq \{j_1, \dots , j_M\} \subseteq \{i_1, \dots , i_{N'}\}.$$

Then

$$\sum_{n=1}^{N} \lambda_{i_n} \leq \sum_{n=1}^{M} \lambda_{j_n} \leq \sum_{n=1}^{N'} \lambda_{i_n}$$

and the result follows on letting $N \to \infty$. □

Since the value of the sum does not depend on the enumeration we are justified in using a notation which does not specify an enumeration and write simply

$$\sum_{i \in I} \lambda_i.$$

More generally, if we allow λ_i to take negative values, then we can set

$$\sum_{i \in I} \lambda_i = \left(\sum_{i \in I^+} \lambda_i \right) - \left(\sum_{i \in I^-} (-\lambda_i) \right)$$

where

$$I^{\pm} = \{i \in I : \pm\lambda_i \geq 0\},$$

https://doi.org/10.1017/CBO9780511810633.008 Published online by Cambridge University Press

allowing that the sum over I is undefined when the sums over I^+ and I^- are both infinite. There is no difficulty in showing for $\lambda_i, \mu_i \geq 0$ that

$$\sum_{i\in I}(\lambda_i + \mu_i) = \sum_{i\in I}\lambda_i + \sum_{i\in I}\mu_i.$$

By induction, for any finite set J and for $\lambda_{ij} \geq 0$, we have

$$\sum_{i\in I}\left(\sum_{j\in J}\lambda_{ij}\right) = \sum_{j\in J}\left(\sum_{i\in I}\lambda_{ij}\right).$$

The following two results on sums are simple versions of fundamental results for integrals. We take the opportunity to prove these simple versions in order to convey some intuition relevant to the general case.

Lemma 6.1.2 (Fubini's theorem – discrete case). *Let I and J be countable sets and let $\lambda_{ij} \geq 0$ for all $i \in I$ and $j \in J$. Then*

$$\sum_{i\in I}\left(\sum_{j\in J}\lambda_{ij}\right) = \sum_{j\in J}\left(\sum_{i\in I}\lambda_{ij}\right).$$

Proof. Let $j_1, j_2, j_3, \ldots$ be an enumeration of J. Then

$$\sum_{i\in I}\left(\sum_{j\in J}\lambda_{ij}\right) \geq \sum_{i\in I}\left(\sum_{k=1}^{n}\lambda_{ij_k}\right) = \sum_{k=1}^{n}\left(\sum_{i\in I}\lambda_{ij_k}\right) \uparrow \sum_{j\in J}\left(\sum_{i\in I}\lambda_{ij}\right)$$

as $n \to \infty$. Hence

$$\sum_{i\in I}\left(\sum_{j\in J}\lambda_{ij}\right) \geq \sum_{j\in J}\left(\sum_{i\in I}\lambda_{ij}\right)$$

and the result follows by symmetry. $\square$

Lemma 6.1.3 (Monotone convergence – discrete case). *Suppose for each $i \in I$ we are given an increasing sequence $(\lambda_i(n))_{n\geq 0}$ with limit λ_i, and that $\lambda_i(n) \geq 0$ for all i and n. Then*

$$\sum_{i\in I}\lambda_i(n) \uparrow \sum_{i\in I}\lambda_i \quad \text{as } n \to \infty.$$

https://doi.org/10.1017/CBO9780511810633.008 Published online by Cambridge University Press

Proof. Set $\delta_i(1) = \lambda_i(1)$ and for $n \geq 2$ set

$$\delta_i(n) = \lambda_i(n) - \lambda_i(n-1).$$

Then $\delta_i(n) \geq 0$ for all i and n, so as $n \to \infty$, by Fubini's theorem

$$\begin{aligned}\sum_{i\in I} \lambda_i(n) &= \sum_{i\in I} \left(\sum_{k=1}^{n} \delta_i(k) \right) \\ &= \sum_{k=1}^{n} \left(\sum_{i\in I} \delta_i(k) \right) \uparrow \sum_{k=1}^{\infty} \left(\sum_{i\in I} \delta_i(k) \right) \\ &= \sum_{i\in I} \left(\sum_{k=1}^{\infty} \delta_i(k) \right) = \sum_{i\in I} \lambda_i. \end{aligned}$$

□

6.2 Basic facts of measure theory

We state here for easy reference the basic definitions and results of measure theory. Let E be a set. A *σ-algebra* $\mathcal{E}$ on E is a set of subsets of E satisfying

(i) $\emptyset \in \mathcal{E}$;

(ii) $A \in \mathcal{E} \Rightarrow A^c \in \mathcal{E}$;

(iii) $(A_n \in \mathcal{E}, n \in \mathbb{N}) \Rightarrow \bigcup_n A_n \in \mathcal{E}$.

Here A^c denotes the complement $E \backslash A$ of A in E. Thus $\mathcal{E}$ is closed under countable set operations. The pair $(E, \mathcal{E})$ is called a *measurable space*. A *measure* μ on $(E, \mathcal{E})$ is a function $\mu : \mathcal{E} \to [0, \infty]$ which has the following *countable additivity* property:

$$(A_n \in \mathcal{E}, n \in \mathbb{N}, A_n \text{ disjoint}) \Rightarrow \mu\left(\bigcup_n A_n\right) = \sum_n \mu(A_n).$$

The triple $(E, \mathcal{E}, \mu)$ is called a *measure space*. If there exist sets $E_n \in \mathcal{E}$, $n \in \mathbb{N}$ with $\bigcup_n E_n = E$ and $\mu(E_n) < \infty$ for all n, then we say μ is *σ-finite*.

