Chapter 24
The Poisson Point Process

Poisson point processes can be used as a cornerstone in the construction of very different stochastic objects such as, for example, infinitely divisible distributions, Markov processes with complex dynamics, objects of stochastic geometry and so forth.

In this chapter, we briefly develop the general framework of random measures and construct the Poisson point process and characterize it in terms of its Laplace transform. As an application we construct a certain subordinator and show that the Poisson point process is the invariant measure of systems of independent random walks. Via the connection with subordinators, in the third section, we construct two distributions that play prominent roles in population genetics: the Poisson–Dirichlet distribution and the GEM distribution.

For a nice exposition including many examples, see also [99].

24.1 Random Measures

In the following, let E be a locally compact Polish space (for example, $E = \mathbb{R}^d$ or $E = \mathbb{Z}^d$) with Borel σ-algebra $\mathcal{B}(E)$. Let

$$\mathcal{B}_b(E) = \big\{B \in \mathcal{B}(E) : B \text{ is relatively compact}\big\}$$

be the system of *bounded* Borel sets and $\mathcal{M}(E)$ the space of Radon measures on E (see Definition 13.3).

Definition 24.1 Denote by $\mathbb{M} = \sigma(I_A : A \in \mathcal{B}_b(E))$ the smallest σ-algebra on $\mathcal{M}(E)$ with respect to which all maps

$$I_A : \mu \mapsto \mu(A), \qquad A \in \mathcal{B}_b(E),$$

are measurable.

A. Klenke, *Probability Theory*, Universitext, DOI 10.1007/978-1-4471-5361-0_24,

Denote by $\mathcal{B}^+(E)$ the set of measurable maps $E \to [0, \infty]$ and by $\mathcal{B}_b^{\mathbb{R}}(E)$ the set of bounded measurable maps $E \to \mathbb{R}$ with compact support. For every $f \in \mathcal{B}^+(E)$, the integral $I_f(\mu) := \int f \, d\mu$ is well-defined and for every $f \in \mathcal{B}_b^{\mathbb{R}}(E)$, $I_f(\mu)$ is well-defined and finite.

Theorem 24.2 *Let τ_v be the vague topology on $\mathcal{M}(E)$. Then*

$$\mathbb{M} = \mathcal{B}(\tau_v) = \sigma\big(I_f : f \in C_c(E)\big) = \sigma\big(I_f : f \in C_c^+(E)\big).$$

Proof This is left as an exercise. (See [82, Lemma 4.1].) $\square$

Let $\widetilde{\mathcal{M}}(E)$ be the space of *all* measures on E endowed with the σ-algebra

$$\widetilde{\mathbb{M}} = \sigma\big(I_A : A \in \mathcal{B}_b(E)\big).$$

Choose a countable dense set $F \subset E$, and for any $x \in F$ choose a compact neighborhood K_x. Then we get (compare Exercise 13.1.8)

$$\mathcal{M}(E) = \bigcap_{x \in F} \big\{\mu \in \widetilde{\mathcal{M}}(E) : \mu(K_x) < \infty\big\}.$$

Hence $\mathcal{M}(E) \in \widetilde{\mathbb{M}}$. Clearly, $\mathbb{M} = \widetilde{\mathbb{M}}\big|_{\mathcal{M}(E)}$ is the trace σ-algebra of $\widetilde{\mathbb{M}}$ on $\mathcal{M}(E)$. Here we need the slightly larger space in order to define random measures in such a way that all almost surely well-defined operations on random measures again yield random measures.

Definition 24.3 A *random measure* on E is a random variable X on some probability space $(\Omega, \mathcal{A}, \mathbf{P})$ with values in $(\widetilde{\mathcal{M}}(E), \widetilde{\mathbb{M}})$ and with $\mathbf{P}[X \in \mathcal{M}(E)] = 1$.

