Applied Stochastic Processes

Vincent Tassion

Spring Semester 2025 (updated: March 11, 2025)

Contents

Acknolwledgment

This notes are based on a first version written by Trevor Winstral.

Chapter 1

Markov Chains: Definitions and construction

Setup:

- S finite or countable set, equipped with the sigma algebra $\mathcal{P}(S)$.
- $(\Omega, \mathcal{F}, \mathbb{P})$ fixed probability space.

Goals:

- Define and motivate Markov Chains via transition probabilities.
- Present the connection with linear algebra and graph theory.
- Simulation of MC from uniforms.

1.1 Transition probabilities and Markov Chains

Definition 1.1. We call *distribution on S* a probability measure μ on S. It is identified with a collection $\mu = (\mu_x)_{x \in S}$ of numbers satisfying

(i) $\forall x \in S \ \mu_x \geq 0$, and

(ii) $\sum_{x \in S} \mu_x = 1$.

Example 1.1 (Uniform distribution). If S is finite, the uniform distribution μ is defined by

$$\forall x \in S \quad \mu_x = \frac{1}{|S|}.$$

.

Example 1.2 (Dirac distribution). For fixed $z \in S$, the Dirac distribution $\delta^z = (\delta^z_x)_{x\in S}$ at z is defined by

$$\forall x \in S \quad \delta^z_x = \begin{cases} 1 & \text{if } x = z, \\ 0 & \text{if } x \neq z. \end{cases}$$

Definition 1.2. A *transition probability* is a collection $P = (p_{x,y})_{x,y\in S}$ such that:

(i) $\forall x, y \in S \quad p_{x,y} \geq 0$, and

(ii) $\forall x \in S \quad \sum_{y\in S} p_{x,y} = 1$.

Equivalently, P is a transition probability if for every fixed $x \in S$, $p_{x,\cdot} := (p_{x,y})_{y\in S}$ is a distribution on S. There are a few different representations of transition probabilities.

Graph representation We can see (S, P) as a weighted oriented graph with the property that the weights leaving any vertex must be nonnegative and sum to 1: the vertex set is S, the edges are all the pairs $(x, y) \in S^2$, and the weights are p_{xy}.

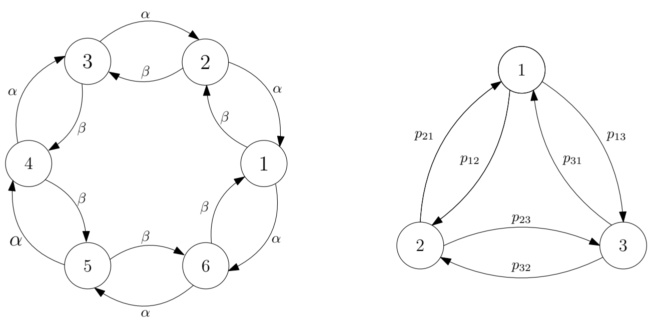


Figure 1.1: Transition probabilities as weighted graphs.

Matrix interpretation Assume S is finite, say $S = \{1, \ldots, N\}$. Then $P = (p_{ij})_{1\leq i,j\leq N}$ is a matrix with nonnegative entries (by Item (i)), and such that each line sums to one (by Item (ii)). Such a matrix is called a stochastic matrix. When S is a general finite set, we can always enumerate its elements to see P as a $|S| \times |S|$ matrix.

Operator interpretation Write $L^\infty(S)$ for the set of bounded function on S, equipped with the norm $\|f\|_\infty = \sup_{x\in S}|f(x)|$. Let P be a transition probability. To every function $f \in L^\infty(S)$, we associate a function Pf defined by

$$\forall x \in S \quad (Pf)_x = \sum_{y\in S} p_{x,y} f_y.$$

Since $|\sum_{y\in S} p_{x,y} f_y| \leq \sum_{y\in S} p_{x,y}|f_y| \leq \|f\|_\infty$, the function Pf is well defined, bounded, and satisfies $\|Pf\|_\infty \leq \|f\|_\infty$. This allows us to identify P with the continuous linear operator $f \mapsto Pf$ acting on $L^\infty(S)$. Items (i) and (ii) correspond to the properties that $P \geq 0$ (i.e. $Pf \geq 0$ for all $f \geq 0$) and $P1 = 1$.

Definition 1.3. Let P be a transition probability, μ a distribution on S. A sequence $X = (X_n)_{n\geq 0}$ of random variables with values in S is a Markov Chain with initial distribution μ and transition probability P if for every $n \geq 0$ and $x_0, \dots, x_n \in S$

$$\boxed{\mathbb{P}(X_0 = x_0, \dots, X_n = x_n) = \mu_{x_0} p_{x_0,x_1} \cdots p_{x_{n-1},x_n}.}$$

In this case, we write $X \sim \mathrm{MC}(\mu, P)$.

Definition 1.4. A sequence $X = (X_n)_{n\geq 0}$ of random variables with values in S is a Markov Chain if there exists a distribution μ and a transition probability P such that $X \sim \mathrm{MC}(\mu, P)$.

1.2 n-Step Transition Probabilities

In this section, we fix a transition probability P on S.

Definition 1.5. Let $n \geq 0$. The n-step transition probability $P^n = (p^{(n)}_{xy})_{x,y\in S}$ associated to P is defined by $p^0_{xy} = 1_{x=y}$ and

$$\boxed{p^{(n)}_{xy} = \sum_{z_1,\dots,z_{n-1}\in S} p_{xz_1} p_{z_1z_2} \cdots p_{z_{n-1}y}.}$$

for every $n \geq 1$, $x, y \in S$.

In the matrix interpretation of transition probabilities, P^n coincides with the n-th power of P. In the operator interpretation, P^n is the n-fold composition of P by itself.

From the Markov Chain perspective, $p_{xy}^{(n)}$ is the probability to move from x to y in n steps, as stated in the following proposition.

Proposition 1.1. *Let $x, y \in S$, $n \geq 0$. If $X \sim \mathrm{MC}(\delta^x, P)$, then*

$$\boxed{p_{x,y}^{(n)} = \mathbb{P}(X_n = y).}$$

Notice that the proposition above implies that P^n is itself a transition probability.

Proof. By first using the definition of the n-step transition probability and then the definition of a Markov Chain, we have

$$\begin{aligned} p_{xy}^{(n)} &= \sum_{z_0, z_1, \ldots, z_{n-1} \in S} \delta_{z_0}^x p_{z_0 z_1} p_{z_1 z_2} \cdots p_{z_{n-1} y} \\ &= \sum_{z_0, \ldots, z_{n-1} \in S} \mathbb{P}(X_0 = z_0, X_1 = z_1 \ldots, X_{n-1} = z_{n-1}, X_n = y) = \mathbb{P}(X_n = y), \end{aligned}$$

where for the last equality we used the disjoint union

$$\{X_n = y\} = \bigcup_{z_0, \ldots, z_{n-1} \in S} \{X_0 = z_0, \ldots, X_{n-1} = z_{n-1}, X_n = y\}.$$

□

1.3 Simulation of a distribution

Lemma 1.2. *Let μ be a distribution on S. There exists a measurable mapping $\Phi_\mu : [0,1] \to S$ such that*

$$\forall x \in S \qquad \mathbb{P}(\Phi_\mu(U) = x) = \mu_x,$$

where $U \sim \mathcal{U}([0,1])$ is an arbitrary uniform random variable.

Proof. We consider a partition $[0,1) = \cup_{x \in S} I_x$, where each I_x is an interval of length μ_x. One way to construct such intervals is to enumerate the elements of the state space $S = \{x_j, \, j \in J\}$ where $J = \{0, \ldots, N\}$ in the finite case and $J = \mathbb{N}$ in the infinite case and define

$$I_{x_j} = [\mu_{x_0} + \ldots + \mu_{x_{j-1}}, \mu_{x_0} + \ldots + \mu_{x_j})$$

for every $j \in J$. Let $\Phi : [0,1] \to S$ defined by $\Phi(u) = x$ if $u \in I_x$, and $\Phi(1) = x_0$ where $x_0 \in S$ is arbitrary. This way, if $U \sim \mathcal{U}([0,1])$ we have for every $x \in S$

$$\mathbb{P}(\Phi(U) = x) = \mathbb{P}(U \in I_x) = \mu_x,$$

as desired. Notice that the chosen value of $\Phi(1)$ does not matter since $U < 1$ almost surely. □

1.4 Simulation of a Markov Chain

Proposition 1.3. *Let μ be a distribution and P a transition probability on S. Let $\Phi_\mu : [0,1] \to S$ as in Lemma 1.2. For each $x \in S$, let $\Phi_x : [0,1] \to S$ measurable such that*

$$\Phi_x(U) \sim p_x. \qquad (\text{ ie } \forall x, y \in S \quad \mathbb{P}(\Phi_x(U)) = p_{x,y}),$$

for arbitrary $U \sim \mathcal{U}([0,1])$. Let $U_0, U_1, \dots$ be a sequence of iid uniform random variables on $[0,1]$. The sequence $X_0, X_1, \dots$ defined inductively by

$$\begin{cases} X_0 = \Phi_\mu(U_0) \\ X_n = \Phi_{X_{n-1}}(U_n) \quad n \geq 1 \end{cases}$$

is a Markov Chain with initial distribution μ and transition probability P.

Proof. Let $n \geq 0$ and $x_0, \dots, x_n \in S$. By first applying the definition of X, and then using independence of the U_i's, we get

$$\begin{aligned} \mathbb{P}(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) &= \mathbb{P}\left(\Phi_\mu(U_0) = x_0, \Phi_{x_0}(U_1) = x_1, \dots, \Phi_{x_{n-1}}(U_n) = x_n\right) \\ &= \mu_{x_0} p_{x_0 x_1} \cdots p_{x_{n-1} x_n}. \end{aligned}$$

□

1.5 One-step Markov Property and Homogeneity.

A central property of Markov Chain is its absence of memory. Furthermore, the chains we are considering are homogeneous in time: if $X_n = x$, the probability to jump from x to y does not depend on the time n. These two properties are formalized as follow:

Proposition 1.4. *Let X be a Markov Chain.*

[1-step Markov Property] *For all $n \geq 0$ and $x_0, \ldots, x_{n+1} \in S$*

$$\boxed{\mathbb{P}\left(X_{n+1} = x_{n+1} \mid X_0 = x_0, \ldots, X_n = x_n\right) = \mathbb{P}\left(X_{n+1} = x_{n+1} \mid X_n = x_n\right).}$$

[Homogeneity] *For all $m, n \geq 0$ and $x, y \in E$*

$$\boxed{\mathbb{P}\left(X_{n+1} = y \mid X_n = x\right) = \mathbb{P}\left(X_{m+1} = y \mid X_m = x\right).}$$

Note: By convention when we write $\mathbb{P}(A \mid B)$ we assume $\mathbb{P}(B) > 0$.

Remark 1.5. Conversely, a sequence of random variables satisfying the 1-Step Markov Property and Homogeneity is a Markov Chain (see Exercise Sheet 2).

Proof. Let $n \geq 0$, $x, y \in S$. By summing over all the possible values for $X_0, \ldots, X_{n-1}$, we have

$$\begin{aligned}
\mathbb{P}\left(X_n = x, X_{n+1} = y\right) &= \sum_{z_0, \ldots, z_{n-1} \in S} \mathbb{P}\left(X_0 = z_0, \ldots, X_{n-1} = z_{n-1}, X_n = x, X_{n+1} = y\right) \\
&= \sum_{z_0, \ldots, z_{n-1} \in S} \mu_{z_0} p_{z_0 z_1} \cdots p_{z_{n-1} x} \cdot p_{xy} \\
&= \mathbb{P}(X_n = x) \cdot p_{xy}.
\end{aligned}$$

By dividing both sides by $\mathbb{P}(X_n = x)$ (assuming it is positive), we obtain

$$\mathbb{P}\left(X_{n+1} = y \,|\, X_n = x\right) = p_{xy}.$$

Since the right hand side does not depend on n, the equation above already establishes Homogeneity.

For the 1-Step Markov Property, let us consider $x_0, \ldots, x_{n+1} \in S$ satisfying

$$\mathbb{P}\left(X_0 = x_0, \ldots, X_n = x_n\right) > 0.$$

By using the definition of a Markov Chain,

$$\begin{aligned}
\mathbb{P}\left(X_{n+1} = x_{n+1} \mid X_0 = x_0, \ldots, X_n = x_n\right) &= \frac{\mathbb{P}\left(X_0 = x_0, \ldots, X_{n+1} = x_{n+1}\right)}{\mathbb{P}\left(X_0 = x_0, \ldots, X_n = x_n\right)} \\
&= \frac{\mu_{x_0} p_{x_0 x_1} \cdots p_{x_n x_{n+1}}}{\mu_{x_0} p_{x_0 x_1} \cdots p_{x_{n-1} x_n}} \\
&= p_{x_n x_{n+1}} = \mathbb{P}\left(X_{n+1} = x_{n+1} \mid X_n = x_n\right).
\end{aligned}$$

□

1.6 Law of a Markov Chain

Motivation: Given μ, P, is there a unique MC(μ, P)? Can two different (μ, P) give rise to two different MC? To formalize and answer such questions we introduce the law of the process.

Definition 1.6. A trajectory is a sequence $x = (x_n)_{n\geq 0}$ of elements of S. We equip the set $S^{\mathbb{N}}$ of trajectories with the product sigma algebra

$$\mathcal{P}(S)^{\otimes\mathbb{N}} = \sigma(\{C_{J,A} \ : \ J \subset \mathbb{N} \text{ finite }, \ A \subset \mathcal{P}(S)^J\})$$

where $C_{J,A} = \{x \in S^{\mathbb{N}} \ : \ \forall j \in J \ x_j \in A_j\}$.

Proposition 1.6. *A mapping $X : \Omega \to S^{\mathbb{N}}$ is measurable if and only each coordinate $X_n : \Omega \to S$ is measurable.*

In probabilistic terms: $X = (X_n)_{n\in\mathbb{N}}$ is a random variable in $S^{\mathbb{N}}$ if and only if each coordinate X_n is a random variable in S.

Proof. For each $n \in \mathbb{N}$ and $a \in S$, let

$$C_{n,a} = \{x \in S^{\mathbb{N}} \ : \ x_n = a\}.$$

One can check that the $C_{n,a}$, $n \in \mathbb{N}$, $a \in S$ generate the product sigma algebra. Hence for every mapping $X : \Omega \to S^{\mathbb{N}}$

$$\begin{aligned} X \text{ measurable} &\Longleftrightarrow \forall n \in \mathbb{N} \ \forall a \in S \quad \{X \in C_{n,a}\} \in \mathcal{F} \\ &\Longleftrightarrow \forall n \in \mathbb{N} \ \forall a \in S \quad \{X_n = a\} \in \mathcal{F} \\ &\Longleftrightarrow \forall n \in \mathbb{N} \quad X_n \text{ measurable.} \end{aligned}$$

□

In particular a Markov Chain is a random variable with values in the product space $S^{\mathbb{N}}$ (since each coordinate is a random variable). Hence, to each Markov Chain X we can associate its law P_X, which is a probability measure on the space $S^{\mathbb{N}}$ of trajectories. The following proposition shows that the initial distribution μ together with the transition probability P characterize the law of the Markov Chain.

Proposition 1.7. *Let μ, μ' be two distributions, P, P' two transition probabilities. Let X, X' be two Markov Chains such that $X \sim \mathrm{MC}(\mu, P)$ and $X' \sim \mathrm{MC}(\mu', P')$.*

$$(\mu, P) = (\mu', P') \implies \mathsf{P}_X = \mathsf{P}_{X'}$$

Proof. For every $n \in \mathbb{N}$, and $a = (a_0, \ldots, a_n) \in S^{n+1}$, let

$$C_{n,a} = \{x \in S^{\mathbb{N}} \, : \, x_0 = a_0, \ldots, x_n = a_n\}.$$

The set $C_{n,a}$ is a particular cylinder of the product sigma-algebra,where the first n coordinates of the trajectory are fixed. One can check that the collection $\{C_{n,a}, \, n \in \mathbb{N}, \, a \in S^{n+1}\}$ is a π-system generating the product sigma algebra. Furthermore, for every $n \in \mathbb{N}$ and $a \in S^{n+1}$, we have

$$\begin{aligned} \mathsf{P}_X(C_{n,a}) &= \mathbb{P}(X \in C_{n,a}) \\ &= \mathbb{P}(X_0 = a_0, \ldots, X_n = a_n) \\ &= \mathbb{P}(X'_0 = a_0, \ldots, X'_n = a_n) = \mathsf{P}_{X'}(C_{n,a}). \end{aligned}$$

By Dynkin Lemma, since the measures P_X and $\mathsf{P}_{X'}$ coincide on a π-system generating the sigma-algebra, they must be equal.

□

Remark 1.8. The reverse implication does not hold in general. It holds under the condition that for every $x \in S$, there exists $n \in \mathbb{N}$ and $a_0, \ldots a_n$ such that $\mu_{a_0} p_{a_0 a_1} \cdots p_{a_{n-1} a_n} > 0$.

Chapter 2

Markov Property

Goals:

- Understanding of the setup with a collection of probability measures.
- Intuition in the Markov property.
- Formal statement and applications.

2.1 Setup

Theorem 2.1. *Let S be finite or countable set, equipped with the sigma algebra $\mathcal{P}(S)$. Let $P = (p_{xy})_{x,y \in S}$ be a transition probability on S. There exist:*

- *a measurable space $(\Omega, \mathcal{F})$,*
- *a collection of probability measures $(\mathbf{P}_\mu)_{\mu \text{ dist.}}$ on $(\Omega, \mathcal{F})$, and*
- *a sequence of random variables $X = (X_n)_{n \geq 0}$ on $(\Omega, \mathcal{F})$, such that*

$$X \sim \mathrm{MC}(\mu, P) \quad \textit{under } \mathbf{P}_\mu.$$

for every distribution μ on S.

Proof. We first fix a distribution ν on S with $\nu_x > 0$ for every $x \in S$. If S is finite, we can choose the uniform distribution. If $S = \{s_1, s_2, \ldots\}$ is infinite, we can choose $\nu_{s_i} = 2^{-i}$, $i \geq 1$. Fix some abstract probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a Markov Chain $X \sim \mathrm{MC}(\nu, \mathrm{P})$ on this space.

For every z in S, set $\mathbf{P}_z = \mathbb{P}\left(\cdot \mid X_0 = x\right)$. This way, for every $z \in S$, $n \geq 0$, and $x_0, \dots, x_n \in S$, we have

$$\begin{aligned}
\mathbf{P}_z(X_0 = x_0, \dots, X_n = x_n) &= \frac{\mathbb{P}(X_0 = z, X_0 = x_0, \dots, X_n = x_n)}{\mathbb{P}(X_0 = z)} \\
&= \frac{\mathbf{1}_{z=x_0} \nu_z p_{x_0 x_1} \cdots p_{x_{n-1} x_n}}{\nu_z} \\
&= \mathbf{1}_{z=x_0} p_{x_0 x_1} \cdots p_{x_{n-1} x_n}.
\end{aligned}$$

For every distribution μ, define $\mathbf{P}_\mu = \sum_{z \in S} \mu_z P_z$. This defines a probability measure (since the μ_z sum to 1) and for every $n \geq 0$ and $x_0, \dots, x_n \in S$, we have

$$\begin{aligned}
\mathbf{P}_\mu(X_0 = x_0, \dots, X_n = x_n) &= \sum_{z \in S} \mu_z \mathbf{P}_z(X_0 = x_0, \dots, X_n = x_n) \\
&= \sum_{z \in S} \mu_z \mathbf{1}_{z=x_0} p_{x_0 x_1} \cdots p_{x_{n-1} x_n} \\
&= \mu_{x_0} p_{x_0 x_1} \cdots p_{x_{n-1} x_n}
\end{aligned}$$

□

Setup for the chapter:

- S finite or countable set, equipped with the sigma algebra $\mathcal{P}(S)$.
- P transition probability.
- $(\Omega, \mathcal{F}, (\mathbf{P}_\mu)_{\mu \text{ distribution on } S})$ probability spaces.
- $X = (X_n)_{n \geq 0}$ random variables such that for every distribution μ

$$X \sim \mathrm{MC}(\mu, P) \quad \text{under } \mathbf{P}_\mu.$$

2.2 Simple Markov Property

Notation. For every $n \in \mathbb{N}$, write $\mathcal{F}_n = \sigma(X_0, \dots, X_n)$.