Example 6.2.1

Let I be a countable set and denote by $\mathcal{I}$ the set of all subsets of I. Recall that $\lambda = (\lambda_i : i \in I)$ is a measure in the sense of Section 1.1 if $\lambda_i \in [0, \infty)$ for all i. For such λ we obtain a measure on the measurable space $(I, \mathcal{I})$ by setting

$$\lambda(A) = \sum_{i\in A} \lambda_i.$$

In fact, we obtain in this way all σ-finite measures μ on $(I, \mathcal{I})$.

Example 6.2.2

Let $\mathcal{A}$ be any set of subsets of E. The set of all subsets of E is a σ-algebra containing $\mathcal{A}$. The intersection of any collection of σ-algebras is again a σ-algebra. The collection of σ-algebras containing $\mathcal{A}$ is therefore non-empty and its intersection is a σ-algebra $\sigma(\mathcal{A})$, which is called the σ-algebra *generated by* $\mathcal{A}$.

Example 6.2.3

In the preceding example take $E = \mathbb{R}$ and

$$\mathcal{A} = \{(a, b) : a, b \in \mathbb{R}, a < b\}.$$

The σ-algebra $\mathcal{B}$ generated by $\mathcal{A}$ is called the *Borel σ-algebra* of $\mathbb{R}$. It can be shown that there is a unique measure μ on $(\mathbb{R}, \mathcal{B})$ such that

$$\mu(a, b) = b - a \qquad \text{for all } a, b.$$

This measure μ is called *Lebesgue measure*.

Let $(E_1, \mathcal{E}_1)$ and $(E_2, \mathcal{E}_2)$ be measurable spaces. A function $f : E_1 \to E_2$ is *measurable* if $f^{-1}(A) \in \mathcal{E}_1$ whenever $A \in \mathcal{E}_2$. When the range $E_2 = \mathbb{R}$ we take $\mathcal{E}_2 = \mathcal{B}$ by default. When the range E_2 is a countable set I we take $\mathcal{E}_2$ to be the set of all subsets $\mathcal{I}$ by default.

Let $(E, \mathcal{E})$ be a measurable space. We denote by $m\mathcal{E}$ the set of measurable functions $f : E \to \mathbb{R}$. Then $m\mathcal{E}$ is a vector space. We denote by $m\mathcal{E}^+$ the set of measurable functions $f : E \to [0, \infty]$, where we take on $[0, \infty]$ the σ-algebra generated by the open intervals (a, b). Then $m\mathcal{E}^+$ is a cone

$$(f, g \in m\mathcal{E}^+, \alpha, \beta \geq 0) \Rightarrow \alpha f + \beta g \in m\mathcal{E}^+.$$

Also, $m\mathcal{E}^+$ is closed under countable suprema:

$$(f_i \in m\mathcal{E}^+, i \in I) \Rightarrow \sup_i f_i \in m\mathcal{E}^+.$$

It follows that, for a sequence of functions $f_n \in m\mathcal{E}^+$, both $\limsup_n f_n$ and $\liminf_n f_n$ are in $m\mathcal{E}^+$, and so is $\lim_n f_n$ when this exists. It can be shown that there is a unique map $\widetilde{\mu} : m\mathcal{E}^+ \to [0, \infty]$ such that

(i) $\widetilde{\mu}(1_A) = \mu(A)$ for all $A \in \mathcal{E}$;

(ii) $\widetilde{\mu}(\alpha f + \beta g) = \alpha\widetilde{\mu}(f) + \beta\widetilde{\mu}(g)$ for all $f, g \in m\mathcal{E}^+, \alpha, \beta \geq 0$;

(iii) $(f_n \in m\mathcal{E}^+, n \in \mathbb{N}) \Rightarrow \widetilde{\mu}(\sum_n f_n) = \sum_n \widetilde{\mu}(f_n)$.

For $f \in m\mathcal{E}$, set $f^{\pm} = (\pm f) \vee 0$, then $f^+, f^- \in m\mathcal{E}^+$, $f = f^+ - f^-$ and $|f| = f^+ + f^-$. If $\widetilde{\mu}(|f|) < \infty$ then f is said to be *integrable* and we set

$$\widetilde{\mu}(f) = \mu(f^+) - \mu(f^-).$$

We call $\widetilde{\mu}(f)$ the *integral* of f. It is conventional to drop the tilde and denote the integral by one of the following alternative notations:

$$\mu(f) = \int_E f d\mu = \int_{x \in E} f(x)\mu(dx).$$

In the case of Lebesgue measure μ, one usually writes simply

$$\int_{x \in \mathbb{R}} f(x)dx.$$

6.3 Probability spaces and expectation

The basic apparatus for modelling randomness is a *probability space* $(\Omega, \mathcal{F}, \mathbb{P})$. This is simply a measure space with total mass $\mathbb{P}(\Omega) = 1$. Thus $\mathcal{F}$ is a σ-algebra of subsets of Ω and $\mathbb{P} : \mathcal{F} \to [0, 1]$ satisfies

(i) $\mathbb{P}(\Omega) = 1$;

(ii) $\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2)$ for A_1, A_2 disjoint;

(iii) $\mathbb{P}(A_n) \uparrow \mathbb{P}(A)$ whenever $A_n \uparrow A$.