Theorem 24.4 *Let X be a random measure on E. Then the set function $\mathbf{E}[X] : \mathcal{B}(E) \to [0, \infty]$, $A \mapsto \mathbf{E}[X(A)]$ is a measure. We call $\mathbf{E}[X]$ the* intensity measure *of X. We say that X is integrable if $\mathbf{E}[X] \in \mathcal{M}(E)$.*

Proof Clearly, $\mathbf{E}[X]$ is finitely additive. Let $A, A_1, A_2, \ldots \in \mathcal{B}(E)$ with $A_n \uparrow A$. Consider the random variables $Y_n := X(A_n)$ and $Y = X(A)$. Then $Y_n \uparrow Y$ and hence, by monotone convergence, $\mathbf{E}[X](A_n) = \mathbf{E}[Y_n] \overset{n\to\infty}{\longrightarrow} \mathbf{E}[Y] = \mathbf{E}[X](A)$. Hence $\mathbf{E}[X]$ is lower semicontinuous and is thus a measure (by Theorem 1.36). $\square$

Theorem 24.5 *Let $\mathbf{P}_X$ be the distribution of a random measure X. Then $\mathbf{P}_X$ is uniquely determined by the distributions of either of the families*

$$\big((I_{f_1}, \ldots, I_{f_n}) : n \in \mathbb{N};\, f_1, \ldots, f_n \in C_c^+(E)\big) \tag{24.1}$$

or

$$\big((I_{A_1}, \ldots, I_{A_n}) : n \in \mathbb{N};\, A_1, \ldots, A_n \in \mathcal{B}_b(E) \textit{ pairwise disjoint}\big). \tag{24.2}$$

Proof The class of sets

$$\mathcal{I} = \big\{(I_{f_1}, \ldots, I_{f_n})^{-1}(A) : n \in \mathbb{N};\ f_1, \ldots, f_n \in C_c^+(E),\ A \in \mathcal{B}\big([0,\infty)^n\big)\big\}$$

is a π-system and by Theorem 24.2 it generates $\mathbb{M}$. Hence the measure $\mathbf{P}_X$ is characterized by its values on $\mathcal{I}$.

Similarly, the claim follows for

$$\big((I_{A_1}, \ldots, I_{A_n}) : n \in \mathbb{N};\ A_1, \ldots, A_n \in \mathcal{B}_b(E)\big).$$

If $A_1, \ldots, A_n \in \mathcal{B}_b(E)$ are arbitrary, then there exist $2^n - 1$ pairwise disjoint sets $B_1, \ldots, B_{2^n-1}$ with $A_i = \bigcup_{k: B_k \subset A_i} B_k$ for all $i = 1, \ldots, n$. The distribution of $(I_{A_1}, \ldots, I_{A_n})$ can be computed from that of $(I_{B_1}, \ldots, I_{B_{2^n-1}})$. □

In the following, let $i = \sqrt{-1}$ be the imaginary unit.

Definition 24.6 Let X be a random measure on E. Denote by

$$\mathcal{L}_X(f) = \mathbf{E}\left[\exp\left(-\int f\,dX\right)\right], \quad f \in \mathcal{B}^+(E),$$

the *Laplace transform* of X and by

$$\varphi_X(f) = \mathbf{E}\left[\exp\left(i\int f\,dX\right)\right], \quad f \in \mathcal{B}_b^{\mathbb{R}}(E),$$

the *characteristic function* of X.

Theorem 24.7 *The distribution $\mathbf{P}_X$ of a random measure X is characterized by its Laplace transform $\mathcal{L}_X(f)$, $f \in C_c^+(E)$, as well as by its characteristic function $\varphi_X(f)$, $f \in C_c(E)$.*

Proof This is a consequence of Theorem 24.5 and the uniqueness theorem for characteristic functions (Theorem 15.8) and for Laplace transforms (Exercise 15.1.2) of random variables on $[0,\infty)^n$. □

Definition 24.8 We say that a random measure X on E has *independent increments* if, for any choice of finitely many pairwise disjoint measurable sets $A_1, \ldots, A_n$, the random variables $X(A_1), \ldots, X(A_n)$ are independent.