As we have seen in Section 1.5, a Markov Chain satisfies two key properties: absence of memory and homogeneity. The simple Markov Property can be seen as the combination of these two properties. In words, it states that for every fixed time $k \in \mathbb{N}$ and state $x \in S$, the following holds:

"Condition on $X_n = x$, $(X_{k+n})_{n\geq 0}$ is a $\mathrm{MC}(\delta^x, P)$, independent of $\mathcal{F}_k$."

This is formalized in the theorem below.

Theorem 2.2 (Simple Markov Property (MP)). *Let μ be a distribution, $x \in S$, and $k \in \mathbb{N}$. For every $f : S^{\mathbb{N}} \to \mathbb{R}$ measurable and bounded, for every Z $\mathcal{F}_k$-measurable, bounded random variable, we have*

$$\boxed{\mathbf{E}_\mu\big(f((X_{k+n})_{n\geq 0}) \cdot Z \mid X_k = x\big) = \mathbf{E}_x\big(f((X_n)_{n\geq 0})\big)\mathbf{E}_\mu\big(Z \mid X_k = x\big).} \tag{2.1}$$

Lemma 2.3. *Let μ be a distribution on S. Let $x \in S, k \in \mathbb{N}$. For every $N \geq 0$, $a_0, \dots a_k \in S$, $b_0, \dots, b_N \in S$, we have*

$$\begin{aligned}\mathbf{P}_\mu(X_k = b_0, \dots, X_{k+N} = b_N, X_0 = a_0, \dots, X_k = a_k \mid X_k = x)& \\ = \mathbf{P}_x(X_0 = b_0, \dots, X_N = b_N)\mathbf{P}_\mu(X_0 = a_0, \dots, X_k = a_k \mid X_k = x)&\end{aligned}$$

Proof. Without loss of generality, we may assume $x = b_0 = a_k$ (otherwise both sides vanish, and the equality is trivially true). By definition, and using $\delta^x_{b_0} = 1$, we have

$$\begin{aligned}\mathbf{P}_\mu(X_k &= b_0, \dots, X_{k+N} = b_N, X_0 = a_0, \dots, X_k = a_k) \\ &= \mu_{a_0} p_{a_0 a_1} \cdots p_{a_{k-1} a_k} \delta^x_{b_0} p_{b_0 b_1} \cdots p_{b_{N-1} b_N} \\ &= \mathbf{P}_\mu(X_0 = a_0, \dots, X_k = a_k)\mathbf{P}_x(X_0 = b_0, \dots, X_k = b_k)\end{aligned}$$

The statement follows by dividing both sides by $\mathbf{P}_\mu(X_k = a_k) = \mathbf{P}_\mu(X_k = x)$.

□

The lemma above establishes Theorem 2.2 when f is of the form $f(\xi) = \mathbf{1}_{\xi_0 = y_0, \dots \xi_N = y_N}$ and $Z = \mathbf{1}_{X_0 = x_0, \dots, X_k = x_k}$. The extension to general functions follows from standard measure-theoretic approximation arguments, detailed below.

Proof of Theorem 2.2. Let Z be an $\mathcal{F}_k$-measurable, bounded random variable. By linearity, Lemma 2.3 implies that

$$\mathbf{E}_\mu\left(\mathbf{1}_A((X_{k+n})_{n\geq 0}) \cdot Z \mid X_k = x\right) = \mathbf{E}_x\left(\mathbf{1}_A((X_n)_{n\geq 0})\right)\mathbf{E}_\mu\left(Z \mid X_k = x\right). \tag{2.2}$$

for every $A \subset S^{\mathbb{N}}$ of the form $A = \{\xi \in S^{\mathbb{N}} : \xi_0 = y_0, \dots \xi_N = y_N\}$, for $N \geq 0$ and $y_0, \dots, y_N \in S$. The collection of such sets form a π-system generating the product σ-algebra on $S^{\mathbb{N}}$. Furthermore, the collection of sets A satisfying (2.2) is a λ-system. Hence, by Dynkin's Lemma, Equation (2.2) is satisfied for all $A \subset S^{\mathbb{N}}$ measurable.

Now, let $f : S^{\mathbb{N}} \to \mathbb{R}$ measurable and bounded. Equation (2.1) is proved by first approximating f by step functions f_k, and then using linearity. □

Corollary 2.4. *Let μ be a distribution on S, $x \in S$, $k \in \mathbb{N}$. For all $f : S^{\mathbb{N}} \to \mathbb{R}$ measurable and bounded, we have*

$$\boxed{\mathbf{E}_\mu \left(f((X_{k+n})_{n\geq 0}) \mid X_k = x \right) = \mathbf{E}_x \left(f((X_n)_{n\geq 0}) \right).}$$

Proposition 2.5 (Chapman Kolmogorov (CK))**.**

$$\forall m, n \geq 0 \quad \forall x, y \in S \quad \boxed{p_{xy}^{(m+n)} = \sum_{z\in S} p_{xz}^{(m)} p_{zy}^{(n)}.}$$

Proof. Fix m, n and $x, y \in S$.

$$\begin{aligned} p_{xy}^{(m+n)} &= \mathbf{P}_x \left(X_{m+n} = y \right) = \sum_{z\in S} \mathbf{P}_x \left(X_{m+n} = y \mid X_m = z \right) \mathbf{P}_x \left(X_m = z \right) \\ &\overset{\text{(MP)}}{=} \sum_{z\in S} \mathbf{P}_z \left(X_n = y \right) \mathbf{P}_x \left(X_m = z \right) = \sum_{z\in S} p_{xz}^{(m)} p_{zy}^{(n)}. \end{aligned}$$

□

2.3 Strong Markov Property

Definition 2.1. Let $T : \Omega \to \mathbb{N} \cup \{+\infty\}$ be a random variable with values in $\mathbb{N} \cup \{+\infty\}$. We say that T is an <u>$(\mathcal{F}_n)$-stopping time</u> if

$$\forall n \in \mathbb{N} \quad \{T = n\} \in \mathcal{F}_n.$$

Example 2.1 (Hitting Times). Let $A \subset S$, $x \in S$, the hitting times

$$H_A = \min\{n \geq 1 : X_n \in A\} \quad \text{and} \quad H_x = \min\{n \geq 1 : X_n = x\}$$

are stopping times.

Definition 2.2. Let T be a stopping time. The <u>stopped sigma-algebra</u> is defined by

$$\mathcal{F}_T = \{A \in \mathcal{F} : \forall n \in \mathbb{N} : \{T = n\} \cap A \in \mathcal{F}_n\}.$$

In words, the strong Markov property says the following:

"Conditioned on $\{T < \infty, X_T = x\}$, $(X_{T+n})_{n\geq 0}$ is a MC(δ^x, P) independent of $\mathcal{F}_T$"

This is formalized in the following theorem, called the strong Markov property.

Theorem 2.6 (Strong Markov Property (SMP))**.** *Let μ be a distribution on S, T an $(\mathcal{F}_n)$-stopping time. Let $x \in S$, then for all $f : S^{\mathbb{N}} \to \mathbb{R}$ measurable and bounded, and Z $\mathcal{F}_T$-measurable and bounded, we have:*

$$\boxed{\mathbf{E}_\mu\left(f((X_{T+n})_{n\geq 0}) \cdot Z \mid T < \infty, X_T = x\right) = \mathbf{E}_x\left(f((X_n)_{n\geq 0})\right) \mathbf{E}_\mu\left(Z \mid T < \infty, X_T = x\right).}$$

Proof. We will multiply each side of the equation by $\mathbf{P}_\mu\left(T < \infty, X_T = x\right)$.

$$\begin{aligned}
\mathbf{E}_\mu\left(f((X_{T+n})_{n\geq 0}) Z \mathbf{1}_{T<\infty, X_T=x}\right) &= \sum_{k\geq 0} \mathbf{E}_\mu\left(f((X_{k+n})_{n\geq 0}) Z \mathbf{1}_{T=k, X_T=k}\right) \\
&= \sum_{k\geq 0} \mathbf{E}_\mu\left(f((X_{k+n})_{n\geq 0}) Z \mathbf{1}_{T=k} \mid X_k = x\right) \mathbf{P}_\mu\left(X_k = x\right) \\
&\overset{\text{(MP)}}{=} \sum_{k\geq 0} \mathbf{E}_x\left(f((X_n)_{n\geq 0})\right) \mathbf{E}_\mu\left(Z \mathbf{1}_{T=k, X_k=x}\right) \\
&= \mathbf{E}_x\left(f((X_n)_{n\geq 0})\right) \sum_{k\geq 0} \mathbf{E}_\mu\left(Z \mathbf{1}_{T=k, X_k=x}\right) = \mathbf{E}_x\left(f((X_n)_{n\geq 0})\right) \mathbf{E}_\mu\left(Z \mathbf{1}_{T<\infty, X_T=x}\right).
\end{aligned}$$

□

2.4 Inter-visit times

Definition 2.3. Fix $x \in S$. The sequence $(T_i)_{i\geq 1}$ of inter-visit times at x is defined by induction by setting $T_1 = H_x$ and for all $i \geq 1$

$$T_{i+1} = \begin{cases} \min\{n \geq 1 : X_{T_1+\ldots+T_i+n} = x\} & \text{if } T_i < \infty, \\ +\infty & \text{otherwise.} \end{cases}$$

For every $i \geq 1$, the time $S = T_1 + \cdots + T_i$ represents the time of the i-th visit of x. This is a stopping time since for every $n \in \mathbb{N}$

$$\{S \leq n\} = \left\{ \sum_{k=1}^{n} \mathbf{1}_{X_k=x} \geq i \right\} \in \mathcal{F}_n.$$

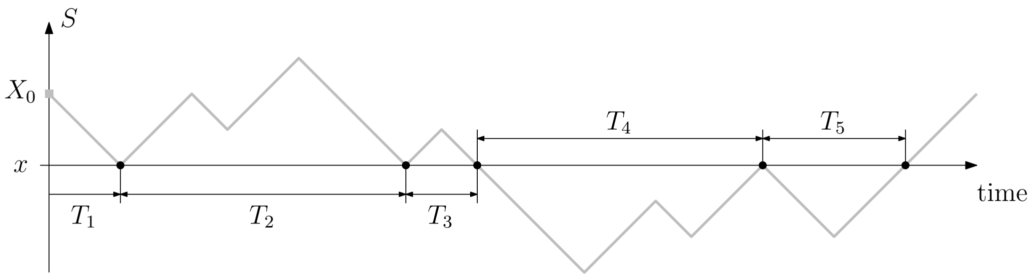


Figure 2.1: Illustration of the inter-visit times a x.

Proposition 2.7. *Let μ be a distribution, $x \in S$. For every $i \geq 1$, we have*

$$\mathbf{P}_\mu(T_i < \infty) = \mathbf{P}_\mu(H_x < \infty)\big(\mathbf{P}_x(H_x < \infty)\big)^{i-1}. \tag{2.3}$$

Proof. We prove the result by induction on i. Equation (2.3) holds for $i = 1$ because $T_1 = H_x$. Now let $i \geq 1$ and assume that (2.3) holds. In order to have $T_{i+1} < \infty$, we must have $S < \infty$, where $S = T_1 + \cdots + T_i$ is the time of the i-th visit of x. Therefore

$$\mathbf{P}_\mu(T_{i+1} < \infty) = \mathbf{P}_\mu\left(T_{i+1} < \infty, S < \infty\right) = \mathbf{P}_\mu\left(T_{i+1} < \infty | S < \infty\right) \cdot \mathbf{P}_\mu(S < \infty). \tag{2.4}$$

Since $S < \infty$ if and only if $T_i < \infty$, we can apply the induction hypothesis to compute the second term in the product above:

$$\mathbf{P}_\mu(S < \infty) = \mathbf{P}_\mu(T_i < \infty) = \mathbf{P}_\mu(H_x < \infty)\big(\mathbf{P}_x(H_x < \infty)\big)^{i-1}.$$

To compute the conditional probability in Equation (2.4), we apply the strong Markov property with the stopping time S. Since it is our first use, let us detail how it is applied. Using that $X_S = x$ if $S < \infty$, and expressing the event $\{T_{i+1} < \infty\}$ in terms of the process $(X_{S+n})_{n\geq 0}$, we get

$$\begin{aligned}
\mathbf{P}_\mu\left(T_{i+1} < \infty | S < \infty\right) &= \mathbf{P}_\mu\left(T_{i+1} < \infty | S < \infty,\ X_S = x\right). \\
&= \mathbf{P}_\mu\left(\exists n \geq 0 \text{ s.t. } X_{S+n} = x | S < \infty,\ X_S = x\right) \\
&\overset{\text{SMP}}{=} \mathbf{P}_x(\exists n \geq 0 \text{ s.t. } X_n = x) \\
&= \mathbf{P}_x(T_1 < \infty).
\end{aligned}$$

This concludes the proof.

□

2.5 Renewal property of the visit times

Proposition 2.8. *Let $x, y \in S$ be such that x is recurrent and $\mathbf{P}_y(H_x < \infty) = 1$. Under $\mathbf{P}_\mu$, the inter-arrival times (after the first visit of x) $T_2, T_3, \ldots$ at x are iid with law given by*

$$\forall t \in \mathbb{N} \quad \mathbf{P}_\mu\left(T_i = t\right) = \mathbf{P}_x\left(H_x = t\right).$$

for every $i \geq 2$.

Remark 2.9. We emphasize that the lemma concerns the inter-visit times T_i starting at $i = 2$. Indeed, the time T_1 corresponds to the time needed to reach x from y, while $T_2, T_3, \cdots$ represent the successive times to reach x from x. Therefore, in general, the distribution of T_1 is not the same as the following times if $y \neq x$. However, if $y = x$, we have $T_1, T_2, \cdots$ iid under $\mathbf{P}_x$.

Proof. We prove by induction on i that for every $i \geq 1$ we have $\mathbf{P}_\mu(T_1, \ldots, T_i < \infty) = 1$, and

$$\forall f_2, \ldots, f_i : \mathbb{N} \to \mathbb{R} \text{ bounded} \quad \mathbf{E}_\mu(f_2(T_2) \cdots f_i(T_i)) = \mathbf{E}_x(f_2(H_x)) \cdots E_x(f_i(H_x)).$$

The statement holds trivially for $i = 1$ (the equation above is an empty statement in this case). Let $i \geq 1$ and assume that the statement holds for i. One can check that the random time $T = T_1 + \cdots + T_i$ is a stopping time. Furthermore we have $\mathbf{P}_\mu(T < \infty, X_T = x) = 1$ (by the induction hypothesis). By the strong Markov property, for every $f_2, \ldots, f_{i+1} : \mathbb{N} \to \mathbb{R}$ bounded, we have

$$\begin{aligned}
\mathbf{E}_\mu\left(f_2(T_2) \cdots f_{i+1}(T_{i+1})\right) &= \mathbf{E}_\mu\left(f_2(T_2) \cdots f_{i+1}(T_{i+1}) | T < \infty, X_T = x\right) \\
&\overset{\text{(SMP)}}{=} \mathbf{E}_\mu\left(f_2(T_2) \cdots f_i(T_i)\right) \mathbf{E}_x\left(f_{i+1}(\min\{n \geq 1 : X_n = x\})\right) \\
&= E_x(f_2(H_x)) \cdots E_x(f_{i+1}(H_x)),
\end{aligned}$$

where we use the induction hypothesis in the last line. □

Chapter 3

Classification of states

Setup:

- S finite or countable set, equipped with the sigma algebra $\mathcal{P}(S)$.
- P transition probability.
- $(\Omega, \mathcal{F}, (\mathbf{P}_\mu)_{\mu \text{ distribution on } S})$ probability spaces.
- $X = (X_n)_{n \geq 0}$ random variables such that for every distribution μ

$$X \sim \mathrm{MC}(\mu, P) \quad \text{under } \mathbf{P}_\mu.$$

Goals:

- Notion of recurrence/transience: link with visit times.
- Decomposition of the state spaces into classes gathering sites with similar properties.

3.1 Recurrence/Transience

Definition 3.1. Let $x \in S$, we say that x is:

- recurrent if $\mathbf{P}_x(H_x < \infty) = 1$.
- transient if $\mathbf{P}_x(H_x < \infty) < 1$.

Notation: For $x \in S$ let

$$\boxed{V_x = \sum_{n \geq 1} \mathbf{1}_{X_n = x}}$$

denote the total number of visits of x by the chain after the first step.

Theorem 3.1. *(Dichotomy Theorem) Let $x \in S$.*

- *If x is recurrent, then $V_x = +\infty$ $\mathbf{P}_x$-a.s..*
- *If x is transient, then $\mathbf{E}_x(V_x) < \infty$.*

Remark 3.2. The theorem excludes the case $\mathbf{P}_x(V_x < \infty) > 0$ and $\mathbf{E}_x(V_x) = +\infty$.

Proof. Let $x \in S$, and let $T_1, T_2 \dots$ be the inter-visit times at x. For every i, by definition we have $T_i < \infty$ if and only if $V_x \geq i$. Hence, by Proposition 2.7 we have

$$\forall i \geq 0 \quad \mathbf{P}_x(V_x \geq i) = \mathbf{P}_x(T_i < \infty) = \rho_x^i, \tag{3.1}$$

where $\rho_x := \mathbf{P}_x(H_x < \infty)$.

If x is recurrent (i.e. $\rho_x = 1$), by continuity of the measure, we get

$$\mathbf{P}_x(V_x = \infty) = \lim_{i \to \infty} \mathbf{P}_x(V_x \geq i) = 1.$$

Now, let us assume that x is transient, i.e. $\rho_x < 1$. Equation (3.1) implies that $1 + V_x$ is a geometric random variable with parameter $1 - \rho_x > 0$, and its expectation is

$$\mathbf{E}_x(V_x) = \frac{\rho_x}{1 - \rho_x} < \infty.$$

□

Corollary 3.3. *Let $x \in S$ be transient. Then for every distribution μ on S, we have $\mathbf{E}_\mu(V_x) < \infty$.*

Proof. We assume that $\mathbf{P}_\mu(H_x < \infty) > 0$ (the result is trivial if $\mathbf{P}_\mu(H_x < \infty) = 0$ because this implies $\mathbf{E}_\mu(V_x) = 0$). Let $\widetilde{V}_x = \sum_{n \geq 0} \mathbf{1}_{X_n = x}$ be the total number of visits of x (including the first step). By the strong Markov property, we have

$$\begin{aligned}
\mathbf{E}_\mu(V_x) = \mathbf{E}_\mu(V_x \mathbf{1}_{H_x < \infty}) &= \mathbf{E}_\mu(V_x \mid H_x < \infty)\mathbf{P}_\mu(H_x < \infty) \\
&\overset{\text{SMP}}{=} \mathbf{E}_x(\widetilde{V}_x)\mathbf{P}_\mu(H_x < \infty) \\
&= (1 + \mathbf{E}_x(V_x))\mathbf{P}_\mu(H_x < \infty) < \infty.
\end{aligned}$$

□

3.2 Positive/Null Recurrence

Notation: For $x \in S$, write $m_x = \mathbf{E}_x(H_x)$.