In (iii) we write $A_n \uparrow A$ to mean $A_1 \subseteq A_n \subseteq \dots$ with $\bigcup_n A_n = A$. A measurable function X defined on $(\Omega, \mathcal{F})$ is called a *random variable*. We use random variables $Y : \Omega \to \mathbb{R}$ to model random quantities, where for a Borel set $B \subseteq \mathbb{R}$ the probability that $Y \in B$ is given by

$$\mathbb{P}(Y \in B) = \mathbb{P}\big(\{\omega : Y(\omega) \in B\}\big).$$

Similarly, given a countable state-space I, a random variable $X : \Omega \to I$ models a random state, with *distribution*

$$\lambda_i = \mathbb{P}(X = i) = \mathbb{P}\big(\{\omega : X(\omega) = i\}\big).$$

To every non-negative or integrable real-valued random variable Y is associated an average value or *expectation* $\mathbb{E}(Y)$, which is the integral of Y with respect to $\mathbb{P}$. Thus we have

(i) $\mathbb{E}(1_A) = \mathbb{P}(A)$ for $A \in \mathcal{F}$;

(ii) $\mathbb{E}(\alpha X + \beta Y) = \alpha\mathbb{E}(X) + \beta\mathbb{E}(Y)$ for $X, Y \in m\mathcal{F}^+$, $\alpha, \beta \geq 0$;

(iii) $(Y_n \in m\mathcal{F}^+, n \in \mathbb{N}, Y_n \uparrow Y) \Rightarrow \mathbb{E}(Y_n) \uparrow \mathbb{E}(Y)$.

When X is a random variable with values in I and $f : I \to [0, \infty]$ the expectation of $Y = f(X) = f \circ X$ is given explicitly by

$$\mathbb{E}\big(f(X)\big) = \sum_{i \in I} \lambda_i f_i$$

where λ is the distribution of X. For a real-valued random variable Y the probabilities are sometimes given by a measurable *density* function ρ in terms of Lebesgue measure:

$$\mathbb{P}(Y \in B) = \int_B \rho(y)dy.$$

Then for any measurable function $f : \mathbb{R} \to [0, \infty]$ there is an explicit formula

$$\mathbb{E}\big(f(Y)\big) = \int_{\mathbb{R}} f(y)\rho(y)dy.$$

6.4 Monotone convergence and Fubini's theorem

Here are the two theorems from measure theory that come into play in the main text. First we shall state the theorems, then we shall discuss some places where they are used. Proofs may be found, for example, in *Probability with Martingales* by D. Williams (Cambridge University Press, 1991).

Theorem 6.4.1 (Monotone convergence). *Let $(E, \mathcal{E}, \mu)$ be a measure space and let $(f_n)_{n \geq 1}$ be a sequence of non-negative measurable functions. Then, as $n \to \infty$*

$$\big(f_n(x) \uparrow f(x) \text{ for all } x \in E\big) \Rightarrow \mu(f_n) \uparrow \mu(f).$$

Theorem 6.4.2 (Fubini's theorem). *Let $(E_1, \mathcal{E}_1, \mu_1)$ and $(E_2, \mathcal{E}_2, \mu_2)$ be two σ-finite measure spaces. Suppose that $f : E_1 \times E_2 \to [0, \infty]$ satisfies*

(i) *$x \mapsto f(x, y) : E_1 \to [0, \infty]$ is $\mathcal{E}_1$ measurable for all $y \in E_2$;*

(ii) *$y \mapsto \int_{x \in E_1} f(x, y)\mu_1(dx) : E_2 \to [0, \infty]$ is $\mathcal{E}_2$ measurable.*

Then

(a) *$y \mapsto f(x, y) : E_2 \to [0, \infty]$ is $\mathcal{E}_2$ measurable for all $x \in E_1$;*

(b) *$x \mapsto \int_{y \in E_2} f(x, y)\mu_2(dy) : E_1 \to [0, \infty]$ is $\mathcal{E}_1$ measurable;*

(c) $\displaystyle\int_{x \in E_1} \left(\int_{y \in E_2} f(x, y)\mu_2(dy)\right)\mu_1(dx) = \int_{y \in E_2} \left(\int_{x \in E_1} f(x, y)\mu_1(dx)\right)\mu_2(dy).$

https://doi.org/10.1017/CBO9780511810633.008 Published online by Cambridge University Press

The measurability conditions in the above theorems rarely need much consideration. They are powerful results and very easy to use. There is an equivalent formulation of monotone convergence in terms of sums: for non-negative measurable functions g_n we have

$$\mu\Big(\sum_{n=1}^{\infty} g_n\Big) = \sum_{n=1}^{\infty} \mu(g_n).$$

To see this just take $f_n = g_1 + \dots + g_n$. This form of monotone convergence has already appeared in Section 6.2 as a defining property of the integral. This is also a special case of Fubini's theorem, provided that $(E, \mathcal{E}, \mu)$ is σ-finite: just take $E_2 = \{1, 2, 3, \dots\}$ and $\mu_2(\{n\}) = 1$ for all n.