Corollary 24.9 *The distribution of a random measure X on E with independent increments is uniquely determined by the family $(\mathbf{P}_{X(A)}, A \in \mathcal{B}_b(E))$.*

Proof This is an immediate consequence of Theorem 24.5. □

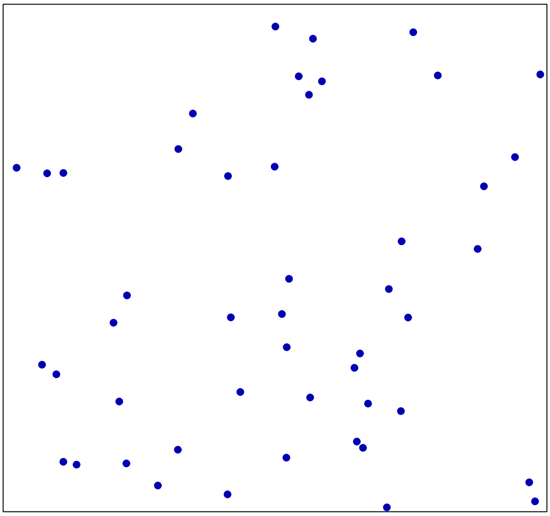

Fig. 24.1 Poisson point process on the unit square with intensity measure 50λ

Definition 24.10 Let $\mu \in \mathcal{M}(E)$. A random measure X with independent increments is called a *Poisson point process (PPP)* with intensity measure μ if, for any $A \in \mathcal{B}_b(E)$, we have $\mathbf{P}_{X(A)} = \mathrm{Poi}_{\mu(A)}$. In this case, we write $\mathrm{PPP}_\mu := \mathbf{P}_X \in \mathcal{M}_1(\mathcal{M}(E))$ and say that X is a PPP_μ.

See Fig. 24.1 for a simulation of a Poisson point process on the unit square.

Remark 24.11 The definition of the PPP (and its construction in the following theorem) still works if $(E, \mathcal{E}, \mu)$ is only assumed to be a σ-finite measure space. However, the characterization in terms of Laplace transforms is a bit simpler in the case of locally compact Polish spaces considered here. $\Diamond$

Theorem 24.12 *For every $\mu \in \mathcal{M}(E)$, there exists a Poisson point process X with intensity measure μ.*

Proof μ is σ-finite since $\mu \in \mathcal{M}(E)$. Hence there exist $E_n \uparrow E$ with $\mu(E_n) < \infty$ for every $n \in \mathbb{N}$. Define $\mu_1 = \mu(E_1 \cap \cdot)$ and $\mu_n = \mu((E_n \setminus E_{n-1}) \cap \cdot)$ for $n \geq 2$. If $X_1, X_2, \ldots$ are independent Poisson point processes with intensity measures $\mu_1, \mu_2, \ldots$, then $X = \sum_{n=1}^{\infty} X_n$ has intensity measure $\mathbf{E}[X] = \mu$ and hence X is a random measure (see Exercise 24.1.1). Furthermore, it is easy to see that X has independent increments and that

$$\mathbf{P}_{X(A)} = \mathbf{P}_{X_1(A)} * \mathbf{P}_{X_2(A)} * \ldots = \mathrm{Poi}_{\mu_1(A)} * \mathrm{Poi}_{\mu_2(A)} * \ldots = \mathrm{Poi}_{\mu(A)}.$$

Hence we have $X \sim \mathrm{PPP}_\mu$.

Therefore, it is enough to consider the case $\mu(E) \in (0, \infty)$. Define

$$\nu = \frac{\mu(\cdot)}{\mu(E)} \in \mathcal{M}_1(E).$$

Let $N, Y_1, Y_2, \dots$ be independent random variables with $N \sim \mathrm{Poi}_{\mu(E)}$ and $\mathbf{P}_{Y_i} = \nu$ for all $i \in \mathbb{N}$. Define

$$X(A) = \sum_{n=1}^{N} \mathbb{1}_A(Y_n) \quad \text{for } A \in \mathcal{B}(E).$$