Definition 3.2. Let $x \in S$ be a recurrent state. We say that x is:

- positive recurrent if $m_x < \infty$,
- null recurrent if $m_x = +\infty$.

The terminology positive/null recurrent is explained in the following section: we will see that the positive recurrent states are the ones which are visited a "positive density" of times, while null recurrent states are visited with a "null density" of times. See the discussion below Theorem 3.4 for more details.

3.3 Density of visits

Notation: For $x \in S$ and $n \geq 0$, let

$$\boxed{V_x^{(n)} = \sum_{k=1}^{n} \mathbf{1}_{X_k = x}}$$

denote the number of visits to x up to time n. Given some distribution μ, The ratio $\frac{1}{n}\mathbf{E}_\mu(V_x^{(n)})$ can be interpreted as the average density of time that the chain spends at x before time n.

Theorem 3.4. *(Density of visits) Let μ be a distribution on S and let $x \in S$ be such that $\mathbf{P}_\mu(H_x < \infty) = 1$. Then*

$$\boxed{\lim_{n \to \infty} \frac{\mathbf{E}_\mu(V_x^{(n)})}{n} = \frac{1}{m_x}.}$$

This theorem can be interpreted as follows:

"In expectation, the density of time spent by the chain at x is $\frac{1}{m_x}$."

If x is transient, or null recurrent ($m_x = \infty$), this density is null. If y is positive recurrent, this density is positive.

Remark 3.5. Notice that for $x, y \in S$ and $n \geq 1$

$$\mathbf{E}_y(V_x^{(n)}) = \sum_{k=1}^{n} \mathbf{E}_y(\mathbf{1}_{X_k=x}) = \sum_{k=1}^{n} p_{yx}^{(k)}.$$

Therefore the theorem above applied to $\mu = \delta^y$ implies

$$\boxed{\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} p_{yx}^{(k)} = \frac{1}{m_x}.}$$

Proof. **Case 1:** x transient. By Corollary 3.3, we have $\mathbf{E}_\mu(V_x) < \infty$. Therefore

$$\frac{\mathbf{E}_\mu(V_x^{(n)})}{n} \leq \frac{\mathbf{E}_\mu\left(V_x\right)}{n} \to 0.$$

Case 2: x recurrent. By Proposition 2.8, we know that the inter-visit times $T_2, T_3 \ldots$ at x are i.i.d. under $\mathbf{P}_\mu$ and fulfill $\mathbf{E}_\mu(T_i) = \mathbf{E}_x\left(H_x\right) = m_x$. Then we can use the Law of Large Numbers and $\mathbf{P}_\mu\left(T_1 < \infty\right) = 1$. We find $\mathbf{P}_\mu$-a.s.,

$$\lim_{i\to\infty} \frac{T_1 + \ldots + T_i}{i} = m_x.$$

Note that this includes the case of $m_x = \infty$, by the following truncation argument: if $m_x = \infty$, consider $K > 0$. By the law of large numbers, $\mathbf{P}_\mu$-almost surely,

$$\liminf_{n\to\infty} \frac{T_2 + \ldots + T_n}{n} \geq \lim_{n\to\infty} \frac{(T_2 \wedge K) + \ldots + (T_n \wedge K)}{n} = \mathbf{E}_\mu(T_2 \wedge K).$$

By monotone convergence, we can let K tend to infinity, and we obtain

$$\lim_{n\to\infty} \frac{T_2 + \ldots + T_n}{n} = \infty$$

$\mathbf{P}_\mu$-almost surely.

Now we write $N_n = V_x^{(n)}$ (the number of visits to x at time n). Following directly from the definition of N_n we have that for any $n > 0$ that

$$T_1 + \ldots + T_{N_n} \leq n < T_1 + \ldots + T_{N_n+1}.$$

Hence, for every $n > 0$, on the event $\{N_n > 0\}$, we have

$$\frac{N_n}{T_1 + \ldots T_{N_n+1}} < \frac{V_x^{(n)}}{n} \leq \frac{N_n}{T_1 + \ldots + T_{N_n}}.$$

The upper and lower bounds each converge to $\frac{1}{m_x}$ almost surely. Hence, we can conclude that $\mathbf{E}_\mu\left(\frac{V_x^{(n)}}{n}\right) \to \frac{1}{m_x}$ by the Dominated Convergence Theorem (using the domination $\frac{V_y^{(n)}}{n} \leq 1$). $\square$

3.4 Communication Classes

Here we will see P as a weighted oriented graph.

Definition 3.3. Let $x, y \in S$. We say that y can be reached from x if there exists an $n \geq 0$ such that $p_{xy}^{(n)} > 0$ and we write $x \to y$. Furthermore, we say that x and y communicate if $y \to x$ and $x \to y$, and we write $x \leftrightarrow y$.

Remark 3.6. (Probabilistic interpretation)

$$x \to y \iff \exists n \geq 0 \ \mathbf{P}_x(X_n = y) > 0 \iff \mathbf{P}_x(\exists n \geq 0 \ X_n = y) > 0.$$

Proposition 3.7. $\leftrightarrow$ *is an equivalence relation on* S.

Proof. Follows from Chapman-Kolmogorov equations. □

Definition 3.4. The equivalence classes of $\leftrightarrow$ are called communication classes of P .

If P has a single unique communication class, we say that P is irreducible.

A communication class C is said to be closed if for every $x, y \in S$

$$x \in C, x \to y \implies y \in C.$$

Proposition 3.8. *Let* C *be a communication class.*

$$C \text{ is closed} \iff \forall x \in C \quad \mathbf{P}_x(\forall n \geq 0 \ X_n \in C) = 1.$$

"If one starts in C, one never leaves."

Proof.

$$\begin{aligned}
(C \text{ is not closed}) &\iff \exists x \in C \ \exists y \in S \setminus C \ x \to y \\
&\iff \exists x \in C \ \exists y \in S \setminus C \ \mathbf{P}_x(\exists n \geq 0 \ X_n = y) > 0 \\
&\iff \exists x \in C \ \mathbf{P}_x(\exists n \geq 0 \ \exists y \in S \setminus C \ X_n = y) > 0 \\
&\iff \exists x \in C \ \mathbf{P}_x(\exists n \geq 0 \ X_n \in S \setminus C) > 0 \\
&\iff \exists x \in C \ \mathbf{P}_x(\forall n \geq 0 \ X_n \setminus C) < 1.
\end{aligned}$$

□

3.5 Closure property of recurrence

Theorem 3.9. *Let $x, y \in S$ such that $x \to y$. If x is recurrent then y is recurrent and $\mathbf{P}_x(H_y < \infty) = \mathbf{P}_y(H_x < \infty) = 1$. In particular $x \leftrightarrow y$.*

Proof. We want to use that every time the chain visits x, it has a non-zero probability to visit y after that, visiting x infinitely often should ensure that y is also visited infinitely often. Assume $y \neq x$ and x recurrent. Let $z_1, \ldots, z_{k-1}$ be distinct elements of S, not equal to x or y such that $p_{xz_1} \cdots p_{z_{k-1}y} > 0$. Then we have

$$\begin{aligned} 0 &= \mathbf{P}_x(H_x = \infty) \geq \mathbf{P}_x(X_1 = z_1, \ldots, X_{k-1} = z_{k-1}, X_k = y, \forall n > 0\ X_{k+n} \neq x) \\ &\overset{(\mathrm{MP})}{=} \underbrace{\mathbf{P}_x(X_1 = z_1, \ldots, X_k = y)}_{>0} \underbrace{\mathbf{P}_y(\forall n > 0\ X_n \neq x)}_{\mathbf{P}_y(H_x = \infty)}. \end{aligned}$$

Thus $\mathbf{P}_y(H_x < \infty) = 1$. Next, we have to show that y is recurrent. Choose m, n such that $p_{xy}^{(n)}, p_{yx}^{(m)} > 0$, we have

$$\mathbf{E}_y(V_y) = \sum_{k>0} p_{yy}^{(k)} \geq \sum_{k>0} p_{yy}^{(m+k+n)} \overset{(\mathrm{CK})}{\geq} \underbrace{p_{yx}^{(m)}}_{>0} \underbrace{\left(\sum_{k>0} p_{xx}^{(k)} \right)}_{=\infty} \underbrace{p_{xy}^{(n)}}_{>0}.$$

Hence, y is recurrent. To show that $\mathbf{P}_x(H_y < \infty) = 1$, use the same argument as above, but with the roles of x and y swapped ($y \to x$, y recurrent), as before. □

Remark 3.10. Let $x \in S$ recurrent and $x \neq y$ then

$$x \to y \iff \mathbf{P}_x(H_y < \infty) > 0 \iff \mathbf{P}_x(H_y < \infty) = 1.$$

Corollary 3.11. *A recurrent class is always closed.*

Proof. C recurrent, $x \in C$, if $x \to y$ then we must have $y \to x$ (otherwise x wouldn't be recurrent), therefore $y \in C$. □

The theorem above gives us a simple criterion for transience:

Corollary 3.12. *If $x \to y$ but $y \nrightarrow x$, then x is transient.*

3.6 Classification of states

Theorem 3.13. *(Classification of states) Let $C \subset S$ be a communication class. Then exactly one of the following holds:*

(i) For all $x \in C$, x is transient.

(ii) For all $x \in C$, x is null recurrent.

(iii) For all $x \in C$, x is positive recurrent.

Proof. Fix $x, y \in S$ with $x \leftrightarrow y$. We prove that y is of the same type (transient, null recurrent or positive recurrent) as x.

If x is transient then y is also transient by Theorem 3.9.

Let us now assume that x positive recurrent. Fix $k \geq 0$ with $p_{xy}^{(k)} > 0$. By Chapman-Kolmogorov, we have for all $j > 0$

$$p_{xy}^{(k+j)} \geq p_{xx}^{(j)} p_{xy}^{(k)}.$$

Thus

$$\underbrace{\frac{1}{n} \sum_{i=1}^{n} p_{xy}^{(i)}}_{\to \frac{1}{m_y}} \geq \underbrace{\left(\frac{1}{n} \sum_{j=1}^{n-k} p_{xx}^{(j)} \right)}_{\to \frac{1}{m_x}} \underbrace{p_{xy}^{(k)}}_{>0}.$$

Therefore, $\frac{1}{m_y} > 0$ and y is positive recurrent. □

Definition 3.5. A communication class $C \subset S$ is said to be transient (resp. recurrent, null recurrent, positive recurrent) if all its elements $x \in C$ are transient (resp. recurrent, null recurrent, positive recurrent).

A consequence of the theorem above is that we can partition the state space S as

$$\boxed{S = T \cup R_1 \cup R_2 \cup \cdots,}$$

where T is the set of transient states (T is equal to the union of all the transient classes), and $R_1, R_2, \ldots,$ are the recurrent classes.

We can classify the behavior of the chain by differentiating if X_n starts in some R_k and if X_n starts in T. In the former case the chain remains in R_k forever. If X_n starts in T, either it remains in T forever, or at some point it moves into an R_k and remains there forever.

Definition 3.6. When P is irreducible, all the sites $x \in S$ are in the same class, and we simply say that P is transient (resp. recurrent, null recurrent, positive recurrent) in the corresponding cases.

3.7 Finite classes

Proposition 3.14. *Let R be a recurrent class, if R is finite, then R is positive recurrent. In particular, if S is finite, then every recurrent state is positive recurrent.*

Proof. Fix $x \in R$, since R is closed we have for every $n > 0$

$$1 = \mathbf{P}_x(X_n \in R) = \sum_{y \in R} p_{xy}^{(n)}.$$

Hence,

$$1 = \sum_{y \in R} \frac{1}{n} \sum_{k=1}^{n} p_{xy}^{(k)} \to \sum_{y \in R} \frac{1}{m_y}.$$

Thus, there must be a $y \in R$ such that $m_y < \infty$, implying that the entire class is positive recurrent.

□

3.8 Finite state space

Proposition 3.15. *If S is finite, then there exists a recurrent state $x \in S$.*

Proof. Almost surely, we have

$$\sum_{x \in S} V_x = \sum_{x \in S} \sum_{n \geq 0} \mathbf{1}_{X_n = x} = \sum_{n \geq 0} \sum_{x \in S} \mathbf{1}_{X_n = x} = \sum_{n \geq 0} 1 = \infty.$$

Fix some distribution μ. By taking the expectation above, and using Fubini Theorem (for nonnegative random variables), we get

$$\sum_{x \in S} \mathbf{E}_\mu(V_x) = \mathbf{E}_\mu\left(\sum_{x \in S} V_x\right) = \infty.$$

Thus we know there exists $x \in S$ such that $\mathbf{E}_\mu(V_x) = \infty$. Therefore by Corollary 3.3, the state x must be recurrent. □

Chapter 4

Convergence to equilibrium

Framework: S finite or countable set, $P = (p_{xy})_{x,y\in S}$ transition probability, $(\Omega, F, (\mathbf{P}_x)_{x\in S})$ probability spaces, $X = (X_n)_{n\geq 0} \sim \mathrm{MC}(\delta^x, P)$ under $\mathbf{P}_x$, $\mathbf{P}_\mu = \sum \mu_x \mathbf{P}_x$.

Goals:

- Definition stationary/reversible distributions.
- Criteria for existence of stationary distributions.
- Behavior of X_n for n large?

4.1 Stationary Distributions

Notation: Let μ be a distribution on S. We define the distribution μP by setting

$$\forall y \in S \quad (\mu P)_y = \sum_{x\in S} \mu_x p_{xy}.$$

(One can check that that it indeed defines a distribution.)

Write μ_n for the law of X_n under $\mathbf{P}_\mu$. It follows from the simple Markov property that the sequence (μ_n) satisfies the induction

$$\begin{cases} \mu_0 = \mu, \\ \mu_{n+1} = \mu_n P \quad \text{for all } n \geq 0. \end{cases}$$

For n large, we expect μ_n to be close to a fixed point of the map $\lambda \to \lambda P$. Such a distribution π is invariant under the dynamics of the process, and the relationship to the long-time behavior of the Markov Chain will be rigorously analyzed in this chapter.

Definition 4.1. Let π be a distribution on S, we say that π is stationary (for P) if

$$\boxed{\pi = \pi P.}$$

When S is finite and if we see P as a matrix, then a stationary distribution corresponds to a left eigenvector π of P for the eigenvalue 1.

Probabilistic interpretation If π is a stationary distribution, then for all $n \geq 0$

$$\boxed{P_\pi[X_n = x] = \pi_x.}.$$

4.2 Reversibility

Definition 4.2. A distribution π on S is said to be reversible (for P) if for any $x, y \in S$

$$\boxed{\pi_x\, p_{xy} = \pi_y\, p_{yx}.}$$

The equation above is equivalent to

$$\mathbf{P}_\pi[X_0 = x, X_1 = y] = \mathbf{P}_\pi[X_0 = y, X_1 = x].$$

Namely, the starting distribution π is reversible if under $\mathbf{P}_\pi$, the probability of starting at y and going to x is equal to the probability of starting at x and going to y. More generally, one can prove (exercise) by induction that π is reversible if and only if for every $n \geq 1$ and $x_0, \ldots, x_n \in S$

$$\mathbf{P}_\pi\left[X_0 = x_0, \ldots, X_n = x_n\right] = \mathbf{P}_\pi\left[X_0 = x_n, \ldots, X_n = x_0\right].$$

"The probability of a trajectory is equal to its time-reversal."

Proposition 4.1. *Let π be a distribution on S. If π is reversible, then π is stationary.*

Proof. Let π be a reversible distribution. For every $y \in S$, we have

$$(\pi P)_y = \sum_{x \in S} \pi_x p_{xy} \overset{\text{reversibility}}{=} \sum_{x \in S} \pi_y p_{yx} = \pi_y \sum_{x \in S} p_{yx} = \pi_y.$$

□

4.3 Stationary Distributions for Irreducible Chains

Recall that $m_x = \mathbf{E}_x[H_x]$, where H_x is the hitting time of x.

Theorem 4.2. *Assume that P is irreducible.*

- *If P is transient or null recurrent, then there is no stationary distribution.*
- *If P is positive recurrent, then there exists a unique stationary distribution given by*

$$\boxed{\pi_x = \frac{1}{m_x}}.$$

Proof. **Case 1:** P transient. Assume for contradiction that there exists a stationary distribution π. For every $x \in S$ and every $n \geq 0$ we have

$$\pi_x = \mathbf{P}_\pi[X_n = x].$$

Write L_x for the last visit time of x. The dichotomy theorem together with the strong Markov property imply that L_x is finite $\mathbf{P}_\pi$-almost surely. Therefore

$$\mathbf{P}_\pi[X_n = x] \leq \mathbf{P}_\pi[L_x \geq n] \xrightarrow{n\to\infty} 0.$$

Therefore, $\pi_x = 0$ for every $x \in S$, this is a contradiction to $\sum_{x\in S} \pi_x = 1$.

Case 2: P null recurrent. Assume for contradiction that there exists a stationary distribution π. As in the transient case we show $\pi_x = 0$ for every x. For every $x \in S$ and for all $n > 0$, we have

$$\pi_x = \frac{1}{n}\sum_{k=1}^{n} \mathbf{P}_\pi\left[X_k = x\right] = \frac{E_\pi[V_x^{(n)}]}{n} = \sum_{y\in S} \pi_y \frac{\mathbf{E}_y[V_x^{(n)}]}{n}. \tag{4.1}$$

Since $\mathbf{P}_y[H_x < \infty] = 1$ for every $y \in S$, by the density of visit theorem, we have

$$\lim_{n\to\infty} \frac{\mathbf{E}_y[V_x^{(n)}]}{n} = \frac{1}{m_x} = 0.$$

By the Dominated Convergence Theorem (using the domination $\frac{\mathbf{E}_y[V_x^{(n)}]}{n} \leq 1$), we can take the limit $n \to \infty$ in (4.1) to conclude $\pi_x = \frac{1}{m_x} = 0$.

Case 3: P positive recurrent. The same argument as in the null recurrent case shows that there is a unique candidate for a stationary distribution, given by

$$\pi_x = \frac{1}{m_x}.$$

To conclude, one needs to prove that this measure is indeed a stationary distribution.