We used monotone convergence in Theorem 1.10.1 to see that for a non-negative random variable Y we have

$$\mathbb{E}(Y) = \lim_{N\to\infty} \mathbb{E}(Y \wedge N).$$

We used monotone convergence in Theorem 2.3.2 to see that for random variables $S_n \geq 0$ we have

$$\mathbb{E}\Big(\sum_n S_n\Big) = \sum_n \mathbb{E}(S_n)$$

and

$$\begin{aligned}\mathbb{E}\Big(\exp\Big\{-\sum_n S_n\Big\}\Big) &= \mathbb{E}\Big(\lim_{N\to\infty} \exp\Big\{-\sum_{n\leq N} S_n\Big\}\Big)\\ &= \lim_{N\to\infty} \mathbb{E}\Big(\exp\Big\{-\sum_{n\leq N} S_n\Big\}\Big).\end{aligned}$$

In the last application convergence is not monotone increasing but monotone decreasing. But if $0 \leq X_n \leq Y$ and $X_n \downarrow X$ then $Y - X_n \uparrow Y - X$. So $\mathbb{E}(Y - X_n) \uparrow \mathbb{E}(Y - X)$ and if $\mathbb{E}(Y) < \infty$ we can deduce $\mathbb{E}(X_n) \downarrow \mathbb{E}(X)$.

Fubini's theorem is used in Theorem 3.4.2 to see that

$$\int_0^{\infty} p_{ii}(t)dt = \int_0^{\infty} \mathbb{E}_i\Big(1_{\{X_t=i\}}\Big)dt = \mathbb{E}_i \int_0^{\infty} 1_{\{X_t=i\}}dt.$$

Thus we have taken $(E_1, \mathcal{E}_1, \mu_1)$ to be $[0, \infty)$ with Lebesgue measure and $(E_2, \mathcal{E}_2, \mu_2)$ to be the probability space with the measure $\mathbb{P}_i$.

6.5 Stopping times and the strong Markov property

The strong Markov property for continuous-time Markov chains cannot properly be understood without measure theory. The problem lies with

https://doi.org/10.1017/CBO9780511810633.008 Published online by Cambridge University Press

the notion of 'depending only on', which in measure theory is made precise as measurability with respect to some σ-algebra. Without measure theory the statement that a set A depends only on $(X_s : s \leq t)$ does not have a precise meaning. Of course, if the dependence is reasonably explicit we can exhibit it, but then, in general, in what terms would you require the dependence to be exhibited? So in this section we shall give a precise measure-theoretic account of the strong Markov property.

Let $(X_t)_{t\geq 0}$ be a right-continuous process with values in a countable set I. Denote by $\mathcal{F}_t$ the σ-algebra generated by $\{X_s : s \leq t\}$, that is to say, by all sets $\{X_s = i\}$ for $s \leq t$ and $i \in I$. We say that a random variable T with values in $[0, \infty]$ is a *stopping time* of $(X_t)_{t\geq 0}$ if $\{T \leq t\} \in \mathcal{F}_t$ for all $t \geq 0$. Note that this certainly implies

$$\{T < t\} = \bigcup_n \{T \leq t - 1/n\} \in \mathcal{F}_t \quad \text{for all} \quad t \geq 0.$$

We define for stopping times T

$$\mathcal{F}_T = \big\{A \in \mathcal{F} : A \cap \{T \leq t\} \in \mathcal{F}_t \text{ for all } t \geq 0\big\}.$$

This turns out to be the correct way to make precise the notion of sets which 'depend only on $\{X_t : t \leq T\}$'.

Lemma 6.5.1. *Let S and T be stopping times of $(X_t)_{t\geq 0}$. Then both X_T and $\{S \leq T\}$ are $\mathcal{F}_T$-measurable.*

Proof. Since $(X_t)_{t\geq 0}$ is right-continuous, on $\{T < t\}$ there exists an $n \geq 0$ such that for all $m \geq n$, for some $k \geq 1$, $(k-1)2^{-m} \leq T < k2^{-m} \leq t$ and $X_{k2^{-m}} = X_T$. Hence

$$\begin{aligned}
&\{X_T = i\} \cap \{T \leq t\} \\
&\qquad = \big(\{X_t = i\} \cap \{T = t\}\big) \\
&\qquad\quad \cup \left(\bigcup_{n=0}^{\infty} \bigcap_{m=n}^{\infty} \bigcup_{k=1}^{[2^m t]} \{X_{k2^{-m}} = i\} \cap \{(k-1)2^{-m} \leq T < k2^{-m}\}\right) \in \mathcal{F}_t
\end{aligned}$$

so X_T is $\mathcal{F}_T$-measurable.