The random variables $\mathbb{1}_A(Y_1), \mathbb{1}_A(Y_2), \dots$ are independent and $\mathrm{Ber}_{\nu(A)}$-distributed; hence we have $X(A) \sim \mathrm{Poi}_{\mu(A)}$ (see Theorem 15.14(iii)). Let $n \in \mathbb{N}$ and let $A_1, \dots, A_n \in \mathcal{B}(E)$ be pairwise disjoint. Then

$$\psi(t) := \mathbf{E}\left[\exp\left(i \sum_{l=1}^{n} t_l \mathbb{1}_{A_l}(Y_1)\right)\right] = 1 + \sum_{l=1}^{n} \nu(A_l)\big(e^{it_l} - 1\big), \quad t \in \mathbb{R}^n,$$

is the characteristic function of $(\mathbb{1}_{A_1}(Y_1), \dots, \mathbb{1}_{A_n}(Y_1))$. Further, let φ be the characteristic function of $(X(A_1), \dots, X(A_n))$ and let φ_l be the characteristic function of $X(A_l)$ for $l = 1, \dots, n$. Hence

$$\varphi_l(t_l) = \exp\big(\mu(A_l)\big(e^{it_l} - 1\big)\big).$$

By Theorem 15.14(iii), we have

$$\begin{aligned}
\varphi(t) &= \mathbf{E}\left[\exp\left(i \sum_{l=1}^{n} t_l X(A_l)\right)\right] \\
&= \exp\big(\mu(E)\big(\psi(t) - 1\big)\big) \\
&= \exp\left(\sum_{l=1}^{n} \mu(A_l)\big(e^{it_l} - 1\big)\right) = \prod_{l=1}^{n} \varphi_l(t_l).
\end{aligned}$$

Thus $X(A_1), \dots, X(A_n)$ are independent. This implies $X \sim \mathrm{PPP}_\mu$. □

Exercise 24.1.1 Let $X_1, X_2, \dots$ be random measures and $\lambda_1, \lambda_2, \dots \in [0, \infty)$. Define $X := \sum_{n=1}^{\infty} \lambda_n X_n$. Show that X is a random measure if and only if we have $\mathbf{P}[X(B) < \infty] = 1$ for all $B \in \mathcal{B}_b(E)$. Infer that if X is a random variable with values in $(\widetilde{\mathcal{M}}(E), \widetilde{\mathbb{M}}(E))$ and $\mathbf{E}[X] \in \mathcal{M}(E)$, then X is a random measure.

Exercise 24.1.2 Let τ_w be the topology of weak convergence on $\mathcal{M}_1(E)$ and let $\sigma(\tau_w)$ be the Borel σ-algebra on $\mathcal{M}_1(E)$. Show that $\mathbb{M}\big|_{\mathcal{M}_1(E)} = \sigma(\tau_w)$.

24.2 Properties of the Poisson Point Process

Theorem 24.13 *Let $\mu \in \mathcal{M}(E)$ be atom-free; that is, $\mu(\{x\}) = 0$ for every $x \in E$. Let X be a random measure on E with $\mathbf{P}[X(A) \in \mathbb{N}_0 \cup \{\infty\}] = 1$ for every $A \in \mathcal{B}(E)$. Then the following are equivalent:*

(i) $X \sim \mathrm{PPP}_\mu$.
(ii) *X almost surely has no double points; that is,*

$$\mathbf{P}\big[X\big(\{x\}\big) \ge 2 \quad \textit{for some } x \in E\big] = 0,$$

and

$$\mathbf{P}\big[X(A) = 0\big] = e^{-\mu(A)} \quad \textit{for all } A \in \mathcal{B}_b(E). \tag{24.3}$$

Proof (i) $\Longrightarrow$(ii) This is obvious.
(ii) $\Longrightarrow$(i) If $A_1, \ldots, A_n \in \mathcal{B}_b(E)$ are pairwise disjoint, then

$$\begin{aligned}\mathbf{P}\big[X(A_1) = 0, \ldots, X(A_n) = 0\big] &= \mathbf{P}\big[X(A_1 \cup \ldots \cup A_n) = 0\big] \\ &= e^{-\mu(A_1 \cup \ldots \cup A_n)} \\ &= \prod_{l=1}^{n} e^{-\mu(A_l)} = \prod_{l=1}^{n} \mathbf{P}\big[X(A_l) = 0\big].\end{aligned}$$