First, let us fix $k \geq 1$. By Theorem 3.4 (density of visits) we have for every $y \in S$

$$\begin{aligned}
\frac{1}{m_y} &= \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} p_{yy}^{(j)} \\
&\overset{\text{(CK)}}{=} \lim_{n\to\infty} \sum_{x\in S} \left(\frac{1}{n} \sum_{j=k}^{n} p_{yx}^{(j-k)} \right) p_{xy}^{(k)} \\
&\overset{\text{(Fatou)}}{\geq} \sum_{x\in S} \liminf_{n\to\infty} \left(\frac{1}{n} \sum_{j=k}^{n} p_{yx}^{(j-k)} \right) p_{xy}^{(k)} \\
&= \sum_{x\in S} \frac{1}{m_x} \cdot p_{xy}^{(k)}.
\end{aligned}$$

Analogously, for a fixed $x \in S$, we have

$$1 = \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \mathbf{P}_x \left[X_j \in S\right] = \lim_{n\to\infty} \sum_{y\in S} \frac{1}{n} \sum_{j=1}^{n} \mathbf{P}_x \left[X_j = y\right] \overset{\text{(Fatou)}}{\geq} \sum_{y\in S} \frac{1}{m_y}.$$

We now prove that the two inequalities above are actually equalities. First, we sum the first inequality over y and get

$$\sum_{y\in S} \frac{1}{m_y} \geq \sum_{y\in S} \left(\sum_{x\in S} \frac{1}{m_x} p_{xy}^{(k)} \right) = \sum_{x\in S} \frac{1}{m_x}.$$

Thus the inequality must be an equality. Namely, for every $k > 0$ and for all $y \in S$, we have

$$\frac{1}{m_y} = \sum_{x\in S} \frac{1}{m_x} p_{xy}^{(k)}. \tag{4.2}$$

We can use this to show that the second inequality is actually an equality. Fix $y \in S$ and note that $\frac{1}{m_y} > 0$ by positive recurrence. We have

$$\begin{aligned}
\frac{1}{m_y} &= \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \left(\sum_{x\in S} \frac{1}{m_x} p_{xy}^{(k)} \right) \\
&= \lim_{n\to\infty} \sum_{x\in S} \frac{1}{m_x} \left(\frac{1}{n} \sum_{k=1}^{n} p_{xy}^{(k)} \right) \\
&\overset{\text{(DCT)}}{=} \sum_{x\in S} \frac{1}{m_x} \frac{1}{m_y}.
\end{aligned}$$

Hence, $\pi_x = \frac{1}{m_x}$ defines a distribution, which is stationary (this follows from Equation (4.2) with $k = 1$). □

4.4 Periodicity

Definition 4.3. Let $x \in S$. The period of x is defined by

$$\boxed{d_x = \gcd\{n > 0 : p_{xx}^{(n)} > 0\}}.$$

By convention $\gcd(\emptyset) = \infty$.

The following proposition asserts that the period is constant on the communication classes.

Proposition 4.3. *Let $x, y \in S$. If $x \leftrightarrow y$, then $d_x = d_y$.*

Proof. Let $x \neq y$. We prove that $d_y | d_x$.

Let us fix $k, \ell \geq 0$ such that $p_{yx}^{(k)}, p_{xy}^{(\ell)} > 0$. Since $p_{yy}^{(k+\ell)} \geq p_{yx}^{(k)} p_{xy}^{(\ell)} > 0$ we have that $d_y | k + \ell$. Now we show that d_y is a common divisor of $\{n > 0 : p_{xx}^{(n)} > 0\}$, this will imply our claim. For every $n > 0$ satisfying $p_{xx}^{(n)} > 0$, we have

$$p_{yy}^{(k+\ell+n)} \geq p_{yx}^{(k)} p_{xx}^{(n)} p_{xy}^{(\ell)} > 0,$$

hence $d_y | k + \ell + n$. Since $d_y | k + \ell$, we also have $d_y | n$. □

Consequence: If P is irreducible, we have

$$\forall x, y \in S \quad d_x = d_y.$$

Definition 4.4. We say that P is aperiodic if for every $x \in S$

$$\boxed{d_x = 1.}$$

Proposition 4.4. *Let x be in S. We have $d_x = 1$ if and only if there is an $n_0 \geq 1$ such that for every $n \geq n_0$ we have that $p_{xx}^{(n)} > 0$.*

We use the following lemma from number theory.

Lemma 4.5. *Let $A \subset \mathbb{N} \setminus \{0\}$ be stable under addition (i.e. $x, y \in A \implies x + y \in A$). Then*

$$\gcd(A) = 1 \iff \exists n_0 \in \mathbb{N} : \{n \in \mathbb{N} : n \geq n_0\} \subset A.$$

Proof.

$\boxed{\Leftarrow}$ Follows from the fact that $\gcd(n_0, n_0+1) = 1$.

$\boxed{\Rightarrow}$ Assume $\gcd(A) = 1$. Let $a \in A$ be arbitrary and $a = \prod_{i=1}^{k} p_i^{\alpha_i}$ be its prime factorization. Since $\gcd(A) = 1$, one can find $b_1, \ldots, b_k \in A$ such that for all i $p_i \nmid b_i$. This implies

$$\gcd(a, b_1, \ldots, b_k) = 1.$$

Write $d = \gcd(b_1, \ldots, b_k)$. By Bezout's Theorem, we can pick $u_1, \ldots, u_k \in \mathbb{Z}$ such that

$$u_1 b_1 + \ldots + u_k b_k = d.$$

Now, choose an integer λ large enough such that $u_i + \lambda a \geq 0$ for every i and define

$$b = (u_1 + \lambda a)b_1 + \ldots + (u_k + \lambda a)b_k = d + \lambda(b_1 + \ldots + b_k)a.$$

The first expression shows that $b \in A$, and the second implies that $\gcd(a, b) = gcd(a, d) = 1$. To summarize, we found $a, b \in A$ such that $\gcd(a, b) = 1$.

Without loss of generality, we may assume $a < b$. Since $\gcd(a, b) = 1$, the set $B = \{b, 2b, \ldots, ab\}$ covers all of the residue classes modulo a. Since $a < b$, this implies that $B + \{ka, k \in \mathbb{N}\}$ includes every number $z \geq ab$. This concludes the proof by choosing $n_0 = ab$. □

Proof of Proposition 4.4. The set $A_x = \{n > 0 : p_{xx}^{(n)} > 0\}$ under addition, because $p_{xx}^{(m+n)} \geq p_{xx}^{(m)} p_{xx}^{(n)}$ for every $m, n > 0$. The proof follows by applying the lemma to $A = A_x$. □

4.5 Product Chain

Our goal in the next two sections is to define two Markov Chains X a $\mathrm{MC}(\mu, P)$ and $\widetilde{X}$ a $\mathrm{MC}(\nu, P)$ on the same probability space such that $X_n = \widetilde{X_n}$ for n large.

To achieve this, we first consider two independent chains X and Y. We then show that the chains meet almost surely (under some assumptions on P) at some random time T. Then we ask that the chains follow the same trajectory for $t > T$.

Notation: Let μ, ν be two distributions on S, we write $\mu \otimes \nu$ for the distribution on S^2, defined by

$$\forall (x, y) \in S^2 \qquad (\mu \otimes \nu)_{(x,y)} = \mu_x \nu_y.$$

Proposition 4.6. *Let $X \sim \mathrm{MC}(\mu, P)$ and $Y \sim \mathrm{MC}(\nu, P)$ be two independent Markov Chains. The sequence of random variables $(X, Y) := ((X_n, Y_n))_{n \geq 0}$ is a Markov Chain on*

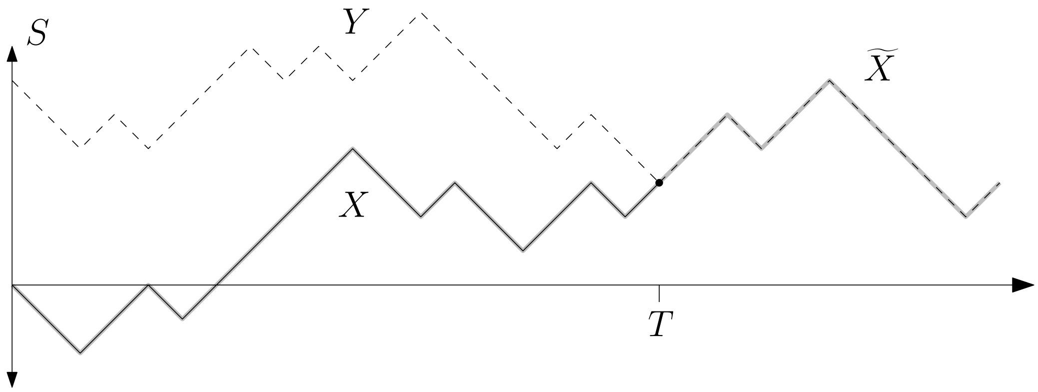


Figure 4.1: A coupling of two simple random walks started from 6 and 0

S^2 with initial distribution $\mu \otimes \nu$ and transition probability $\overline{P}$ defined by

$$\boxed{\overline{p}_{\omega,\omega'} = p_{xx'}p_{yy'}}.$$

Remark 4.7. To see that $\overline{P} = (\overline{p}_{w,w'})_{w,w' \in S^2}$ is a transition probability, calculate

$$\sum_{w' \in S} \overline{p}_{ww'} = \sum_{x',y' \in S} p_{xx'}p_{yy'} = 1.$$

Proposition 4.8. *If P is irreducible and aperiodic then $\overline{P}$ is irreducible and aperiodic.*

Remark 4.9. Aperiodic is important. Indeed P irreducible does not imply that $\overline{P}$ is irreducible in general. For example, consider $S = \{1, 2\}$ and $p_{12} = p_{21} = 1$. In this case, P is irreducible, but $\overline{P}$ is not irreducible.

Proof. Let $w = (x, y)$ and $w' = (x', y') \in S^2$. By irreducibility we can choose $k, \ell \geq 0$ such that $p^{(k)}_{xx'}, p^{(\ell)}_{yy'} > 0$. Then for every $n \geq \max(k, \ell)$ we have

$$\overline{p}^{(n)}_{ww'} = p^{(n)}_{xx'}p^{(n)}_{yy'} \geq p^{(k)}_{xx'}p^{(n-k)}_{x'x'}p^{(\ell)}_{yy'}p^{(n-\ell)}_{y'y'} > 0.$$

This holds as the two terms $p^{(n-k)}_{x'x'}$ and $p^{(n-\ell)}_{y'y'}$ are strictly positive for n large enough. □

Proposition 4.10. *If π is stationary for P then $\pi \otimes \pi$ is stationary for $\overline{P}$.*

Proof. For every $(y, y') \in S^2$ we have

$$\pi_y \pi_{y'} = \sum_{x \in S} \pi_x p_{xy} \sum_{x' \in S} \pi_{x'} p_{x'y'} = \sum_{(x,x') \in S^2} \pi_x \pi_{x'} p_{xy} p_{x'y'}.$$

□

4.6 Coupling Markov Chains

In this whole section, we fix $X \sim \mathrm{MC}(\mu, P)$ and $Y \sim \mathrm{MC}(\nu, P)$ two independent Markov Chains on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Definition 4.5. We define the stopping time (for the product chain (X, Y))

$$\boxed{T = \min\{n \geq 0 : X_n = Y_n\}.}$$

Remark 4.11. To see that T is indeed a stopping time, notice that $T = H_A$ with $A = \{(x, y) \in S^2 : x = y\}$.

Proposition 4.12. *For every $n \geq 0$*

$$\boxed{\sum_{x \in S} |\mathbb{P}[X_n = x] - \mathbb{P}[Y_n = x]| \leq 2\mathbb{P}[T > n].}$$

Lemma 4.13. *The sequence of random variable $\widetilde{X} = (\widetilde{X}_n)_{n \geq 0}$ defined by*

$$\widetilde{X}_n = \begin{cases} Y_n & \text{for } n < T \\ X_n & \text{for } n \geq T \end{cases}.$$

is a Markov Chain on S with initial distribution ν and transition probability P.

Proof. Define $\widetilde{Y}$ by

$$\widetilde{Y}_n = \begin{cases} X_n & \text{for } n < T \\ Y_n & \text{for } n \geq T \end{cases}.$$

Let $n \geq 0$. Writing $X_{[n]}$ for $(X_1, \ldots, X_n)$, we show that $(X_{[n]}, Y_{[n]})$ and $(\widetilde{Y}_{[n]}, \widetilde{X}_{[n]})$ have the same distribution. This implies that $\widetilde{X}_{[n]}$ has the same distribution as $Y_{[n]}$, which concludes the proof. To achieve this, we fix $x = (x_0, \ldots, x_n)$ and $y = (y_0, \ldots, y_n) \in S^n$, and prove that

$$\mathbb{P}[X_{[n]} = x, Y_{[n]} = y] = \mathbb{P}[\widetilde{Y}_{[n]} = x, \widetilde{X}_{[n]} = y]. \tag{4.3}$$

If $x_i \neq y_i$ for every $i \leq n$, then the trajectories x and y do not intersect and (4.3) is a direct consequence of the definition of $(\widetilde{X}, \widetilde{Y})$. Now, we assume that $x_i = y_i$ for some index $i \leq n$ and we prove that (4.3) also holds in this case. Define

$$t = \min\{i \ : \ x_i = y_i\}.$$

In particular we have $x_t = y_t$. If $X_{[n]} = x, Y_{[n]} = y$ then $T = t$. Furthermore, by using $x_t = y_t$ and the independence between X and Y, we find

$$\begin{aligned} \mathbb{P}[X_{[n]} = x, Y_{[n]} = y] &= \mathbb{P}[X_{[n]} = (x_0, \ldots, x_t, y_{t+1}, \ldots, y_n), Y_{[n]} = (y_0, \ldots, y_t, x_{t+1}, \ldots, x_n)] \\ &= \mathbb{P}[\widetilde{Y}_{[n]} = x, \widetilde{X}_{[n]} = y]. \end{aligned}$$

which concludes the proof. □

Proof of Proposition 4.12. We use the coupling between X and $\widetilde{X}$ to conclude the proof. For every $n \geq 0$

$$\begin{aligned} \sum_{x \in S} |\mathbb{P}\left[X_n = x\right] - \ \left[Y_n = x\right]| &= \sum_{x \in S} \left| \mathbb{P}\left[X_n = x\right] - \mathbb{P}\left[\widetilde{X}_n = x\right] \right| \\ &= \sum_{x \in S} \left| \mathbb{P}\left[X_n = x, T \leq n\right] + \mathbb{P}\left[X_n = x, T > n\right] \right. \\ &\qquad \left. - \mathbb{P}\left[\widetilde{X}_n = x, T \leq n\right] - \mathbb{P}\left[\widetilde{X}_n = x, T > n\right] \right| \\ &\leq \sum_{x \in S} \mathbb{P}\left[X_n = x, T > n\right] + \mathbb{P}\left[\widetilde{X}_n = x, T > n\right] \\ &= 2\mathbb{P}\left[T > n\right]. \end{aligned}$$

□

4.7 Convergence to equilibrium

Theorem 4.14. *Assume that P is irreducible, aperiodic, and admits a stationary distribution π. Then for every distribution μ on S and $x \in S$*

$$\boxed{\lim_{n \to \infty} \mathbf{P}_\mu \left[X_n = x\right] = \pi_x.}$$

Equivalently: Under $\mathbf{P}_\mu : X_n \overset{(law)}{\to} X_\infty$ where $X_\infty \sim \pi$.

Equivalently: For all $f : S \to \mathbb{R}$ bounded: $\lim_{n \to \infty} \mathbf{E}_\mu \left[f(X_n)\right] = \int_S f d\pi$.

Proof. Consider the product chain $(X_n, Y_n)_{n\geq 0}$ as before, where X has initial distribution μ and Y starts with the invariant distribution π .

By Proposition 4.8, the product transition probability $\overline{P}$ is irreducible. Furthermore, by Proposition 4.10, it admits a stationary distribution. By Theorem 4.2, this implies that $\overline{P}$ is positive recurrent. Fix an arbitrary vertex $a \in S$ and consider the hitting time $H_{(a,a)}$ for the product chain. Since $\overline{P}$ is irreducible and recurrent, Theorem (3.9) (closure property of recurrence) implies that the hitting time $H_{(a,a)}$ is finite almost surely. Therefore, the stopping time $T = \min\{n \geq 0 : X_n = Y_n\}$ is also finite almost surely, because $T \leq H_{(a,a)}$. By applying Proposition 4.12, we have that for every $x \in S$

$$|\mathbb{P}[X_n = x] - \pi_x| = |\mathbb{P}[X_n = x] - \mathbb{P}[Y_n = x]| \leq 2\mathbb{P}[T > n] \xrightarrow{n\to\infty} 0.$$

□

4.8 Null recurrent and transient cases

Theorem 4.15. *Assume that P is irreducible, aperiodic, and null recurrent or transient. Then for every distribution μ and every $x \in S$*

$$\boxed{\lim_{n\to\infty} \mathbf{P}_\mu[X_n = x] = 0.}$$

Lemma 4.16. *Assume that $\overline{P}$ is irreducible and recurrent. For every μ distribution on S, any $i \geq 0$, and every $x \in S$*

$$\lim_{n\to\infty} |\mathbf{P}_\mu[X_n = x] - \mathbf{P}_\mu[X_{n+i} = x]| = 0$$

Proof. Fix $i \geq 0$ and consider the distribution $\mu_i = \mu P^i$ (i.e. μ_i is the law of X_i under $\mathbf{P}_\mu$). Let $X \sim \mathrm{MC}(\mu)$ and $Y \sim \mathrm{MC}(\mu_i)$ be two independent Markov Chains. For each $n \geq 0$, the distribution of Y_n is $\mu_i P^n = \mu P^{i+n}$ (by Chapman Kolmogorov equations), therefore

$$\forall x \in S \quad \mathbb{P}[Y_n = x] = \mathbb{P}[X_{n+i} = x].$$

The stopping time $T = \min\{n \geq 0 : X_n = Y_n\}$ is finite almost surely as $\overline{P}$ is irreducible and recurrent. By Proposition 4.12, we have $\lim_{n\to\infty} |\mathbb{P}[X_n = x] - \mathbb{P}[Y_n = x] = 0|$, i.e.

$$\lim_{n\to\infty} |\mathbb{P}[X_n = x] - \mathbb{P}[X_{n+i} = x]| = 0.$$

□

Proof of Theorem 4.15. We distinguish two cases, depending whether $\overline{P}$ is transient or recurrent.

Case 1: Assume $\overline{P}$ transient. Let $X, Y \sim \mathrm{MC}(\mu, P)$ independent. Fix $x \in S$, since (x,x) is transient, the last visit $L = \max\{n \geq 0 : (X_n, Y_n) = (x,x)\}$ is finite almost surely (by the Dichotomy Theorem). Hence,

$$\mathbb{P}[X_n = x]^2 = \mathbb{P}\left[X_n = x, Y_n = x\right] \leq \mathbb{P}\left[L \geq n\right] \xrightarrow{n\to\infty} 0.$$

Case 2: Assume $\overline{P}$ is null recurrent. Fix $x \in S$ and $\varepsilon > 0$. Since x is a null recurrent state, by Theorem 3.4 (density of visits), we can choose k such that

$$\frac{1}{k}\sum_{i=0}^{k-1} p_{xx}^{(i)} < \epsilon.$$

For every $n \geq 0$, define the stopping time $H = \min\{j \geq n : X_j = x\}$ (representing the first hit time of x after time n). Since the chain does not visit x between time n and time H, we have

$$\frac{1}{k}\sum_{i=1}^{k} \mathbf{P}_\mu\left[X_{n+i} = x\right] \leq \frac{1}{k}\sum_{i=0}^{k-1} \mathbf{P}_\mu\left[X_{H+i} = x\right] \overset{\text{(StMP)}}{=} \frac{1}{k}\sum_{i=0}^{k-1} \mathbf{P}_x\left[X_i = x\right] \leq \varepsilon.$$

In order to conclude, we use Lemma 4.16: for n large $\mathbf{P}_\mu\left[X_n = x\right]$ is closed to the average $\frac{1}{k}\sum_{i=1}^{k} \mathbf{P}_\mu\left[X_{n+i} = x\right]$, which is small by the equation above. More precisely, for every $n \geq 0$, we have

$$\begin{aligned}\mathbf{P}_\mu\left[X_n = x\right] &= \frac{1}{k}\sum_{i=1}^{k} \mathbf{P}_\mu\left[X_n = x\right] \\ &\leq \underbrace{\frac{1}{k}\sum_{i=1}^{k} \left|\mathbf{P}_\mu\left[X_n = x\right] - \mathbf{P}_\mu\left[X_{n+i} = x\right]\right|}_{\xrightarrow[n\to\infty]{\text{Lemma 4.16}} 0} + \underbrace{\frac{1}{k}\sum_{i=1}^{k} \mathbf{P}_\mu\left[X_{n+i} = x\right]}_{\leq \varepsilon}.\end{aligned}$$

Since $\overline{P}$ is irreducible and recurrent, Lemma 4.16 concludes that

$$\limsup_{n\to\infty} \mathbf{P}_\mu\left[X_n = x\right] \leq \epsilon.$$

□

4.9 Monte-Carlo Markov Chain: Hardcore model

Reference: see Chapter 7 of [2].