We have

$$\{S > T\} \cap \{T \leq t\} = \bigcup_{s \in \mathbb{Q}, s \leq t} \big(\{T \leq s\} \cap \{S > s\}\big) \in \mathcal{F}_t$$

so $\{S > T\} \in \mathcal{F}_T$, and so $\{S \leq T\} \in \mathcal{F}_T$. $\square$

Lemma 6.5.2. *For all $m \geq 0$, the jump time J_m is a stopping time of $(X_t)_{t\geq 0}$.*

Proof. Obviously, $J_0 = 0$ is a stopping time. Assume inductively that J_m is a stopping time. Then

$$\{J_{m+1} \leq t\} = \bigcup_{s\in\mathbb{Q}, s\leq t} \{J_m \leq s\} \cap \{X_s \neq X_{J_m}\} \in \mathcal{F}_t$$

for all $t \geq 0$, so J_{m+1} is a stopping time and the induction proceeds. $\square$

We denote by $\mathcal{G}_m$ the σ-algebra generated by $Y_0, \dots, Y_m$ and $S_1, \dots, S_m$, that is, by events of the form $\{Y_k = i\}$ for $k \leq m$ and $i \in I$ or of the form $\{S_k > s\}$ for $k \leq m$ and $s > 0$.

Lemma 6.5.3. *Let T be a stopping time of $(X_t)_{t\geq 0}$ and let $A \in \mathcal{F}_T$. Then for all $m \geq 0$ there exist a random variable T_m and a set A_m, both measurable with respect to $\mathcal{G}_m$, such that $T = T_m$ and $1_A = 1_{A_m}$ on $\{T < J_{m+1}\}$.*

Proof. Fix $t \geq 0$ and consider

$$\mathcal{A}_t = \{A \in \mathcal{F}_t : A \cap \{t < J_{m+1}\} = A_m \cap \{t < J_{m+1}\} \quad \text{for some } A_m \in \mathcal{G}_m\}.$$

Since $\mathcal{G}_m$ is a σ-algebra, so is $\mathcal{A}_t$. For $s \leq t$ we have

$$\begin{aligned} &\{X_s = i\} \cap \{t < J_{m+1}\} \\ &= \left(\bigcup_{k=0}^{m-1} \{Y_k = i, J_k \leq s < J_{k+1}\} \cup \{Y_m = i, J_m \leq s\} \right) \cap \{t < J_{m+1}\} \end{aligned}$$

so $\{X_s = i\} \in \mathcal{A}_t$. Since these sets generate $\mathcal{F}_t$, this implies that $\mathcal{A}_t = \mathcal{F}_t$.

For T a stopping time and $A \in \mathcal{F}_T$ we have $B(t) := \{T \leq t\} \in \mathcal{F}_t$ and $A(t) := A \cap \{T \leq t\} \in \mathcal{F}_t$ for all $t \geq 0$. So we can find $B_m(t)$, $A_m(t) \in \mathcal{G}_m$ such that

$$\begin{aligned} B(t) \cap \{T < J_{m+1}\} &= B_m(t) \cap \{T < J_{m+1}\}, \\ A(t) \cap \{T < J_{m+1}\} &= A_m(t) \cap \{T < J_{m+1}\}. \end{aligned}$$

Set

$$T_m = \sup_{t\in\mathbb{Q}} t 1_{B_m(t)}, \quad A_m = \bigcup_{t\in\mathbb{Q}} A_m(t)$$

then T_m and A_m are $\mathcal{G}_m$-measurable and

$$\begin{aligned} T_m 1_{\{T<J_{m+1}\}} &= \sup_{t\in\mathbb{Q}} t 1_{B_m(t)\cap\{T<J_{m+1}\}} \\ &= \left(\sup_{t\in\mathbb{Q}} t 1_{\{T\leq t\}}\right) 1_{\{T<J_{m+1}\}} = T 1_{\{T<J_{m+1}\}} \end{aligned}$$

https://doi.org/10.1017/CBO9780511810633.008 Published online by Cambridge University Press

and

$$\begin{aligned} A_m \cap \{T < J_{m+1}\} &= \bigcup_{t \in \mathbb{Q}} A_m(t) \cap \{T < J_{m+1}\} \\ &= \bigcup_{t \in \mathbb{Q}} \big(A \cap \{T \le t\}\big) \cap \{T < J_{m+1}\} = A \cap \{T < J_{m+1}\} \end{aligned}$$

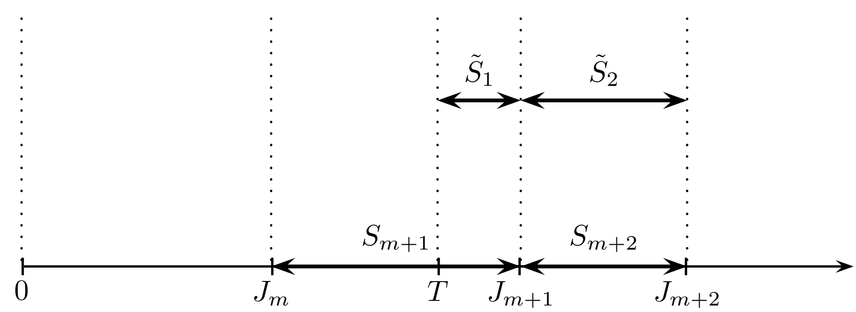
as required. □

Theorem 6.5.4 (Strong Markov property). *Let $(X_t)_{t\ge 0}$ be Markov(λ, Q) and let T be a stopping time of $(X_t)_{t\ge0}$. Then, conditional on $T < \zeta$ and $X_T = i$, $(X_{T+t})_{t\ge0}$ is Markov(δ_i, Q) and independent of $\mathcal{F}_T$.*