Hence the random variables $\widetilde{X}(A) := X(A) \wedge 1$ are independent for disjoint sets A. The rest of the proof is similar to that of Theorem 5.34. Let $A \in \mathcal{B}_b(E)$. Choose an $A_0 \subset A$ with $\mu(A_0) = \mu(A)/2$ (this is possible by Exercise 8.3.1 since μ is atom-free) and define $A_1 = A \setminus A_0$. Similarly, choose $A_{i,0}, A_{i,1} \subset A_i$ for $i = 0, 1$ and inductively define disjoint sets $A_{i,0}, A_{i,1} \subset A_i$ for $i \in \{0,1\}^{n-1}$ with $\mu(A_i) = 2^{-n}\mu(A)$ for every $i \in \{0,1\}^n$. Define

$$N_n(A) := \sum_{i \in \{0,1\}^n} \widetilde{X}(A_i).$$

As X does not have double points, we have $N_n(A) \uparrow X(A)$ almost surely. On the other hand, by assumption, $N_n(A) \sim b_{2^n, 1-\exp(-2^{-n}\mu(A))}$ for $n \in \mathbb{N}$; hence the characteristic functions converge:

$$\begin{aligned}\varphi_{N_n(A)}(t) &= \Big(1 + \big(1 - e^{-2^{-n}\mu(A)}\big)\big(e^{it} - 1\big)\Big)^{2^n} \\ &\overset{n\to\infty}{\longrightarrow} \exp\big(\mu(A)\big(e^{it} - 1\big)\big) = \varphi_{\mathrm{Poi}_{\mu(A)}}(t).\end{aligned}$$

Therefore, we have $\mathbf{P}_{N_n(A)} \overset{n\to\infty}{\longrightarrow} \mathrm{Poi}_{\mu(A)}$ and thus $X(A) \sim \mathrm{Poi}_{\mu(A)}$.

If $A_1, \ldots, A_k \in \mathcal{B}_b(E)$ are pairwise disjoint, then the sets $N_n(A_1), \ldots, N_n(A_k)$ (constructed in a way similar to that above) are independent. Hence also the limits $X(A_l) = \lim_{n\to\infty} N_n(A_l)$, $l = 1, \ldots, k$ are independent. □

Theorem 24.14 *Let $\mu \in \mathcal{M}(E)$ and let X be a Poisson point process with intensity measure μ. Then X has Laplace transform*

$$\mathcal{L}_X(f) = \exp\left(\int \mu(dx)\big(e^{-f(x)} - 1\big)\right), \quad f \in \mathcal{B}^+(E),$$

and characteristic function

$$\varphi_X(f) = \exp\left(\int \mu(dx)\big(e^{if(x)} - 1\big)\right), \quad f \in \mathcal{B}_b^{\mathbb{R}}(E).$$

Proof It is enough to show the claim for simple functions $f = \sum_{l=1}^n \alpha_l \mathbb{1}_{A_l}$ with complex numbers $\alpha_1, \ldots, \alpha_n$ and with pairwise disjoint sets $A_1, \ldots, A_n \in \mathcal{B}_b(E)$. (For general f, the claim follows by the usual approximation arguments.) For such f, however,

$$\begin{aligned}
\mathbf{E}\big[\exp\big(-I_f(X)\big)\big] &= \mathbf{E}\left[\prod_{l=1}^n e^{-\alpha_l X(A_l)}\right] = \prod_{l=1}^n \mathbf{E}\big[e^{-\alpha_l X(A_l)}\big] \\
&= \prod_{l=1}^n \exp\big(\mu(A_l)\big(e^{-\alpha_l} - 1\big)\big) \\
&= \exp\left(\sum_{l=1}^n \mu(A_l)\big(e^{-\alpha_l} - 1\big)\right) \\
&= \exp\left(\int \mu(dx)\big(e^{-f(x)} - 1\big)\right).
\end{aligned}$$