We consider a 8x8 square grid, i.e. the graph $G = (V, E)$ where $V = \{1, \ldots, 8\}^2$ and $E = \{\{x, y\} \subset V \ : \ \|x - y\|_1 = 1\}$. In the hardcore model, particles are placed randomly on the vertices in such a way that

- there is at most one particle on each vertex; and
- no two neighbours are occupied by a particle.

Formally, a configuration is an element $\xi \in \{0, 1\}^V$. Such a configuration associates to each vertex $v \in V$ a value $\xi(v) = 0$ or $\xi(v) = 1$, where $\xi(v) = 1$ is interpreted as the presence of a particle at v. Such a configuration is called admissible if $\min(\xi(v), \xi(w)) = 0$ for every edge $\{v, w\} \in E$.

Question: How to simulate Y, a uniform random variable in

$$S = \{\xi \in \{0, 1\}^V \ : \ \xi \text{ is admissible}\}?$$

We will construct a Markov chain on S with stationary distribution π, the uniform distribution on S. We start on a fixed admissible configuration $X_0 = \eta \in S$. For every $n \geq 0$, we define X_{n+1} from X_n as follows:

- Pick a vertex v uniformly at random in V.
- If a neighbour of v is occupied in X_n, we do nothing and set $X_{n+1} = X_n$.
- If none of the neighbours of v is occupied in X_n, then we set $X_{n+1}(v)$ to be the result of a fair coin, and we leave all the other values unchanged: we set $X_{n+1}(w) = X_n(w)$, for all $w \neq v$.

Proposition 4.17. *For every $\xi \in S$ we have*

$$\lim_{n \to \infty} \mathbb{P}[X_n = \xi] = \frac{1}{|S|}.$$

Proof. The chain defined above is a Markov Chain with transition probability P defined by

$$p_{\xi,\psi} = \begin{cases} \frac{1}{2|V|} & \text{if } \psi \text{ and } \xi \text{ differ exactly at one vertex,} \\ 1 - \frac{k}{2|V|} & \text{if } \xi = \psi, \\ 0 & \text{otherwise.} \end{cases}$$

where $k = k(\xi)$ is the number of admissible configurations ψ that differ from ξ exactly at one vertex. The definition of $p_{\xi,\psi}$ is symmetric in $\xi, \psi \in S$, therefore $p_{\xi,\psi} = p_{\psi,\xi}$, which implies that

$$\forall \xi, \psi \in S \qquad \frac{1}{|S|} p_{\xi,\psi} = \frac{1}{|S|} p_{\psi,\xi}.$$

This implies that the uniform distribution is reversible, and therefore stationary.

Furthermore, the chain is irreducible (one can check that $0 \leftrightarrow \xi$ for all $\xi \in S$) and aperiodic (because $p_{\xi,\xi} > 0$ for every ξ). See Exercise 6.5 for more details. The proof follows by applying Theorem 4.14.

□

Chapter 5

Renewal Processes

Framework: $(\Omega, \mathcal{F}, \mathbb{P})$ probability space. In the whole chapter, we fix

$$T_1, T_2, \ldots \text{ i.i.d. random variables on } \mathbb{R}_+$$

satisfying $\mathbb{P}[T_1 = 0] < 1$. We write

$$\mu = \mathbb{E}[T_1] \in (0, \infty] \qquad \text{and} \qquad F(t) = \mathbb{P}[T_1 \leq t]$$

for the expectation and the distribution function of T_1, respectively.

5.1 Definition

Definition 5.1. Let $i \geq 1$. The random variable T_i is called the i-th inter-arrival time, and we define the i-th arrival time (or i-th renewal time) as

$$\boxed{S_i = T_1 + \cdots + T_i.}$$

Definition 5.2. The continuous time stochastic process $(N_t)_{t \geq 0}$ defined by

$$\forall t \geq 0 \quad N_t = \sum_{k=1}^{\infty} \mathbf{1}_{S_k \leq t}$$

is called the *renewal process with arrival distribution* F.

In words, N_t counts the number of renewal times in the interval $[0, t]$.

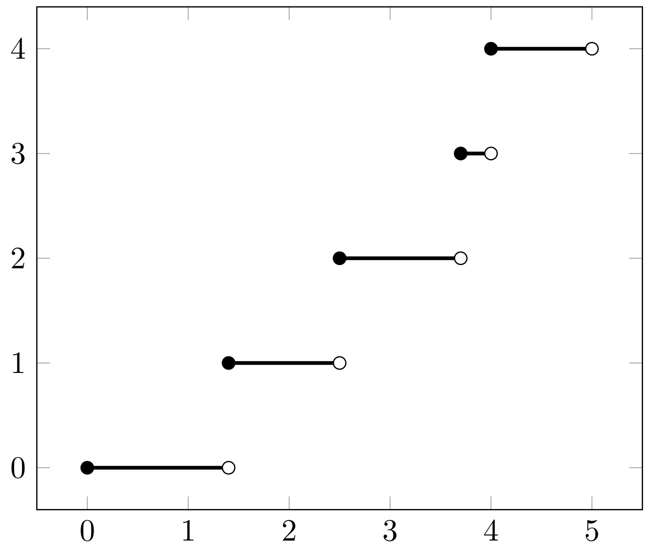
Examples:

(i) $T_1 = 1$ a.s. ("deterministic case")

(ii) $T_1 \sim \mathcal{U}(0,1)$.

5.2 Exponential inter-arrival times

If the inter-arrival times are exponential random variables with parameter λ, then the renewal process N is called a *Poisson Process with parameter* λ. Such process will be analyzed in more depth in Chapter **??**. The name comes from the distribution of N_t, which is a Poisson random variable, as stated in the following proposition.



Proposition 5.1. *Fix* $\lambda > 0$ *and assume that*

$$T_1 \sim \mathrm{Exp}(\lambda)$$

(i.e. $F(t) = 1 - e^{\lambda t}$ *for* $t \geq 0$*). In this case, for every fixed* $t \geq 0$*, we have*

$$N_t \sim \mathrm{Pois}(\lambda t).$$

Proof. We prove by induction on n, that

$$\forall t \geq 0 \quad \mathbb{P}[N_t = n] = \frac{(\lambda t)^n}{n!} e^{-\lambda t} \tag{5.1}$$

For $n = 0$, we have $N_t = 0$ if there is no renewal before time t, therefore,

$$\mathbb{P}[N_t = 0] = \mathbb{P}[T_1 \geq t] = e^{-\lambda t}.$$

Let $n \geq 0$ and assume that (5.1) holds. Fix $t \geq 0$. There are $n+1$ renewal before time t iff $T_1 < t$ and there are exactly n renewal times between T_1 and t. By conditioning on T_1, and using independence, we obtain

$$\begin{aligned}
\mathbb{P}[N_t = n+1] &= \mathbb{P}[T_1 < t, T_1 + \cdots + T_{n+1} \leq t, T_1 + \cdots + T_{n+2} > t] \\
&= \int_0^\infty \mathbb{P}[s < t, s + T_2 + \cdots + T_{n+1} \leq t, s + T_2 + \cdots + T_{n+2} > t] \lambda e^{-\lambda s} ds \\
&= \int_0^t \mathbb{P}[T_2 + \cdots + T_{n+1} \leq t - s, T_2 + \cdots + T_{n+2} > t - s] \lambda e^{-\lambda s} ds \\
&= \int_0^t \mathbb{P}[N_{t-s} = n] \lambda e^{-\lambda s} ds
\end{aligned}$$

By the induction hypothesis, we obtain

$$\mathbb{P}[N_t = n+1] = \int_0^t \frac{(\lambda(t-s))^n}{n!} \lambda e^{-\lambda t} ds = \left[-\frac{(\lambda(t-s))^{n+1}}{(n+1)!} \right]_0^t e^{-\lambda t} ds = \frac{(\lambda t)^{n+1}}{(n+1)!} e^{-\lambda t}.$$

□

5.3 Bernoulli inter-arrival times

In this section, we give another example where the law of N_t can be computed explicitly.

Proposition 5.2. *Fix $\alpha > 0$ and $0 < \beta \leq 1$ and assume that*

$$T_1 = \begin{cases} \alpha & \text{with probability } \beta \\ 0 & \text{with probability } 1 - \beta \end{cases}$$

(i.e. $T_1 \overset{\text{(law)}}{=} \alpha Z$, where $Z \sim \text{Ber}(\beta)$). In this case, for every fixed $t \geq 0$, we have

$$N_t \overset{\text{(law)}}{=} X_0 + \sum_{i=1}^{\lfloor t/\alpha \rfloor} (1 + X_i).$$

where the X_i's are i.i.d. geometric random variables with parameter β.

Proof. The sequence $T_1, T_2, \dots$ is a random sequence of numbers taking values in $\{0, \alpha\}$. Since $T_i = \alpha$ with probablity $\beta > 0$, we know (by Borel-Cantelli Theorem) that the value α appears infinitely many times. Define $X_0 \in \{0, 1, 2 \dots\}$ to be the numbers of 0's before the first α, and for every $i \geq 1$, define X_i as the number of 0's between the i-th and the $i+1$-th α. Notice that $X_0, X_1, \dots$ is an iid sequence of geometric random variables with parameter β. Indeed, by independence, for every $i \geq 0$ and every $k_0, \dots, k_i$ we have

$$\mathbb{P}[X_0 = k_0, \dots, X_i = k_i] = \prod_{j=0}^{i} \mathbb{P}[T_{\ell_j+1} = 0, \dots, T_{\ell_j+k_j-1} = 0, T_{\ell_j+k_j} = \alpha] = \prod_{j=0}^{i} (1-\beta)^{k_j} \beta.$$

where we set $\ell_0 = 0$ and $\ell_j = k_1 + \cdots + k_j$ for $j \geq 1$.

By definition, the number of renewal times before time t is exactly the number of terms in the sequence $(T_1, T_2, \dots)$ before we see $\lfloor t/\alpha \rfloor$ times the value α. Following the definitions above, we get

$$N_t = X_0 + \sum_{i=1}^{\lfloor t/\alpha \rfloor} (1 + X_i).$$

□

5.4 Basic properties

Lemma 5.3 (Monotonicity). *Let $(T'_i)_{i\geq 1}$ be a sequence of iid random variables satisfying*

$$T'_i \leq T_i \quad a.s.$$

Then the renewal process N' define by $N'_t = \sum_{k=1}^{\infty} \mathbf{1}_{T'_1+\cdots+T'_k \leq t}$ satisfies

$$N'_t \geq N_t \quad a.s.$$

for every $t \geq 0$.

Proof. Let $k \geq 1$ and $t \geq 0$. If $T_1 + \cdots + T_k \leq t$ then $T'_1 + \cdots + T'_k \leq t$ a.s. Therefore,

$$\mathbf{1}_{T_1+\cdots+T_k \leq t} \leq \mathbf{1}_{T'_1+\cdots+T'_k \leq t} \qquad \text{a.s.}$$

The results follows by summing the equation above over all $k \geq 1$. □

Proposition 5.4 (Basic properties). *The renewal process N satisfies the following properties. Almost surely,*

(i) $t \mapsto N_t$ *is non-decreasing, right continuous, with values in* $\mathbb{N}$ *and*

(ii) $\lim_{t\to\infty} N_t = \infty$.

Proof.

(i) Write $\mathbb{Q}_+ = \mathbb{Q} \cap (0,\infty)$ for the positive rational numbers. We have

$$\mathbb{P}\left[T_1 > 0\right] = \mathbb{P}\Big[\bigcup_{\alpha\in\mathbb{Q}_+} \{T_1 \geq \alpha\}\Big] = \lim_{\substack{\alpha\to 0\\ \alpha\in\mathbb{Q}_+}} \mathbb{P}[T_1 \geq \alpha].$$

We have

$$\sum_{i>0} \mathbb{P}\left[T_i \geq \alpha\right] = \infty.$$

Therefore, by the Borel-Cantelli lemma, $\mathbb{P}\left[A\right] = 1$, where

$$A = \left\{\omega : T_i(\omega) \geq \alpha \text{ for infinitely many } i\right\}.$$

For every $\omega \in A$, $\lim_{n\to\infty} S_n(\omega) = \infty$, and therefore

$$t \mapsto N_t(\omega) = \sum_{k\geq 1} \mathbf{1}_{S_k(\omega)\leq t}$$

is a non-decreasing function with values in $\mathbb{N}$.

(ii) All the inter-arrival times $T_1, T_2, \ldots$ are finite almost surely. Therefore, all the renewal times $S_1, S_2, \ldots$ are finite almost surely. When this occurs, we have

$$\lim_{t\to\infty} N_t = \lim_{t\to\infty} \sum_{k\geq 1} \mathbf{1}_{S_k\leq t} = +\infty.$$

□

5.5 Exponential moments

Proposition 5.5 (Exponential moments)**.** *There exists $c > 0$ such that*

$$\forall t \geq 0 \quad \mathbb{E}\left[e^{cN_t}\right] \leq e^{\frac{1+t}{c}}$$

Proof. As in the proof of Proposition 5.4, we can pick $\alpha \in (0,1]$ such that $\mathbb{P}\left[T_1 \geq \alpha\right] > 0$. For every $i > 0$, define

$$T_i' = \alpha \mathbf{1}_{T_i \geq \alpha}.$$

We have $T_i' \leq T_i$ a.s. and (T_i') are i.i.d. random variables with

$$T_i' = \begin{cases} \alpha & \text{with probability } \beta \\ 0 & \text{with probability } 1-\beta \end{cases}$$

where $\beta = \mathbb{P}\left[T_1 \geq \alpha\right] > 0$. Define the renewal process N' by

$$N_t' = \sum_{k \geq 1} \mathbf{1}_{T_1' + \cdots + T_k' \leq t}.$$

By Proposition 5.2, we have that

$$N_t' \overset{\text{(law)}}{=} X_0 + \sum_{i=1}^{\lfloor \frac{t}{\alpha} \rfloor} (1 + X_i),$$

where (X_i) are geometric random variables with success parameter β. For $c > 0$ such that $(1-\beta)e^c < 1$ we have

$$\mathbb{E}[e^{c(1+X_i)}] = e^c \left(\frac{\beta}{1-(1-\beta)e^c} \right) \leq e^{\frac{\alpha}{c}}.$$

Hence, we can choose $c > 0$ small enough such that $\mathbb{E}[e^{c(1+X_i)}] \leq e^{\frac{\alpha}{c}}$. Using this bound and independence we obtain for all $t \geq 0$

$$\mathbb{E}\left[e^{cN_t'}\right] \leq \prod_{i=0}^{\lfloor \frac{t}{\alpha} \rfloor} \mathbb{E}\left[e^{c(1+X_i)}\right] \leq e^{\frac{\alpha}{c}(1+\frac{t}{\alpha})} = e^{\frac{\alpha+t}{c}}.$$

This completes the proof since we chose $\alpha \leq 1$. $\square$

Remark 5.6. In particular, for every $t \geq 1$, we have

$$\mathbb{E}\left[e^{c\frac{N_t}{t}}\right] \overset{(Jensen)}{\leq} \mathbb{E}\left[e^{cN_t}\right]^{\frac{1}{t}} \leq e^{\frac{2}{c}}$$

and for every $k \geq 1$

$$\mathbb{E}\left[\left(\frac{N_t}{t}\right)^k\right] \leq \frac{k!}{c^k} e^{\frac{2}{c}}. \tag{5.2}$$

5.6 Law of Large Numbers

Theorem 5.7 (Law of Large Numbers)**.** *Recall that $\mu = \mathbb{E}[T_1]$. We have*

$$\lim_{t\to\infty} \frac{N_t}{t} = \frac{1}{\mu} \ a.s.$$

Remark 5.8. If $\mu = \infty$, then $\lim_{t\to\infty} \frac{N_t}{t} = 0$ a.s.

Proof. By the strong law of large numbers (for non negative random variable), we have

$$\lim_{n\to\infty} \frac{S_{n+1}}{n+1} = \lim_{n\to\infty} \frac{S_n}{n+1} = \mu \qquad \text{a.s.}$$

Notice that for every t

$$S_{N_t} \le t \le S_{N_t+1}.$$

Therefore,

$$\underbrace{\frac{S_{N_t}}{N_t+1}}_{\to\mu} \le \frac{t}{N_t+1} < \underbrace{\frac{S_{N_t+1}}{N_t+1}}_{\to\mu}.$$

Where the convergences are almost sure. Therefore $\lim_{t\to\infty} \frac{1+N_t}{t} = \frac{1}{\mu}$ a.s., which implies that $\lim_{t\to\infty} \frac{N_t}{t} = \frac{1}{\mu}$ a.s.

□

Theorem 5.9 (Central Limit Theorem)**.** *Assume that $\mathbb{E}[T_1^2] < \infty$. Write $\mu = \mathbb{E}[T_1]\,,\ \sigma^2 = Var(T_1)$. Then, assuming $\sigma > 0$, we have*

$$\frac{N_t - \frac{t}{\mu}}{\sigma\sqrt{\frac{t}{\mu^3}}} \xrightarrow[t\to\infty]{(law)} \mathcal{N}(0,1)$$

Proof. See exercises. □

5.7 Renewal function

Definition 5.3. The renewal function is the function $m : \mathbb{R}_+ \to \mathbb{R}_+$ defined by

$$\boxed{\forall t \geq 0 \quad m(t) = \mathbb{E}[N_t].}$$

Remark 5.10. Equation (5.2) applied to $k = 1$ implies that $m(t) < \infty$ for every $t \geq 0$.

Interpretation: The set $\{S_1, S_2, \ldots\}$ of renewal times defines a set of random points in $\mathbb{R}_+$, and

$$m(t) = \mathbb{E}\left[\text{Number of points in the interval } [0,t]\right].$$

Remark 5.11. For the Poisson process with parameter λ, we know (by Proposition 5.1) that $N_t \sim \text{Pois}(\lambda t)$. Therefore, the renewal function is linear in this case:

$$\forall t \geq 0 \quad m(t) = \lambda t.$$

Proposition 5.12. *The renewal function m is non-decreasing, non-negative, and right continuous.*

Proof. Since N_t is non-decreasing in t and non-negative almost surely, the expectation $m(t) = \mathbb{E}[N_t]$ also satisfies these two properties. For the right continuity, observe that almost surely $N_{t+s} - N_t \downarrow 0$ as $s \downarrow 0$. Therefore $m(t+s) - m(t) \to 0$ by monotone convergence. □

5.8 Elementary renewal theorem

Theorem 5.13 (Elementary Renewal Theorem)**.**

$$\boxed{\lim_{t\to\infty} \frac{m(t)}{t} = \frac{1}{\mu}.}$$

Proof. We already have $\lim_{t\to\infty} \frac{N_t}{t} = \frac{1}{\mu}$ a.s. (by Theorem 5.7). Furthermore, we have seen that $\sup_{t\geq 1} \mathbb{E}\left[\left(\frac{N_t}{t}\right)^2\right] < \infty$. Hence $\frac{N_t}{t}$ is uniformly integrable and

$$\lim_{t\to\infty} \frac{m(t)}{t} = \lim_{t\to\infty} \mathbb{E}\left[\frac{N_t}{t}\right] = \mathbb{E}\left[\lim_{t\to\infty} \frac{N_t}{t}\right] = \frac{1}{\mu}.$$

□

5.9 Lattice distributions

Definition 5.4. We say that F is lattice if there exists $a > 0$ and such that

$$\mathbb{P}\left[T_1 \in a\mathbb{Z}\right] = 1. \tag{5.3}$$

In this case the span of F is defined as the largest $a > 0$ such that (5.3) holds. Otherwise, we say that F is non lattice.