Proof. On $\{T < \zeta\}$ set $\widetilde{X}_t = X_{T+t}$ and denote by $(\widetilde{Y}_n)_{n\ge0}$ the jump chain and by $(\widetilde{S}_n)_{n\ge1}$ the holding times of $(\widetilde{X}_t)_{t\ge0}$. We have to show that, for all $A \in \mathcal{F}_T$, all $i_0, \dots, i_n \in I$ and all $s_1, \dots, s_n \ge 0$

$$\begin{aligned} &\mathbb{P}\big(\{\widetilde{Y}_0 = i_0, \dots, \widetilde{Y}_n = i_n, \widetilde{S}_1 > s_1, \dots, \widetilde{S}_n > s_n\} \cap A \cap \{T < \zeta\} \cap \{X_T = i\}\big) \\ &\quad = \mathbb{P}_i(Y_0 = i_0, \dots, Y_n = i_n, S_1 > s_1, \dots, S_n > s_n) \\ &\qquad \times \mathbb{P}\big(A \cap \{T < \zeta\} \cap \{X_T = i\}\big). \end{aligned}$$

It suffices to prove this with $\{T < \zeta\}$ replaced by $\{J_m \le T < J_{m+1}\}$ for all $m \ge 0$ and then sum over m. By Lemmas 6.5.1 and 6.5.2, $\{J_m \le T\} \cap \{X_T = i\} \in \mathcal{F}_T$ so we may assume without loss of generality that $A \subseteq \{J_m \le T\} \cap \{X_T = i\}$. By Lemma 6.5.3 we can write $T = T_m$ and $1_A = 1_{A_m}$ on $\{T < J_{m+1}\}$, where T_m and A_m are $\mathcal{G}_m$-measurable.



On $\{J_m \le T < J_{m+1}\}$ we have, as shown in the diagram

$$\widetilde{Y}_n = Y_{m+n}, \quad \widetilde{S}_1 = S_{m+1} - (T - J_m), \quad \widetilde{S}_n = S_{m+n} \text{ for } n \ge 2.$$

https://doi.org/10.1017/CBO9780511810633.008 Published online by Cambridge University Press

Now, conditional on $Y_m = i$, S_{m+1} is independent of $\mathcal{G}_m$ and hence of $T_m - J_m$ and A_m and, by the memoryless property of the exponential

$$\mathbb{P}\big(S_{m+1} > s_1 + (T_m - J_m) \mid Y_m = i, S_{m+1} > T_m - J_m\big) = e^{-q_i s_1} = \mathbb{P}_i(S_1 > s_1).$$

Hence, by the Markov property of the jump chain

$$\begin{aligned}
&\mathbb{P}\big(\{\widetilde{Y}_0 = i_0, \dots, \widetilde{Y}_n = i_n, \\
&\qquad \widetilde{S}_1 > s_1, \dots, \widetilde{S}_n > s_n\} \cap A \cap \{J_m \le T < J_{m+1}\} \cap \{X_T = i\}\big) \\
&= \mathbb{P}\big(\{Y_m = i_0, \dots, Y_{m+n} = i_n, S_{m+1} > s_1 + (T_m - J_m), \\
&\qquad S_{m+2} > s_2, \dots, S_{m+n} > s_n\} \cap A_m \cap \{S_{m+1} > T_m - J_m\}\big) \\
&= \mathbb{P}_i(Y_0 = i_0, \dots, Y_n = i_n, \\
&\qquad S_1 > s_1, \dots, S_n > s_n)\mathbb{P}\big(A \cap \{J_m \le T < J_{m+1}\} \cap \{X_T = i\}\big)
\end{aligned}$$

as required. $\square$

6.6 Uniqueness of probabilities and independence of σ-algebras

For both discrete-time and continuous-time Markov chains we have given definitions which specify the probabilities of certain events determined by the process. From these specified probabilities we have often deduced explicitly the values of other probabilities, for example hitting probabilities. In this section we shall show, in measure-theoretic terms, that our definitions determine the probabilities of *all* events depending on the process. The constructive approach we have taken should make this seem obvious, but it is illuminating to see what has to be done.

Let Ω be a set. A *π-system* $\mathcal{A}$ on Ω is a collection of subsets of Ω which is closed under finite intersections; thus

$$A_1, A_2 \in \mathcal{A} \Rightarrow A_1 \cap A_2 \in \mathcal{A}.$$

We denote as usual by $\sigma(\mathcal{A})$ the σ-algebra generated by $\mathcal{A}$. If $\sigma(\mathcal{A}) = \mathcal{F}$ we say that $\mathcal{A}$ *generates* $\mathcal{F}$.