□

Corollary 24.15 (Moments of the PPP) *Let $\mu \in \mathcal{M}(E)$ and $X \sim \mathrm{PPP}_\mu$.*

(i) *If $f \in \mathcal{L}^1(\mu)$, then $\mathbf{E}[\int f\, dX] = \int f\, d\mu$.*
(ii) *If $f \in \mathcal{L}^2(\mu) \cap \mathcal{L}^1(\mu)$, then $\mathbf{Var}[\int f\, dX] = \int f^2\, d\mu$.*

Recall that only for finite μ, we have the inclusion $\mathcal{L}^2(\mu) \subset \mathcal{L}^1(\mu)$.

Proof If $f = f^+ - f^- \in \mathcal{L}^1(\mu)$, then for the characteristic function, integral and differentiation interchange, $\frac{d}{dt}\varphi_X(tf^+) = i\varphi_X(tf^+) \int f(x) e^{itf^+(x)} \mu(dx)$ and hence (by Exercise 15.4.4(iii))

$$\mathbf{E}\big[I_{f^+}(X)\big] = \frac{1}{i}\frac{d}{dt}\varphi_X\big(tf^+\big)\Big|_{t=0} = \int f^+\, d\mu.$$

Arguing similarly with f^- and adding up, we get (i).

If $f \in \mathcal{L}^1(\mu) \cap \mathcal{L}^2(\mu)$, then the argument can be iterated (using Theorem 15.34)

$$\frac{d^2}{dt^2}\varphi_X(tf) = -\varphi_X(tf)\left[\int f^2(x)\, e^{itf(x)} \mu(dx) + \left(\int f(x)\, e^{itf(x)} \mu(dx)\right)^2\right],$$

hence we have $\mathbf{E}[I_f(X)^2] = -\frac{d^2}{dt^2}\varphi_X(tf)\big|_{t=0} = I_{f^2}(\mu) + I_f(\mu)^2$. □

Theorem 24.16 (Mapping theorem) *Let E and F be locally compact Polish spaces and let $\phi : E \to F$ be a measurable map. Let $\mu \in \mathcal{M}(E)$ with $\mu \circ \phi^{-1} \in \mathcal{M}(F)$ and let X be a PPP on E with intensity measure μ. Then $X \circ \phi^{-1}$ is a PPP on F with intensity measure $\mu \circ \phi^{-1}$.*

Proof For $f \in \mathcal{B}^+(F)$,

$$\mathcal{L}_{X\circ\phi^{-1}}(f) = \mathcal{L}_X(f \circ \phi) = \exp\left(\int \left(e^{-f(\phi(x))} - 1\right)\mu(dx)\right)$$
$$= \exp\left(\int \left(e^{-f(y)} - 1\right)\left(\mu \circ \phi^{-1}\right)(dy)\right).$$

Now, Theorem 24.14 and Theorem 24.7 yield the claim. □

Theorem 24.17 *Let $\nu \in \mathcal{M}((0,\infty))$ and let $X \sim \mathrm{PPP}_\nu$ on $(0,\infty)$. Further, define $Y := \int x X(dx)$. Then the following are equivalent.*

(i) $\mathbf{P}[Y < \infty] > 0$.
(ii) $\mathbf{P}[Y < \infty] = 1$.
(iii) $\int \nu(dx)(1 \wedge x) < \infty$.

If (i)–(iii) *hold, then Y is an infinitely divisible nonnegative random variable with Lévy measure ν.*

Proof Let $Y_\infty = \int_{[1,\infty)} x X(dx)$ and $Y_t := \int_{(t,1)} x X(dx)$ for $t \in [0,1)$. Evidently, $Y = Y_0 + Y_\infty$. Furthermore, it is clear that

$$\mathbf{P}[Y_\infty < \infty] > 0 \quad \Longleftrightarrow \quad \mathbf{P}[Y_\infty < \infty] = 1 \quad \Longleftrightarrow \quad \nu\big([1,\infty)\big) < \infty. \tag{24.4}$$