5.10 Blackwell's renewal theorem: lattice case

Theorem 5.14 (Blackwell's Renewal Theorem)**.** *Assume that the law of T_1 is lattice with span a, then the sequence $(m(ai))_{i\in\mathbb{N}}$ satisfies*

$$\boxed{\lim_{i\to\infty} m(a\cdot i) - m(a\cdot(i-1)) = \frac{a}{\mu}.}$$

Proof. Via Markov Chains, see exercises. □

5.11 Blackwell's renewal theorem: non-lattice case

Theorem 5.15 (Blackwell's Renewal Theorem)**.** *Assume that the law of T_1 is non-lattice, then for all $h \geq 0$*

$$\boxed{\lim_{t\to\infty} m(t+h) - m(t) = \frac{h}{\mu}.}$$

Proof. Admitted. □

Remark 5.16. Blackwell's theorem is "stronger" than elementary renewal theorem:

$$\frac{m(t)}{t} \approx \frac{m(\lfloor t\rfloor)}{\lfloor t\rfloor} = \frac{1}{\lfloor t\rfloor}\sum_{k=1}^{\lfloor t\rfloor} m(k) - m(k-1) \overset{\text{(Blackwell)}}{\to} \frac{1}{\mu}.$$

Chapter 6

Renewal Equation

Framework: $(\Omega, \mathcal{F}, \mathbb{P})$ probability space. In the whole chapter, we fix

$$T_1, T_2, \ldots \text{ i.i.d. random variables on } \mathbb{R}_+$$

satisfying $\mathbb{P}[T_1 = 0] < 1$. We write

$$\mu = \mathbb{E}[T_1] \in (0, \infty] \qquad \text{and} \qquad F(t) = \mathbb{P}[T_1 \leq t].$$

6.1 Lesbesgue-Stieltjes measure

Theorem 6.1. *Let g be a right continuous non-decreasing function on $\mathbb{R}_+$. There exists a unique measure ν_g on $\mathbb{R}_+$ such that*

$$\forall t \geq 0 \quad \nu_g([0, t]) = g(t).$$

Proof. Admitted (follows from Caratheordory's extension Theorem). □

Notation Let g be a right continuous non-decreasing function on $\mathbb{R}_+$. For $h \in L^1(\nu_g)$ or h measurable and non-negative, write

$$\int_{\mathbb{R}_+} h \, dg = \int_{\mathbb{R}_+} h \, d\nu_g.$$

Example 1: F is a right continuous non-decreasing function on $\mathbb{R}_+$ and ν_F corresponds to the law of T_1: for every $B \subset \mathbb{R}_+$ measurable,

$$\nu_F(B) = \mathbb{P}[T_1 \in B].$$

Furthermore, for every h measurable bounded, we have

$$\int_{\mathbb{R}_+} h\,dF = \mathbb{E}[h(T_1)].$$

Example 2: Proposition 5.12 states that the renewal function m is right-continuous non-decreasing. The corresponding measure ν_m has the following interpretation: for every $B \subset \mathbb{R}_+$ measurable,

$$\nu_m(B) = \mathbb{E}\left[\text{Number of renewals in } B\right].$$

Furthermore, for every h measurable bounded, we have

$$\int_{\mathbb{R}_+} h\,dm = \mathbb{E}\Big[\sum_{k\geq 1} h(S_n)\Big].$$

6.2 Convolution operator

Definition 6.1 (Convolution operator)**.** Let G be a right continuous non-decreasing function on $\mathbb{R}_+$. Let $h : \mathbb{R}_+ \to \mathbb{R}$ measurable be such that for all $t \geq 0$ $\int_0^t |h(t-s)|dG(s) < \infty$ or h measurable non-negative. For every $t \geq 0$, define

$$(h * G)(t) = \int_0^t h(t-s)dG(s).$$

Remark 6.2. If X, Y are two independent random variables on $\mathbb{R}_+$ with distribution functions F_X, F_Y respectively, then

$$F_{X+Y} = F_X * F_Y.$$

The proof is left as an exercise.

This is useful in our context to express the distribution of the n-th renewal time $S_n = T_1 + \dots T_n$ for $n \geq 1$. Using the remark above and an induction, we can express the distribution function of S_n as a n-fold convolution:

$$F_{S_n} = F_{T_1+\dots+T_n} = F^{*n},$$

where we write $F^{*n} = \underbrace{F * \dots * F}_{n \text{ times}}$.

This leads directly to the following expression of the renewal function.

Proposition 6.3. *For every* $t \geq 0$

$$m(t) = \sum_{k \geq 1} F^{*k}(t).$$

Proof. For every $t \geq 0$, we have

$$m(t) = \mathbb{E}\left[\sum_{n \geq 1} \mathbf{1}_{S_n \leq t}\right] = \sum_{n \geq 1} \mathbb{P}\left[S_n \leq t\right] = \sum_{n \geq 1} F^{*n}(t).$$

□

6.3 Renewal equation

Definition 6.2. Let $h : \mathbb{R}_+ \to \mathbb{R}$ be measurable locally bounded (i.e. $\forall t \geq 0$, $h|_{[0,t]}$ is bounded). Let $g : \mathbb{R}_+ \to \mathbb{R}$ such that for all $t \geq 0$ $\int_0^t |g(t-s)| dF(s) < \infty$. We say that g is a solution of the (h, F) renewal equation if

$$\boxed{\forall t \geq 0 \quad g(t) = h(t) + \int_0^t g(t-s) dF(s),}$$

i.e. $g = h + g * F$.

Proposition 6.4. *m is a solution of the* (F, F) *renewal equation, ie.* $m = F + m * F$.

Proof 1.

$$m = \sum_{i > 0} F^{*i} = F + \sum_{i > 1} F^{*(i-1)} * F \overset{\text{monotone cv.}}{=} F + \underbrace{\left(\sum_{i > 1} F^{*(i-1)}\right)}_{m} * F.$$

□

Proof 2. For $t \geq 0$, we have

$$m(t) = \mathbb{E}\left[\sum_{k>0} \mathbf{1}_{T_1+\ldots+T_k \leq t}\right] = \mathbb{P}\left[T_1 \leq t\right] + \underbrace{\mathbb{E}\left[\sum_{k>1} \mathbf{1}_{T_1+\ldots+T_k \leq t}\right]}_{(\star)}$$

$$\begin{aligned}(\star) &\overset{\text{(Fubini)}}{=} \sum_{k>1} \mathbb{E}\left[\mathbf{1}_{T_1+\ldots+T_k \leq t}\right] \overset{\text{(Indep.)}}{=} \sum_{k>1} \int_0^t \mathbb{E}\left[\mathbf{1}_{s+T_2+\ldots+T_k \leq t}\right] dF(s) \\ &= \int_0^t m(t-s) dF(s).\end{aligned}$$

□

6.4 Excess time

For $t \geq 0$, define

$$E_t = S_{N_t+1} - t,$$

the time left to wait until next renewal.

Proposition 6.5 (Excess distribution function)**.** *Fix $x \geq 0$. The function e_x defined by $e_x(t) = \mathbb{P}\left[E_t \leq x\right]$ for all $t \geq 0$ satisfies*

$$e_x = h_x + e_x * F,$$

where $h_x(t) = F(x+t) - F(t)$. (i.e. e_x is a solution of the (h_x, F) renewal equation).

Proof. Fix $x, t \geq 0$. We can separate $e_x(t)$ into two parts, one for the probability if there has already been a renewal before time t, and one if that has not occurred:

$$e_x(t) = \mathbb{P}\left[T_1 > t, E_t \leq x\right] + \mathbb{P}\left[T_1 \leq t, E_t \leq x\right].$$

Now we analyze each term separately. The first term can be directly expressed as

$$\mathbb{P}\left[T_1 > t, T_1 \leq t + x\right] = F(t+x) - F(t).$$

For the second term, we exploit the renewal structure of the process. Observe that E_t is measurable with respect to $(T_1, T_2, \ldots)$: by definition, we have $E_t = \phi_t(T_1, T_2, \ldots)$, where

$$\phi_t(t_1, t_2, \ldots) = \sum_{n \geq 0} \mathbf{1}_{t_1+\cdots+t_n \leq t, t_1+\cdots+t_{n+1} > t}(t_1 + \cdots + t_{n+1} - t).$$

Notice that for every $s \leq t$, $\phi_t(s, t_2, \ldots) = \phi_{t-s}(t_2, \ldots)$. Using this observation, we find

$$\begin{aligned}\mathbb{P}[T_1 \leq t, E_t \leq x] =& \mathbb{P}[T_1 \leq t, \phi_t(T_1, T_2, \ldots) \leq x] \\ =& \int_0^t \mathbb{P}[\phi_t(s, T_2, \ldots) \leq x] \, dF(s) \\ =& \int_0^t \mathbb{P}[\phi_{t-s}(T_2, \ldots) \leq x] \, dF(s) \\ =& \int_0^t e_x(t-s) dF(s) = (e_x * F)(t)\end{aligned}$$

Thus $e_x(t) = h_x(t) + (e_x * F)(t)$. □

6.5 Well-Posedness of the Renewal Equation

Theorem 6.6. *Let $h : \mathbb{R}_+ \to \mathbb{R}$ be measurable, locally bounded. Then there exists a unique $g : \mathbb{R}_+ \to \mathbb{R}$ measurable, locally bounded, solution of*

$$g = h + g * F,$$

given by $g = h + h * m$.

Intuitive Proof. Assume g is a solution, then we have

$$\begin{aligned}g =& h + g * F \\ =& h + (h + g * F) * F \\ & \vdots \\ \overset{(*)}{=}& h + h * F + h * F^{*2} + h * F^{*3} + \ldots \\ =& h + h * m\end{aligned}$$

□

Rigorous Proof. **Existence** $g = h + h * m$ is measurable and locally bounded, because h is. We have

$$\begin{aligned}h + g * F &= h + (h + h * m) * F \\ &= h + h * \underbrace{(F + m * F)}_{=m} = g.\end{aligned}$$

Uniqueness Let g_1, g_2 be two solutions of the (h, F) renewal equation. Then $g_1 - g_2 = (g_1 - g_2) * F$ and therefore, by induction, $g_1 - g_2 = (g_1 - g_2) * F^{*n}$ for every $n \geq 1$. Fix $t \geq 0$. For every $n \geq 1$, we have

$$|g_1(t) - g_2(t)| = \left| \int_0^t (g_1 - g_2)(t - s) dF^{*n}(s) \right| \leq \sup_{[0,t]} |g_1 - g_2| \int_0^t dF^{*n}(s).$$

Where we can see the integral term is equal to $\mathbb{P}[T_1 + \ldots + T_n \leq t]$ which converges to 0 as n tends to infinity. Hence $g_1 = g_2$. □

6.6 Discussion about the asymptotic Behavior

From now and until the end of the chapter, we assume that F is non-lattice.

Question: Let g be the solution of the (h, F) renewal equation, what is the asymptotic behavior of $g(t)$ for $t \to \infty$?

A first answer: We start by considering the case $h = \mathbf{1}_{[a,b]}$ for $0 \leq a \leq b$. Let $g = h + h * m$ be the solution of the (h, F) renewal equation. For every $t > b$, we have $h(t) = 0$, hence

$$\begin{aligned} g(t) &= \int_0^t h(t - s) dm(s) \\ &= \int_{t-b}^{t-a} h(s) dm(s) \\ &= \underbrace{m(t - a) - m(t - b)}_{\overset{\text{(Blackwell)}}{\to} \frac{b-a}{\mu}}. \end{aligned}$$

Hence

$$\lim_{t \to \infty} g(t) = \frac{1}{\mu} \int_0^\infty h(s) ds.$$

How does this generalize?

Idea: Extend to simple functions $\sum \lambda_i \mathbf{1}_{[a_i, b_i]}$ (this is straightforward), and then to a more general class of measurable functions. A good framework for this extension is to consider directly Riemann integrable functions.

6.7 Directly Rieman integrable functions

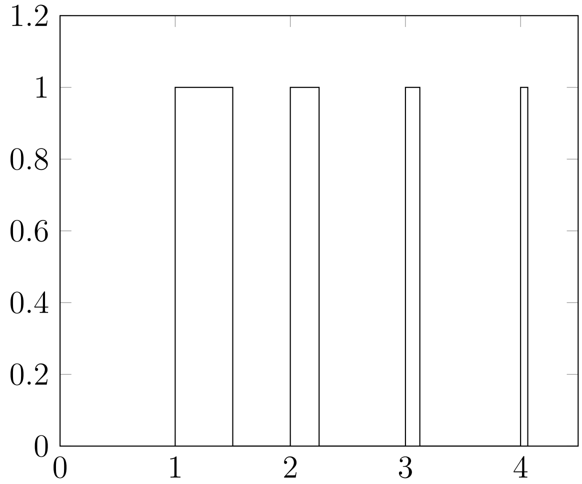


Figure 6.1: An integrable function which is not dRi.

Definition 6.3. $h : \mathbb{R}_+ \to \mathbb{R}_+$ measurable, h is called *directly Riemann Integrable* (dRi) if

$$\forall \delta > 0 \quad \sum_{k=0}^{\infty} \delta \sup_{[k\delta,(k+1)\delta]} h < \infty.$$

and

$$\lim_{\delta \to 0} \delta \sum_{k=0}^{\infty} \sup_{[k\delta,(k+1)\delta]} h = \lim_{\delta \to 0} \delta \sum_{k=0}^{\infty} \inf_{[k\delta,(k+1)\delta]} h.$$

$h : \mathbb{R}_+ \to \mathbb{R}$ is dRi if and only if $h_+ = \max(h, 0)$ and $h_- = \max(-h, 0)$ are dRi.

Remark 6.7. If h is dRi, then it is integrable. The converse is not true: The function $h = \sum_{k>0} \mathbf{1}_{[k,k+2^{-k}]}$ is integrable, but is not dRi.

Proposition 6.8. *Let $h : \mathbb{R}_+ \to \mathbb{R}_+$ be measurable.*

Assume that h is continuous at a.e. $t \in \mathbb{R}$ and there exists H non-increasing such that $0 \le h \le H$ and $\int_0^\infty H < \infty$. Then h is dRi.

Proof. See Prop. 4.1 in [1].

□

Remark 6.9. In particular if h is bounded, continuous at a.e. $t \in \mathbb{R}$, and vanishes outside a compact set, then h is dRi.

6.8 Smith key renewal theorem

Theorem 6.10 (Smith Key Renewal Theorem)**.** *Let h be dRi, F non-lattice. Then $g = h + h * m$ satisfies*

$$\boxed{\lim_{t\to\infty} g(t) = \frac{1}{\mu}\int_0^\infty h(u)du.}$$

Remark 6.11. The case $h = \mathbf{1}_{[0,b]}$ corresponds to the Blackwell Theorem.

Proof. Since h is dRi we have

$$\sum_k \sup_{[k,k+1]} |h| < \infty.$$

Hence $h(t) \to 0$. Therefore it suffices to prove

$$\lim_{t\to\infty} \int_0^t h(t-s)dm(s) = \frac{1}{\mu}\int_0^t h(u)du.$$

Let $\delta > 0$ such that $F(\delta) < 1$.

Assume $h = \sum_{k\geq 0} c_k \mathbf{1}_{[k\delta,(k+1)\delta)}$ with $c_k \geq 0$ and $\sum_{k\geq 0} c_k < \infty$. By monotone convergence

$$h(t-s)dm(s) = \sum_{k\geq 0} \underbrace{c_k[m(t-k\delta) - m(t-k\delta-\delta)]}_{h_k(t)}.$$

Observe that for every $u \geq \delta$

$$\begin{aligned} 1 \geq F(u) = m(u) - \int_0^u F(u-s)dm(s) &= \int_0^u (1-F(u-s))dm(s) \\ &\geq \int_{u-\delta}^u \underbrace{(1-F(u-s))}_{\geq 1-F(\delta)}dm(s) \geq (1-F(\delta))\,(m(u)-m(u-\delta))\,. \end{aligned}$$

In the first equality, it was used that m is the solution of the (F,F) renewal equation. Hence for every t and every k

$$h_k(t) \leq \frac{c_k}{1-F(\delta)},$$

by distinguishing between $t - k\delta \geq \delta$ and $t - k\delta < \delta$, and using that m is non-decreasing, vanishing on $(-\infty, 0)$. By dominated convergence

$$\lim_{t\to\infty} \sum_{k\geq 0} h_k(t) = \sum_{k\geq 0} \underbrace{\lim_{t\to\infty} h_k(t)}_{\overset{\text{(Blackwell)}}{=} c_k \frac{\delta}{\mu}} .$$

Hence $\lim_{t\to\infty} \int_0^t h(t-s)dm(s) = \sum_{k=0}^{\infty} c_k \frac{\delta}{\mu} = \frac{1}{\mu}\int_0^\infty h(u)du$.

Now assume $h \geq 0$ dRi. Let $\delta > 0$ such that $F(\delta) < 1$. Write

$$\underline{h}_\delta = \sum_{k\geq 0} (\inf_{[k\delta,(k+1)\delta]} h) \mathbf{1}_{[k\delta,(k+1)\delta)}$$

$$\overline{h}_\delta = \sum_{k\geq 0} (\sup_{[k\delta,(k+1)\delta]} h) \mathbf{1}_{[k\delta,(k+1)\delta)}.$$

We have for every t

$$\int_0^t h(t-s)dm(s) \leq \int_0^t \overline{h}_\delta(t-s)dm(s) \to \frac{1}{\mu}\int_0^t \overline{h}_\delta(u)du.$$

Hence

$$\limsup_{t\to\infty} \int_0^t h(t-s)dm(s) \leq \frac{1}{\mu}\int_{\mathbb{R}} \overline{h}_\delta(u)du.$$

Since

$$\left| \int_{\mathbb{R}} \overline{h}_\delta(u)du - \int_{\mathbb{R}} h(u)du \right| \leq \sum_{k\geq 0} \delta \left(\overline{h}_\delta(k\delta) - \underline{h}_\delta(k\delta) \right) \overset{\delta\to 0}{\longrightarrow} 0,$$

where the limit is due to h being dRi. We can let δ tend to 0 in the equation above (with lim sup) to obtain

$$\limsup_{t\to\infty} \int_0^t h(t-s)dm(s) \leq \frac{1}{\mu}\int_{\mathbb{R}} h(u)du,$$

and equivalently

$$\liminf_{t\to\infty} \int_0^t h(t-s)dm(s) \geq \frac{1}{\mu}\int_{\mathbb{R}} h(u)du.$$

$$\frac{1}{\mu}\int_{\mathbb{R}} h(u)du \leq \liminf_{t\to\infty} \int_0^t h(t-s)dm(s) \leq \limsup_{t\to\infty} \int_0^t h(t-s)dm(s) \leq \frac{1}{\mu}\int_{\mathbb{R}} h(u)du.$$

□

6.9 Application to the excess time

Assume that $\mu < \infty$. Let E_t be the excess time (time until next renewal) and $e_x(t) = \mathbb{P}[E_t \leq x]$. What is $\lim_{t\to\infty} e_x(t)$? We know that $e_x = h_x + e_x * F$, where $h_x(t) = F(t+x) - F(t)$.