Theorem 6.6.1. *Let $(\Omega, \mathcal{F})$ be a measurable space. Let $\mathbb{P}_1$ and $\mathbb{P}_2$ be probability measures on $(\Omega, \mathcal{F})$ which agree on a π-system $\mathcal{A}$ generating $\mathcal{F}$. Then $\mathbb{P}_1 = \mathbb{P}_2$.*

Proof. Consider

$$\mathcal{D} = \{A \in \mathcal{F} : \mathbb{P}_1(A) = \mathbb{P}_2(A)\}.$$

https://doi.org/10.1017/CBO9780511810633.008 Published online by Cambridge University Press

We have assumed that $\mathcal{A} \subseteq \mathcal{D}$. Moreover, since $\mathbb{P}_1$ and $\mathbb{P}_2$ are probability measures, $\mathcal{D}$ has the following properties:

(i) $\Omega \in \mathcal{D}$;

(ii) $(A, B \in \mathcal{D}$ and $A \subseteq B) \Rightarrow B \backslash A \in \mathcal{D}$;

(iii) $(A_n \in \mathcal{D}, A_n \uparrow A) \Rightarrow A \in \mathcal{D}$.

Any collection of subsets having these properties is called a *d-system*. Since $\mathcal{A}$ generates $\mathcal{F}$, the result now follows from the following lemma. $\square$

Lemma 6.6.2 (Dynkin's π-system lemma). *Let $\mathcal{A}$ be a π-system and let $\mathcal{D}$ be a d-system. Suppose $\mathcal{A} \subseteq \mathcal{D}$. Then $\sigma(\mathcal{A}) \subseteq \mathcal{D}$.*

Proof. Any intersection of d-systems is again a d-system, so we may without loss assume that $\mathcal{D}$ is the smallest d-system containing $\mathcal{A}$. You may easily check that any d-system which is also a π-system is necessarily a σ-algebra, so it suffices to show $\mathcal{D}$ is a π-system. This we do in two stages.

Consider first

$$\mathcal{D}_1 = \{A \in \mathcal{D} : A \cap B \in \mathcal{D} \text{ for all } B \in \mathcal{A}\}.$$

Since $\mathcal{A}$ is a π-system, $\mathcal{A} \subseteq \mathcal{D}_1$. You may easily check that $\mathcal{D}_1$ is a d-system – because $\mathcal{D}$ is a d-system. Since $\mathcal{D}$ is the smallest d-system containing $\mathcal{A}$, this shows $\mathcal{D}_1 = \mathcal{D}$.

Next consider

$$\mathcal{D}_2 = \{A \in \mathcal{D} : A \cap B \in \mathcal{D} \text{ for all } B \in \mathcal{D}\}.$$

Since $\mathcal{D}_1 = \mathcal{D}$, $\mathcal{A} \subseteq \mathcal{D}_2$. You can easily check that $\mathcal{D}_2$ is also a d-system. Hence also $\mathcal{D}_2 = \mathcal{D}$. But this shows $\mathcal{D}$ is a π-system. $\square$

The notion of independence used in advanced probability is the independence of σ-algebras. Suppose that $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and $\mathcal{F}_1$ and $\mathcal{F}_2$ are sub-σ-algebras of $\mathcal{F}$. We say that $\mathcal{F}_1$ and $\mathcal{F}_2$ are *independent* if

$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2) \qquad \text{for all } A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2.$$

The usual means of establishing such independence is the following corollary of Theorem 6.6.1.

Theorem 6.6.3. *Let $\mathcal{A}_1$ be a π-system generating $\mathcal{F}_1$ and let $\mathcal{A}_2$ be a π-system generating $\mathcal{F}_2$. Suppose that*

$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2) \qquad \text{for all } A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2.$$

Then $\mathcal{F}_1$ and $\mathcal{F}_2$ are independent.

Proof. There are two steps. First fix $A_2 \in \mathcal{A}_2$ with $\mathbb{P}(A_2) > 0$ and consider the probability measure

$$\widetilde{\mathbb{P}}(A) = \mathbb{P}(A \mid A_2).$$

We have assumed that $\widetilde{\mathbb{P}}(A) = \mathbb{P}(A)$ for all $A \in \mathcal{A}_1$, so, by Theorem 6.6.1, $\widetilde{\mathbb{P}} = \mathbb{P}$ on $\mathcal{F}_1$. Next fix $A_1 \in \mathcal{F}_1$ with $\mathbb{P}(A_1) > 0$ and consider the probability measure

$$\widetilde{\widetilde{\mathbb{P}}}(A) = \mathbb{P}(A \mid A_1).$$

We showed in the first step that $\widetilde{\widetilde{\mathbb{P}}}(A) = \mathbb{P}(A)$ for all $A \in \mathcal{A}_2$, so, by Theorem 6.6.1, $\widetilde{\widetilde{\mathbb{P}}} = \mathbb{P}$ on $\mathcal{F}_2$. Hence $\mathcal{F}_1$ and $\mathcal{F}_2$ are independent. $\square$

We now review some points in the main text where Theorems 6.6.1 and 6.6.3 are relevant.