If (iii) holds, then $\mathbf{E}[Y_0] = \int_{(0,1)} x\,\nu(dx) < \infty$; hence $Y_0 < \infty$ a.s. (and thus $Y < \infty$ a.s. by (24.4)). On the other hand, if (iii) does not hold, then $Y_\infty = \infty$ a.s. or $\mathbf{E}[Y_0] = \infty$. While Y_∞ can have infinite expectation even if $Y_\infty < \infty$ a.s., for Y_0 this is impossible since, in contrast with Y_∞, Y_0 is composed not of a few large contributions but many small ones so that a law of large numbers is in force. More precisely, by Corollary 24.15, we have

$$\mathbf{Var}[Y_t] = \int_{(t,1)} x^2 \nu(dx) \le \int_{(t,1)} x\,\nu(dx) = \mathbf{E}[Y_t] < \infty \quad \text{for all } t \in (0,1).$$

Hence, by Chebyshev's inequality,

$$\mathbf{P}\left[Y_t < \frac{\mathbf{E}[Y_t]}{2}\right] \le \frac{4\mathbf{Var}[Y_t]}{\mathbf{E}[Y_t]^2} \xrightarrow{t\to 0} 0.$$

Thus $Y_0 = \sup_{t\in(0,1)} Y_t \ge \mathbf{E}[Y_0]/2 = \infty$ almost surely.

Now assume that (i)–(iii) hold. By Theorem 24.14, Y has the Laplace transform

$$\mathbf{E}\left[e^{-tY}\right] = \exp\left(\int \nu(dx)\left(e^{-tx} - 1\right)\right).$$

By the Lévy–Khinchin formula (Theorem 16.14), Y is infinitely divisible with Lévy measure ν. $\square$

Corollary 24.18 *Let $\mu_i \in \mathcal{M}_1([0,\infty))$, $i = 1, 2$, be infinitely divisible distributions with canonical measures $\nu_i \in \mathcal{M}((0,\infty))$ and deterministic parts $\alpha_i \geq 0$ (compare Theorem 16.14). If we have*

$$\alpha_1 \leq \alpha_2 \quad \textit{and} \quad \nu_1\big([x,\infty)\big) \leq \nu_2\big([x,\infty)\big) \quad \textit{for all } x > 0, \tag{24.5}$$

then μ_1 is stochastically smaller than μ_2; i.e., $\mu_1 \leq_{\mathrm{st}} \mu_2$.

Proof (The proof follows [100, Proof of Lemma 6.1].) The idea is to use a coupling argument where based on one Poisson point process we construct the two random variables Y_1, Y_2 with $Y_i \sim \mu_i$, $i = 1, 2$, such that $Y_1 \leq Y_2$ almost surely. By Theorem 17.58, this yields the claim.

Let $G_i(x) := \nu_i([x,\infty))$, $i = 1, 2$, $x > 0$, and

$$\phi_i(y) := G_i^{-1}(y) = \inf\big\{x \geq 0 : G_i(x) \leq y\big\} \quad \text{for } y > 0.$$

If ν_i is finite, then $\phi_i(y) = 0$ for $y \geq \nu_i((0,\infty))$. Let λ denote the Lebesgue measure on $[0,\infty)$. By construction, for the image measure restricted to the positive reals, we have

$$\big(\lambda \circ \phi_i^{-1}\big)\big|_{(0,\infty)} = \nu_i, \quad i = 1, 2.$$

Now assume that X is a PPP on $(0,\infty)$ with intensity measure λ. By Theorem 24.16, the random measures

$$X_i := \left(\int \phi_i(x) X(dx)\right)\bigg|_{(0,\infty)} = X \circ \phi_i^{-1}$$

are PPPs with intensity measures ν_i, $i = 1, 2$. By Theorem 24.17, we thus have

$$Y_i := \alpha_i + \int \phi_i(x) X(dx) \sim \mu_i \quad \text{for } i = 1, 2.$$

However, by assumption, we have $G_1 \geq G_2$ which implies $\phi_1 \leq \phi_2$ and thus $Y_1 \leq Y_2$ a.s. $\square$