Remark 6.12. $\mu = \mathbb{E}[T_1] = \int_0^\infty \mathbb{P}[T_1 > t]\, dt$

With this we have that $h_x(t) \leq 1 - F(t) = \mathbb{P}[T_1 > t]$, and $1 - F(t)$ is non-increasing in t and continuous a.e. (because it is the difference of two monotone functions).

$$\int_0^\infty \mathbb{P}[T_1 > t]\, dt = \mathbb{E}[T_1] = \mu < \infty.$$

Thus (by the proposition) h_x is dRi. Now we can apply the theorem and get that

$$\lim_{t\to\infty} \mathbb{P}[E_t \leq x] = \frac{1}{\mu}\int_0^\infty h_x(t)dt = \frac{1}{\mu}\int_0^\infty F(t+x) - F(t)dt,$$

with $F(t+x) - F(t) = \mathbb{E}\left[\mathbf{1}_{T_1\in(t,t+x]}\right]$, we find that the limit is equal to

$$\frac{1}{\mu}\int_0^\infty \mathbb{E}\left[\mathbf{1}_{T_1\in(t,t+x]}\right] dt = \frac{1}{\mu}\mathbb{E}\left[\int_0^\infty \mathbf{1}_{t\in[T_1-x,T_1)}\right] dt = \frac{1}{\mu}\mathbb{E}\left[\int_{\max\{T_1-x,0\}}^{T_1} dt\right] = \begin{cases} T_1, & T_1 \leq x \\ x, & T_1 > x. \end{cases}$$

Thus for t large: $\mathbb{P}[E_t \leq x] \approx \frac{1}{\mu}\mathbb{E}[\min\{T_1, x\}]$.

Remark 6.13. $G(x) = \frac{1}{\mu}\mathbb{E}[\min\{T_1, x\}]$ is the delay distribution in the proof of Blackwell's Theorem.

Chapter 7

General Poisson Point Processes

Reference Lectures on the Poisson Process (Penrose), Poisson Processes (Kingman)

Framework:

- $(\Omega, \mathcal{F}_\Omega, \mathbb{P})$ probability space.
- (E, d) a Polish space (separable, complete, metric space).
- $\mathcal{E}$ Borel σ-algebra of E.
- μ sigma-finite measure on $(E, \mathcal{E})$, i.e. there exists a partition

$$E = \bigcup_{i \in \mathbb{N}} E_i,$$

  such that each E_i is measurable and satisfies $\mu(E_i) < \infty$.

Examples:

(i) $E = \{0\}$, $\mu = \delta_0$.

(ii) $E = \mathbb{R}_+$, $\mu = \lambda \cdot \mathsf{Leb}_{\mathbb{R}_+}$ "Lebesgue Measure on $\mathbb{R}_+$.

(iii) $E = \mathbb{R}^2$, $\mu(dx) = \frac{1}{\pi} e^{-|x|^2} dx$ 'Gaussian'

Goal: We wish to define a random set of points on $(E, \mathcal{E})$ where

$$\text{"number of points around } x\text{"} \approx \mu(dx).$$

In particular we wish to define a random variable: $\Omega \to$'set of points in a general state space E' (ex: $\mathbb{R}^2$, $[0,1]^2$, a manifold, $\mathbb{Z}$, a space of function,etc...)

7.1 Representing Points?

First question How can we represent points on $E = \mathbb{R}_+$ mathematically?

(i) 'Time point of view', ie $T_1, T_2, \ldots$ where T_i = time between the $(i-1)$'th and i'th point.

(ii) Cadlag formulation with values in $\mathbb{N}$. N_t = number of points in $[0, t]$.

(iii) A set of points $\mathcal{S} = \{S_1, S_2, \ldots\}$

(iv) Measure $M : \mathcal{B}(\mathbb{R}_+) \to \mathbb{N}$ with $M(A)$ = number of points in A.

(i) and (ii) are specific to $\mathbb{R}_+$ and to not extend to general space. (iii) and (iv) are both possible. We will prefer (iv) because it allows us to deal with multiplicity.

Notation We consider the measurable space $(\mathcal{M}, \mathcal{B}(\mathcal{M}))$, where

$$\boxed{\mathcal{M} = \{\text{sigma-finite measures } \eta \text{ on } E \text{ such that } \forall B \in \mathcal{E}\ \eta(B) \in \mathbb{N} \cup \{+\infty\}\}},$$

and $\mathcal{B}(\mathcal{M})$ is the σ-algebra generated by the sets

$$\{\eta \in \mathcal{M} : \eta(B) = k\}$$

for $B \subset E$ measurable and $k \in \mathbb{N}$.

Proposition 7.1 (Representation as Dirac Sum). *Let $\mathcal{M}_{<\infty} = \{\eta \in \mathcal{M} : \eta(E) < \infty\}$, there exist measurable maps $\tau : \mathcal{M}_{<\infty} \to \mathbb{N}$ and $X_i : \mathcal{M}_{<\infty} \to E$ such that*

$$\forall \eta \in \mathcal{M}_{<\infty} \quad \eta = \sum_{i=0}^{\tau(\eta)} \delta_{X_i(\eta)}.$$

Remark 7.2. Thus η corresponds to a collection of points $\{X_1, \ldots, X_\tau\}$.

Notation: For every $k \geq 0$ we write $\mathcal{M}_k$ for the set of measures $\eta \in \mathcal{M}$ with total mass $\eta(E) = k$.

Lemma 7.3. *Let $k \geq 1$. There exists a measurable map $Z : \mathcal{M}_k \to E$ such that*

$$\forall \eta \in \mathcal{M}_k \quad \eta(\{Z\}) \geq 1.$$

Proof. Fix $k \geq 1$ and $\mathcal{Y} = \{y_1, y_2, \dots\}$ at most countable and dense in E. We will construct by induction $Y_1, Y_2, \dots$ some measurable maps from $\mathcal{M}_k$ to $\mathcal{Y}$ such that for every $n \geq 1$

$$\eta\Big(\bigcap_{1\leq m\leq n} \mathsf{B}(Y_m(\eta), \frac{1}{m})\Big) \geq 1,$$

for every $\eta \in \mathcal{M}_k$.

Construction of Y_1: Since the set $\mathcal{Y}$ is dense in E, we have $E = \bigcup_{i>0} B(y_i, 1)$. Therefore, for every $\eta \in \mathcal{M}_k$, by the union bound we have $1 \leq \eta(E) \leq \sum_{i\geq 1} \eta(\mathsf{B}(y_i, 1))$. We can thus define

$$Y_1(\eta) = y_{i_1} \quad \text{where } i_1 = \min\{i : \ \eta(B(y_i, 1)) \geq 1\}.$$

This define a map $Y_1 : \mathcal{M}_k \to \mathcal{Y}$, which is measurable because for every j

$$Y_1^{-1}(\{y_j\}) = \bigcap_{i<j} \{\eta \ : \ \eta(B(y_i, 1)) = 0\} \cap \{\eta \ : \ \eta(B(y_i, 1)) = 1\}.$$

Construction of Y_n: Let $n \geq 1$ and assume that $Y_1, \dots, Y_{n-1}$ have already been constructed. Let $\eta \in \mathcal{M}_k$ and $C = \bigcap_{1\leq m\leq n-1} \mathsf{B}(Y_m(\eta), \frac{1}{m})$. We have

$$1 \leq \eta(C) \leq \sum_{i>0} \eta\left(C \cap \mathsf{B}\left(y_i, \frac{1}{n}\right)\right).$$

Define $Y_n(\eta) = y_{i_n}$ where $i_n = \min\{i : \ \eta(C \cap \mathsf{B}(y_i, \frac{1}{n})) \geq 1\}$. As above, Y_n is measurable.

The sequence $(Y_n)_{n\geq 0}$ constructed above is a Cauchy sequence (indeed for every $n \geq m$ $\mathsf{B}(Y_n, \frac{1}{n}) \cap \mathsf{B}(Y_m, \frac{1}{m}) \neq \emptyset$, hence by the triangle inequality $d(Y_n, Y_m) \leq \frac{2}{m}$). Define $Z_{k+1}(\eta) = \lim_{n\to\infty} Y_n(\eta)$ (Z_{k_1} is measurable as a simple limit of measurable functions). Furthermore $\{Z_{k+1}(\eta)\} = \bigcap_{n>0} B(Y_n, \frac{2}{n})$ and therefore $\eta(\{Z_{k+1}(\eta)\}) \geq 1$.

□

Proof of Proposition 7.1. We have $\mathcal{M}_{<\infty} = \bigcup_{k=0}^{\infty} \mathcal{M}_k$ where $\mathcal{M}_k = \{\eta : \ \eta(E) = k\}$. We prove by induction on $k \geq 0$ that for every $k \geq 0$ there exist $Z_1, \dots, Z_k : \ \mathcal{M}_k \to E$ measurable such that

$$\forall \eta \in \mathcal{M}_k \quad \eta = \sum_{i=1}^{k} \delta_{Z_i}.$$

For $k = 0$ there is nothing to prove. Let $k \geq 0$ and assume that the property holds. Let $\eta \in \mathcal{M}$ such that $\eta(E) = k + 1$. By Lemma 7.3, there exists $Z_{k+1} : \mathcal{M}_{k+1} \to E$ measurable such that $\eta(Z_{k+1}(\eta)) \geq 1$. Define

$$\eta' = \eta - \delta_{Z_{k+1}(\eta)}$$

(η' is measurable in η). Note that $\eta'(E) = k$, and therefore $\eta' \in \mathcal{M}_k$. By induction, there exist $Z_1'(\eta'), \ldots, Z_k'(\eta')$ such that $\eta' = \delta_{Z_1'} + \ldots + \delta_{Z_k'}$. Setting $Z_i(\eta) := Z_i'(\eta')$ for $i \leq k$, we obtain

$$\eta = \sum_{i=1}^{k+1} \delta_{Z_i(\eta)}.$$

□

7.2 Point process

Definition 7.1. A <u>point process</u> on $(E, \mathcal{E})$ is a stochastic process

$$M = (M(B))_{B \in \mathcal{E}}$$

with values in $\mathbb{N} \cup \{\infty\}$, such that $M \in \mathcal{M}$ a.s.

Interpretation: For fixed B, the random integer $M(B)$ intuitively represents the number of points in B. A point process indicates how many points there are in each region B of the space. The condition $M \in \mathcal{M}$ a.s. ensures that all the numbers of points in different regions are compatible with each other.

Remark 7.4. In the definition above, we make a slight abuse of notation and also write M for the random mapping $M : B \mapsto M(B)$.

As usual in probability, the underlying parameter $\omega \in \Omega$ is implicit. Formally, a point process is a collection $M = (M_\omega(B))_{\omega \in \Omega, B \in \mathcal{E}}$ with values in $\mathbb{N} \cup \{\infty\}$ such that

- for every fixed B, $\omega \mapsto M_\omega(B)$ is measurable.
- for almost every $\omega \in \Omega$, the mapping $M_\omega : B \mapsto M_\omega(B)$ is an element of $\mathcal{M}$.

Remark 7.5. One can check that the definition above is equivalent to saying that the mapping $\omega \mapsto M_\omega$ is a random variable with values in $\mathcal{M}$.

Examples of Point Processes

- $M = 0$ a.s. (This corresponds to the random set $\mathcal{S} = \emptyset$ a.s.)
- $E = [0, 1]$, X random variable on $[0, 1]$. $M = \delta_X$ is a point process. (This corresponds to the random set $\mathcal{S} = \{X\}$ a.s.)
- $X_1, \ldots X_n$ i.i.d. random variable on $[0, 1]$, $N = \delta_{X_1} + \ldots + \delta_{X_n}$ is a point process. (This corresponds to the random set $\mathcal{S} = \{X_1, \ldots, X_n\}$ a.s.)

7.3 Poisson Point Processes

Convention $X \sim \text{Pois}(\infty)$ if and only if $X = \infty$ a.s.

Definition 7.2. A Poisson point process with intensity μ on $(E, \mathcal{E})$ (ppp(μ)) is a point process M such that

(i) For all $B_1, \ldots, B_k \subset E$ measurable and disjoint, $M(B_1), \ldots, M(B_k)$ are independent.

(ii) For all $B \subset E$ measurable, $M(B)$ has law $\text{Pois}(\mu(B))$.

Remark 7.6. Let $B \subset E$ measurable. Item (ii) includes the case $\mu(B) = \infty$: it is equivalent to

$$\mu(B) \begin{cases} \sim \text{Pois}(\mu(B)) & \text{if } \mu(B) < \infty, \\ = +\infty \text{ a.s.} & \text{if } \mu(B) = \infty. \end{cases}$$

In particular, by applying the definition to $B = E$, we obtain that the total number of points in the space $\tau := M(E)$ is a Poisson random variable with parameter $\mu(E)$: we have

$$\tau \begin{cases} < \infty \text{ a.s.} & \text{if } \mu(E) < \infty, \\ = +\infty \text{ a.s.} & \text{if } \mu(E) = \infty. \end{cases}$$

Remark 7.7. Thanks to Item (ii), we can calculate the average number of points in a region. For every $B \subset E$ measurable, we have

$$\boxed{\mathbb{E}\left[M(B)\right] = \mu(B)},$$

(on average, there are $\mu(B)$ points in B).

7.4 Representation as a proper process

Theorem 7.8. *Let M be a ppp(μ) on $(E, \mathcal{E})$. Let $\tau = M(E)$ (the total number of points in E). There exist some random variables $X_n \in E$, $n > 0$ such that*

$$M = \sum_{n=1}^{\tau} \delta_{X_n} \quad a.s.$$

Remark 7.9. The theorem gives a "random set" interpretation of Poisson process. We have a correspondence:

$$\begin{aligned} M \in \mathcal{M} \text{ rand. counting measure} \quad &\longleftrightarrow \quad \mathcal{S} = \{X_1, \dots, X_\tau\} \text{ random set} \\ M(B) \quad &\longleftrightarrow \quad |\mathcal{S} \cap B|, \text{ number of points in } B \text{ (with multiplicity).} \end{aligned}$$

Proof of Theorem 7.8. Let $(E_i)_{i\in\mathbb{N}}$ be a partition of E such that $\mu(E_i) < \infty$ for every i. The process $M_i := M(\,\cdot\, \cap E_i)$ takes values in $\mathcal{M}_{<\infty}$. Hence the proposition in the previous section ensures that there exist some random variables $\tau^{(i)}, Z_1^{(i)}, \dots, Z_\tau^{(i)}$ such that

$$M_i = \sum_{j=1}^{\tau^{(i)}} \delta_{Z_j^{(i)}} \text{ a.s.}$$

Use that $M = \sum_{i=1}^{\infty} M_i$, and a reordering of the terms in the sums, we obtain the desired result. □

Question Does there always exist a ppp(μ) on E?

7.5 Existence: Spaces with finite measure

Assume $\mu(E) < \infty$.

Proposition 7.10. *Let Z, $(X_i)_{i\geq 1}$ be independent random variables.*

$$Z \sim \text{Pois}(\mu(E)), \quad X_i \sim \frac{\mu(\,\cdot\,)}{\mu(E)}.$$

Then $M = \sum_{i=1}^{Z} \delta_{X_i}$ is a ppp(μ) on E.

Proof. Let $k \geq 2$ and $B_1, \dots B_{k-1} \subset E$ be disjoint and measurable. Set $B_k = E \setminus \left(\bigcap_{i=1}^k B_i\right)$.

Fix $n_1, \dots, n_k \in \mathbb{N}$ arbitrary. Set $n = n_1 + \dots + n_k$, and define for each $i \in \{1, \dots, k\}$,

$$Y_i = \sum_{j=1}^{n} \mathbb{1}_{X_j \in B_i}$$

Observe that $(Y_1, \dots, Y_k)$ is a multinomial random variable with parameters $(n; \frac{\mu(B_1)}{\mu(E)}, \dots, \frac{\mu(B_k)}{\mu(E)})$ independent of Z. We have

$$\begin{aligned}\mathbb{P}\left[M(B_1) = n_1, \dots, M(B_k) = n_k\right] &= \mathbb{P}\left[Z = n, Y_1 = n_1, \dots, Y_k = n_k\right] \\ &= \frac{\mu(E)^n}{n!} e^{-\mu(E)} \cdot \frac{n!}{n_1! \cdots n_k!} \left(\frac{\mu(B_1)}{\mu(E)}\right)^{n_1} \cdots \left(\frac{\mu(B_k)}{\mu(E)}\right)^{n_k} \\ &= \prod_{i=1}^{k} \frac{\mu(B_i)^{n_i}}{n_i!} e^{-\mu(B_i)}.\end{aligned}$$

By summing over all n_k, we get

$$\mathbb{P}\left[M(B_1) = n_1, \dots, M(B_{k-1}) = n_{k-1}\right] = \prod_{i=1}^{k-1} \frac{\mu(B_i)^{n_i}}{n_i!} e^{-\mu(B_i)}.$$

Hence $M(B_1), \dots, M(B_{k-1})$ are independent $\mathrm{Pois}(\mu(B_i))$ random variables. □

7.6 Superposition

Lemma 7.11. *Let* $\lambda = \sum_{i=1}^{\infty} \lambda_i$, $\lambda_i \geq 0$. $(X_i)_{i>0}$ *independent random variables with* $X_i \sim \mathrm{Poiss}(\lambda_i)$ *for every* $i \geq 1$. *Then the sum* $X = \sum_{i=1}^{\infty} X_i$ *is a Poiss*(λ) *random variable.*

Proof. See Exercises. □

Theorem 7.12. *Let* $M_i, i \geq 1$ *be a sequence of independent ppp*(μ_i) *where* μ_i *and* $\mu = \sum_{i=1}^{\infty} \mu_i$ *are sigma-finite measures. Then* $M = \sum_{i=1}^{\infty} M_i$ *is a ppp*(μ).

Proof. We first check that M is a point process. For every $B \subset E$ measurable, $M(B) = \sum_i M_i(B)$ is a well defined random variable (as a sum of nonnegative random variables). M is a measure almost surely (as a sum of of measures). Let $(E_n)_{n \in \mathbb{N}}$ be a partition of E such that $\mu(E_n) < \infty$ for every i. For all n,

$$\mathbb{E}[M(E_n)] = \sum_{i=1}^{\infty} \mathbb{E}[M_i(E_n)] = \sum_{i=1}^{\infty} \mu_i(E_n) = \mu(E_n) < \infty.$$

Hence $M(E_n) < \infty$ a.s. for every $n \in \mathbb{N}$, which implies that M is a σ-finite measure almost surely.