In Theorem 1.1.1 we showed that our definition of a discrete-time Markov chain $(X_n)_{n\geq 0}$ with initial distribution λ and transition matrix P determines the probabilities of all events of the form

$$\{X_0 = i_0, \ldots, X_n = i_n\}.$$

But subsequently we made explicit calculations for probabilities of events which were not of this form – such as the event that $(X_n)_{n\geq 0}$ visits a set of states A. We note now that the events $\{X_0 = i_0, \ldots, X_n = i_n\}$ form a π-system which generates the σ-algebra $\sigma(X_n : n \geq 0)$. Hence, by Theorem 6.6.1, our definition determines (in principle) the probabilities of all events in this σ-algebra.

In our general discussion of continuous-time random processes in Section 2.2 we claimed that for a right-continuous process $(X_t)_{t\geq 0}$ the probabilities of events of the form

$$\{X_{t_0} = i_0, \ldots, X_{t_n} = i_n\}$$

for all $n \geq 0$ determined the probabilities of all events depending on $(X_t)_{t\geq 0}$. Now events of the form $\{X_{t_0} = i_0, \ldots, X_{t_n} = i_n\}$ form a π-system which generates the σ-algebra $\sigma(X_t : t \geq 0)$. So Theorem 6.6.1 justifies (a precise version) of this claim. The point about right-continuity is that without such an assumption an event such as

$$\{X_t = i \text{ for some } t > 0\}$$

which might reasonably be considered to depend on $(X_t)_{t\geq 0}$, is *not necessarily measurable* with respect to $\sigma(X_t : t \geq 0)$. An argument given in

Section 2.2 shows that this event *is* measurable in the right-continuous case. We conclude that, without some assumption like right-continuity, general continuous-time processes are unreasonable.

Consider now the method of describing a minimal right-continuous process $(X_t)_{t\geq 0}$ via its jump process $(Y_n)_{n\geq 0}$ and holding times $(S_n)_{n\geq 1}$. Let us take $\mathcal{F} = \sigma(X_t : t \geq 0)$. Then Lemmas 6.5.1 and 6.5.2 show that $(Y_n)_{n\geq 0}$ and $(S_n)_{n\geq 1}$ are $\mathcal{F}$-measurable. Thus $\mathcal{G} \subseteq \mathcal{F}$ where

$$\mathcal{G} = \sigma\big((Y_n)_{n\geq 0}, (S_n)_{n\geq 1}\big).$$

On the other hand, for all $i \in I$

$$\{X_t = i\} = \bigcup_{n\geq 0} \{J_n \leq t < J_{n+1}\} \cap \{Y_n = i\} \in \mathcal{G},$$

so also $\mathcal{F} \subset \mathcal{G}$.

A useful π-system generating $\mathcal{G}$ is given by sets of the form

$$B = \{Y_0 = i_0, \dots, Y_n = i_n, S_1 > s_1, \dots, S_n > s_n\}.$$

Our jump chain/holding time definition of the continuous-time chain $(X_t)_{t\geq 0}$ with initial distribution λ and generator matrix Q may be read as stating that, for such events

$$\mathbb{P}(B) = \lambda_{i_0}\pi_{i_0 i_1} \dots \pi_{i_{n-1} i_n} e^{-q_{i_0} s_1} \dots e^{-q_{i_{n-1}} s_n}.$$

Then, by Theorem 6.6.1, this definition determines $\mathbb{P}$ on $\mathcal{G}$ and hence on $\mathcal{F}$.

Finally, we consider the strong Markov property, Theorem 6.5.4. Assume that $(X_t)_{t\geq 0}$ is Markov(λ, Q) and that T is a stopping time of $(X_t)_{t\geq 0}$. On the set $\widetilde{\Omega} = \{T < \zeta\}$ define $\widetilde{X}_t = X_{T+t}$ and let $\widetilde{\mathcal{F}} = \sigma(\widetilde{X}_t : t \geq 0)$; write $(\widetilde{Y}_n)_{n\geq 0}$ and $(\widetilde{S}_n)_{n\geq 0}$ for the jump chain and holding times of $(\widetilde{X}_t)_{t\geq 0}$ and set

$$\widetilde{\mathcal{G}} = \sigma\big((\widetilde{Y}_n)_{n\geq 0}, (\widetilde{S}_n)_{n\geq 0}\big).$$

Thus $\widetilde{\mathcal{F}}$ and $\widetilde{\mathcal{G}}$ are σ-algebras on $\widetilde{\Omega}$, and coincide by the same argument as for $\mathcal{F} = \mathcal{G}$. Set

$$\widetilde{B} = \{\widetilde{Y}_0 = i_0, \dots, \widetilde{Y}_n = i_n, \widetilde{S}_1 > s_1, \dots, \widetilde{S}_n > s_n\}.$$

Then the conclusion of the strong Markov property states that

$$\mathbb{P}(\widetilde{B} \mid T < \zeta, X_T = i) = \mathbb{P}_i(B)$$

with B as above, and that

$$\mathbb{P}(\widetilde{C} \cap A \mid T < \zeta, X_T = i) = \mathbb{P}(\widetilde{C} \mid T < \zeta, X_t = i)\mathbb{P}(A \mid T < \zeta, X_T = i)$$

for all $\widetilde{C} \in \widetilde{\mathcal{F}}$ and $A \in \mathcal{F}_T$. By Theorem 6.6.3 it suffices to prove the independence assertion for the case $\widetilde{C} = \widetilde{B}$, which is what we did in the proof of Theorem 6.5.4.