For $B \subset E$ measurable,

$$M(B) = \sum_i M_i(B) \overset{(d)}{=} \sum_i \text{Pois}(\mu_i(B_n)).$$

By the lemma, $M(B)$ is a Pois($\mu(B)$) random variable. Finally for $B_1, \ldots, B_k \subset E$ measurable and disjoint $(M_i(B_j))_{i\in\mathbb{N},1\leq j\leq k}$ are independent random variables. Therefore

$$M(B_i) = \sum_i M_i(B_1), \ldots, M(B_k) = \sum_i M_i(B_k)$$

are independent by grouping. □

Theorem 7.13. *Assume that μ is a sigma-finite measure on $(E, \mathcal{E})$, then there exists a ppp(μ) on E.*

Proof. $\mu = \sum_{i=1}^{\infty} \mu_i$ where $\mu_i(E) < \infty$. Let (M_i) be independent Poisson processes, where M_i is a ppp(μ_i). By superposition, $M = \sum_{i=1}^{\infty} M_i$ is a ppp(μ). □

7.7 Law of the Poisson process

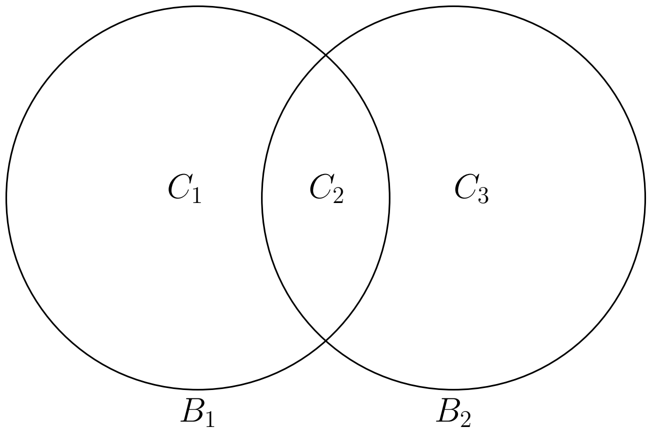
Let M be a ppp(μ) on E, its law P_M is a probability measure on $\mathcal{M}$.

Proposition 7.14. *Let M, M' be two ppp(μ) on $(E, \mathcal{E})$ then $P_M = P_{M'}$.*

Remark 7.15. $P_M = P_{M'}$ if and only if for all $A \subset \mathcal{M}$ measurable $P_M(A) = P_{M'}(A)$ if and only if for all $A \subset \mathcal{M}$ measurable $\mathbb{P}[M \in A] = \mathbb{P}[M' \in A]$.

Proof. Let $B_1, B_2 \subset E$ measurable, $n_1, n_2 \geq 0$. Define $C_1 = B_1 \setminus B_2$, $C_2 = B_1 \cap B_2$, and $C_3 = B_2 \setminus B_1$.

$$\begin{aligned}
\mathbb{P}[M(B_1) = n_1, M(B_2) = n_2] &= \sum_{\substack{m_1+m_2=n_1 \\ m_2+m_3=n_2}} \mathbb{P}[M(C_1) = n_1, M(C_2) = m_2, M(C_3) = m_3] \\
&= \sum_{\substack{m_1+m_2=n_1 \\ m_2+m_3=n_2}} \mathbb{P}[M'(C_1) = m_1, M'(C_2) = m_2, M'(C_3) = m_3] \\
&= \mathbb{P}[M'(B_1) = n_1, M'(B_2) = n_2]
\end{aligned}$$



Where the second equality holds as the C_i are disjoint. Equivalently, for all $B_1, \dots, B_k \subset E$ measurable

$$\mathbb{P}\left[M(B_1) = n_1, \dots, M(B_k) = n_k\right] = \mathbb{P}\left[M'(B_1) = n_1, \dots, M'(B_k) = n_k\right].$$

Therefore $P_M(A) \overset{(*)}{=} P_{M'}(A)$ for every set of the form $A = \{\eta : \ (\eta(B_1), \dots, \eta(B_k)) \in K\}$ for $B_1, \dots, B_k \subset E$ measurable and $K \subset \mathbb{N}^k$. Such sets for a π-system and generate $\mathcal{B}(\mathcal{M})$. Hence, by Dynkin's lemma, $(*)$ holds for every measurable set $A \subset \mathcal{M}$ measurable. $\square$

7.8 Restriction

Notation If ν is a measure on E, $C \subset E$ measurable, then we write $\nu_C := \nu(\cdot \cap C)$ (the measure restricted to C).

Theorem 7.16 (Restriction)**.** *Let $C_1, C_2, \dots \subset E$ measurable and disjoint. If N is a ppp(μ) on E, then $N_{C_1}, N_{C_2} \dots$ are independent ppp with respective intensities $\mu_{C_1}, \mu_{C_2}, \dots$*

Proof. Let $C_0 = E \setminus (\cup_{i \geq 1} C_i)$ (possibly empty). This, way we have a partition $E = \bigcup_{i \geq 0} C_i$ Let $N'_0, N'_1, \dots$ independent ppp with respective intensities $\mu_{C_0}, \mu_{C_1}, \dots$. By superposition $N' = \sum_{i \geq 0} N'_i$ is a ppp(μ) (indeed, $\mu = \sum_{i \geq 0} \mu_{C_i}$).

For every $B \subset E$ measurable and $j \geq 0$

$$\begin{aligned} N'(B \cap C_j) &= \sum_{i > 0} \underbrace{N'_i(B \cap C_j)}_{=0 \text{ a.s. if } i \neq j} \\ &= N'_j(B) \text{ a.s.} \end{aligned}$$

Hence $N'_{C_j} = N'_j$ a.s. Let $f_1, \dots, f_k : \mathcal{M} \to \mathbb{R}_+$ measurable.

$$\mathbb{E}\left[\prod_{i=1}^{k} f_i(N_{C_i})\right] \overset{\text{(uniqueness)}}{=} \mathbb{E}\left[\prod_{i=1}^{k} f_i(N'_{C_i})\right] = \mathbb{E}\left[\prod_{i=1}^{k} f_i(N'_i)\right] = \prod_{i=1}^{k} \mathbb{E}\left[f_i(N'_i)\right].$$

Hence $N_{C_1}, \dots, N_{C_k}$ are independent ppp(μ_{C_i}). □

7.9 Mapping

Let $(F, \mathcal{F})$ be Polish space equipped with its Borel σ-algebra. We consider a measurable map

$$T : E \to F.$$

Given a measure ν on E, we write $T\#\nu$ for the pushforward measure of ν under T (defined by $T\#\nu(B) = \nu(T^{-1}(B))$ for every $B \in \mathcal{E}$).

Theorem 7.17. *Assume that $T\#\mu$ is sigma-finite. Let M be a ppp(μ) on E. The process*

$$T\#M = (M(T^{-1}(B))_{B\in\mathcal{F}}$$

is a ppp($T\#\mu$) on F.

Proof. We first show that $T\#M$ is a point process on F. For every fixed $B \in \mathcal{F}$, we have $T^{-1}(B) \in \mathcal{E}$ (because T is measurable). Therefore, $T\#M(B) = M(T^{-1}(B))$ is a well defined random variable. Let $\mathcal{M}'$ be the space of sigma-finite measures on $(F, \mathcal{F})$ taking values in $\mathbb{N} \cup \{\infty\}$. Notice that $\eta \in \mathcal{M} \implies T\#\eta \in \mathcal{M}'$. Since $M \in \mathcal{M}$ almost surely, we also have $T\#M \in \mathcal{M}'$ almost surely.

Let $B \in \mathcal{F}$. By definition, we have

$$T\#M(B) = M\left(T^{-1}(B)\right) \sim \text{Poisson}(\mu(T^{-1}(B)) = \text{Poisson}(T\#\mu(B)).$$

Let $B_1, \dots, B_k$ be disjoint sets in $\mathcal{F}$. Then, their pre-images $T^{-1}(B_1), \dots, T^{-1}(B_k)$ are disjoint measurable sets in $\mathcal{E}$. The independence of the random variables

$$T\#M(B_1) = M(T^{-1}(B_1)), \dots, T\#M(B_k) = M(T^{-1}(B_k))$$

arises from the fact that M is a Poisson point process. As before, we have that $T\#M(B_1) = M(T^{-1}(B_1)) \sim \text{Poisson}(\mu(T^{-1}(B_1)) = \text{Poisson}(T\#\mu(B_1))$, and the statement follows. □

Remark 7.18. If we decompose $M = \sum_{i=1}^{\tau} \delta_{X_i}$ (as in Theorem 7.8), then $T\#M$ can be written as $T\#M = \sum_{i=1}^{\tau} \delta_{T(X_i)}$. Namely if the process M correspond to the point $X_1, X_2, \dots$ then the process $T\#M$ corresponds to the image of these points $T(X_1), T(X_2) \dots$

Example 7.1. $E = \mathbb{R},\ F = \mathbb{Z},\ T : E \to F; x \to \lfloor x \rfloor,\ \mu = \mathcal{L},\ T\#\mu = |\cdot|.$

7.10 Marking

Motivation Cars on a highway, at time 0 the position of the cars is a ppp(1) on $\mathbb{R}$ (that means on average 1 car per kilometer of highway). We put an observer (Olga) at 0 on $\mathbb{R}$.

Case 1: All of the cars have speed 50km/h, we want to study X = number of cars seen by Olga in 1 hour. What is the law of X? $X \sim \text{Pois}(50)$.

Case 2: The cars have a random speed $\sim \mathcal{U}([50, 100])$. What is the law of X? It may at first seem complicated, but it is not!

Framework Let $(F, \mathcal{F}, \nu)$ Polish, probability space ('space of marks').

Definition 7.3. Let $M = \sum_{i=1}^{\tau} \delta_{X_i}$ be a ppp(μ) on E. $(Y_i)_{i>0}$ i.i.d. random variable with law ν independent of M. The Y*-marked point process associated to* M is the point process on $E \times F$ defined by

$$\boxed{\overline{M} = \sum_{i=1}^{\tau} \delta_{(X_i, Y_i)}.}$$

Remark 7.19. X_i corresponds to the position of the cars in Case 2, and Y_i to their speeds.

Theorem 7.20. *The marked process* $\overline{M}$ *is a ppp*$(\mu \otimes \nu)$.

Proof. See Section 7.13. □

7.11 Thinning

Theorem 7.21. *Let* $p \in [0, 1]$. *Let* $M = \sum_{i=1}^{\tau} \delta_{X_i}$ *be a ppp*(μ) *on* E. *Let* $(Z_i)_{i\geq 1}$ *be an infinite sequence of iid Bernoulli random variables with parameter* p. *The two point processes*

$$M_0 = \sum_{\substack{i\geq 1 \\ Z_i=0}} \delta_{X_i} \quad \text{and} \quad M_1 = \sum_{\substack{i\geq 1 \\ X_i=1}} \delta_{X_i}$$

are two independent ppp with intensities $(1 - p)\mu$ *and* $p\mu$ *respectively.*

Proof. The point process on $E \times \{0,1\}$ defined by

$$\overline{M} = \sum_{i \geq 1} \delta_{(X_i, Z_i)}.$$

is Ber(p)-marking of M. Hence by Theorem 7.20, $\overline{M}$ is a ppp($\mu \otimes \text{Ber}(p)$) on $E \times \{0,1\}$. By restriction, the two processes $\overline{M}|_{E\times\{0\}}$ and $\overline{M}|_{E\times\{1\}}$, are independent processes with intensities $(\mu \otimes \text{Ber}(p))|_{E\times\{0\}}$ and $(\mu \otimes \text{Ber}(p))|_{E\times\{1\}}$ respectively. This concludes the proof since M_j is the projection o f $\overline{M}|_{E\times\{j\}}$ on the coordinate j. □

7.12 Laplace Functional

Lemma 7.22. *Let X be a Pois(λ) random variable, for $\lambda > 0$, then for all $u \geq 0$*

$$\mathbb{E}\left[e^{-uX}\right] = \exp(-\lambda(1 - e^{-u})).$$

Proof. For every $u \geq 0$ we have

$$\mathbb{E}\left[e^{-uX}\right] = \sum_k \frac{\lambda^k}{k!} e^{-\lambda} e^{-ku} = e^{-\lambda} \exp(\lambda e^{-u}).$$

□

Definition 7.4. Let M be a point process on $(E, \mathcal{E})$, for every $u : E \to \mathbb{R}_+$ measurable define

$$\boxed{L_M(u) = \mathbb{E}\left[\exp(-\int u(x) M(dx)\right].}$$

Remark 7.23. $L_M(u)$ is well defined. Indeed $\int_E u(x)M(dx) = \int_E u dN$ is a well defined random variable.

We can interpret $\int u(x)M(dx)$ as $\sum_{x \text{ 'points of N'}} u(x)$ with multiplicities counted.

Theorem 7.24 (Characterization via Laplace Functional)**.** *Let μ be a sigma-finite measure on $(E, \mathcal{E})$. Let M be a point process on E. The following are equivalent*

(i) M is a ppp(μ),

(ii) For all $u : E \to \mathbb{R}_+$ measurable

$$L_M(u) = \exp\left(-\int_E 1 - e^{-u(x)} \mu(dx)\right).$$

Proof. $\boxed{\Longrightarrow}$ Let $u = \sum_{i=1}^k u_i \mathbb{1}_{B_i}$ for $B_1, \dots, B_k$ disjoint, $u_i \geq 0$.

$$\begin{aligned} L_M(u) &= \mathbb{E}\left[\exp\left(-\sum_{i=1}^k u_i M(B_i)\right)\right] \overset{\text{(indep.)}}{=} \prod_{i=1}^k \mathbb{E}\left[e^{u_i M(B_i)}\right] \\ &= \prod_{i=1}^k \exp\left(-\mu(B_i)(1-e^{-u_i})\right) = \exp\left(-\int_E 1 - e^{-u(x)}\mu(dx)\right). \end{aligned}$$

For general $u \geq 0$, consider (u_n) of the form above such that $u_n \uparrow u$. For every n

$$\underbrace{L_M(u_n)}_{\overset{\text{(MCT)}}{\to} L_M(u)} = \underbrace{\exp\left(-\int_E (1-e^{-u_n(x)})\mu(dx)\right)}_{\to \exp\left(-\int_E (1-e^{-u(x)})\mu(dx)\right)}.$$

$\boxed{\Longleftarrow}$ Let $B_1, \dots, B_k$ be disjoint. For all $x = (x_1, \dots, x_k)$ with $x_i \geq 0$. By applying (ii) to $u = \sum_{i=1}^k x_i \mathbb{1}_{B_i}$, we have

$$\begin{aligned} \mathbb{E}\left[e^{-x\cdot(M(B_1),\dots,M(B_k))}\right] &= L_N(u) \\ &= \exp\left(-\int_E 1 - e^{-u(x)}\mu(dx)\right) \\ &= \prod_{i=1}^k \exp\left(-\mu(B_i)(1-e^{-x_i}\right) = \mathbb{E}\left[e^{-x\cdot Y}\right], \end{aligned}$$

where $Y = (Y_1, \dots, Y_k)$ is a random vector of independent variables. Furthermore Y_i are $\text{Pois}(\mu(B_i))$ random variables, since the Laplace transform characterizes the law we have

$$(M(B_1), \dots, M(B_k)) \overset{\text{(law)}}{=} Y.$$

□

7.13 Proof of the marking Theorem

First we show that $\overline{M}$ is a point process. For every $B \subset E$ measurable,

$$\overline{M}(B) = \sum_{i=1}^\tau \underbrace{\mathbb{1}_{(X_i,Y_i)\in B}}_{\text{measurable}}.$$

Let $u : E \times F \to \mathbb{R}_+$ measurable

$$L_{\overline{M}}(u) = \sum_{m \in \mathbb{N} \cup \{\infty\}} \underbrace{\mathbb{E}\left[\mathbb{1}_{\tau=m} \exp\left(-\sum_{k=1}^{m} u(X_k, Y_k)\right)\right]}_{f(m)}.$$

For $m < \infty$, we have

$$\begin{aligned} f(m) &= \int_F \dots \int_F \mathbb{E}\left[\mathbb{1}_{\tau=m} \exp\left(-\sum_{k=1}^{m} u(X_k, y_k)\right)\right] \nu(dy_1) \dots \nu(dy_k) \\ &= \mathbb{E}\left[\mathbb{1}_{\tau=m} \prod_{k=1}^{m} \underbrace{\left(\int_F e^{-u(X_k, y_k)}\right)}_{e^{-v(X_k)}}\right] \end{aligned}$$

where $v(x) = -\log\left(\int_F e^{-u(x,y)}\nu(dy)\right) \geq 0$. Hence for all $m < \infty$, we have

$$f(m) = \mathbb{E}\left[\mathbb{1}_{\tau=m} \exp\left(-\sum_{k=1}^{m} v(x_k)\right)\right].$$

Equivalently and using monotone convergence, the equality above also holds for $m = \infty$. Therefore

$$\begin{aligned} L_{\overline{M}}(u) &= \sum_{m \in \mathbb{N} \cup \{\infty\}} \mathbb{E}\left[\mathbb{1}_{\tau=m} \exp\left(-\sum_{k=1}^{m} v(X_k)\right)\right] = \mathbb{E}\left[\exp\left(-\sum_{k=1}^{\tau} v(X_k)\right)\right] \\ &= L_M(v) = \exp\left(-\int_E 1 - e^{-v(x)}\mu(dx)\right) \\ &= \exp\left(-\int_E \left[1 - \int_F e^{-u(x,y)}\nu(dy)\right]\mu(dx)\right) \\ &= \exp\left(-\int_{E \times F} 1 - e^{-u(x,y)}\nu(dy)\mu(dx)\right). \end{aligned}$$

Hence $\overline{M}$ is a ppp$(\mu \otimes \nu)$.

Chapter 8

Appendix

Lemma 8.1. *Let $A \subset \mathbb{N} \setminus \{0\}$ be stable under addition (i.e. $x, y \in A \implies x + y \in A$). Then*

$$\gcd(A) = 1 \iff \exists n_0 \in \mathbb{N} : \{n \in \mathbb{N} : n \geq n_0\} \subset A.$$

Proof. $\Longleftarrow$: Follows from the fact that $\gcd(n_0, n_0 + 1) = 1$.

$\implies$: Assume $\gcd(A) = 1$. Let $a \in A$ be arbitrary and $a = \prod_{i=1}^{k} p_i^{alpha_i}$ be its prime factorization. Since $\gcd(A) = 1$, one can find $b_1, \ldots, b_k \in A$ such that for all i $p_i \nmid b_i$. This implies

$$\gcd(a, b_1, \ldots, b_k) = 1.$$

Write $d = \gcd(b_1, \ldots, b_k)$. By Bezout's Theorem, we can pick $u_1, \ldots, u_k \in \mathbb{Z}$ such that

$$u_1 b_1 + \ldots + u_k b_k = d.$$

Now, choose an integer λ large enough such that $u_i + \lambda a \geq 0$ for every i and define

$$b = (u_1 + \lambda a) b_1 + \ldots + (u_k + \lambda a) b_k = d + \lambda(b_1 + \ldots + b_k) a.$$

The first expression shows that $b \in A$, and the second implies that $\gcd(a, b) = gcd(a, d) = 1$. To summarize, we found $a, b \in A$ such that $\gcd(a, b) = 1$.

Without loss of generality, we may assume $a < b$. Since $\gcd(a, b) = 1$, the set $B = \{b, 2b, \ldots, ab\}$ covers all of the residue classes modulo a. Since $a < b$, this implies that $B + \{ka, k \in \mathbb{N}\}$ includes every number $z \geq ab$. This concludes the proof by choosing $n_0 = ab$. □

Bibliography

[1] S. r. Asmussen. *Applied probability and queues*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons, Ltd., Chichester, 1987.

[2] O. Häggström. *Finite Markov chains and algorithmic applications*, volume 52 of *London Mathematical Society Student Texts*. Cambridge University Press, Cambridge, 2002.