Example 24.19 By Corollary 16.10, for every nonnegative infinitely divisible distribution μ with Lévy measure ν, there exists a stochastic process $(Y_t)_{t\geq 0}$ with independent stationary increments and $Y_t \sim \mu^{*t}$ (hence with Lévy measure $t\nu$). Here we give a direct construction of this process. Let X be a PPP on $(0,\infty) \times [0,\infty)$ with intensity measure $\nu \otimes \lambda$ (here λ is the Lebesgue measure). Define $Y_0 = 0$ and

$$Y_t := \int_{(0,\infty)\times(0,t]} x\, X\big(d(x,s)\big).$$

By the mapping theorem, we have $X(\bullet \times (s,t]) \sim \mathrm{PPP}_{(t-s)\nu}$; hence $Y_t - Y_s$ is infinitely divisible with Lévy measure $(t-s)\nu$. The independence of the increments is evident. Note that $t \mapsto Y_t$ is right continuous and monotone increasing.

The process Y that we have just constructed is called a *subordinator* with Lévy measure ν. ◊

The procedure in the previous example can be generalized by allowing time sets more general than $[0,\infty)$.

Definition 24.20 A random measure Y is called infinitely divisible if, for any $n \in \mathbb{N}$, there exist i.i.d. random measures $Y_1, \ldots, Y_n$ with $Y = Y_1 + \ldots + Y_n$.

Theorem 24.21 *Let $\nu \in \mathcal{M}((0,\infty) \times E)$ with*

$$\int \mathbb{1}_A(t)(1 \wedge x)\nu\big(d(x,t)\big) < \infty \quad \textit{for all } A \in \mathcal{B}_b(E),$$

and let $\alpha \in \mathcal{M}(E)$. Let X be a PPP_ν *and*

$$Y(A) := \alpha(A) + \int x \mathbb{1}_A(t) X\big(d(x,t)\big) \quad \textit{for } A \in \mathcal{B}(E).$$

Then Y is an infinitely divisible random measure with independent increments. For $A \in \mathcal{B}(E)$, $Y(A)$ has the Lévy measure $\nu(\bullet \times A)$.

We call ν the canonical measure and α the deterministic part of Y.

Proof This is a direct consequence of Theorem 24.16 and Theorem 24.17. □

Remark 24.22 We can write Y as $Y = \alpha + \int x\delta_t X(d(x,t))$, where δ_t is the Dirac measure at $t \in E$. If instead of $x\,\delta_t$, we allow more general measures $\chi \in \mathcal{M}(E)$, then we get a representation

$$Y = \alpha + \int_{\mathcal{M}(E)} \chi X(d\chi),$$

where $X \sim \mathrm{PPP}_\nu$ on $\mathcal{M}(E)$ and $\nu \in \mathcal{M}(\mathcal{M}(E))$ with

$$\int \nu(d\chi)\big(\chi(A) \wedge 1\big) < \infty$$

for any $A \in \mathcal{B}_b(E)$. It can be shown that this is the most general form of an infinitely divisible measure on E. We call ν the canonical measure of Y and α the deterministic part. Y is characterized by its Laplace transform which obeys the Lévy–Khinchin formula:

$$\mathcal{L}_Y(f) = \exp\bigg(-\int f\,d\alpha + \int \nu(d\chi)\big(e^{-\int f\,d\chi} - 1\big)\bigg).$$

◊

Theorem 24.23 (Coloring theorem) *Let F be a further locally compact Polish space, let $\mu \in \mathcal{M}(E)$ be atom-free and let $(Y_x)_{x\in E}$ be i.i.d. random variables, independent of X, with values in F and distribution $\nu \in \mathcal{M}_1(F)$. Then*

$$Z(A) := \int \mathbb{1}_A(x, Y_x) X(dx), \quad A \in \mathcal{B}(E \times F),$$

is a $\mathrm{PPP}_{\mu\otimes\nu}$ *on* $E \times F$.

Proof This is left as an exercise. □

There is an obvious generalization of the coloring theorem: The assumption that μ is atom-free was needed in order that X have no double points. That is, for every unit mass that X produces, there is a different random variable Y_x. However, this can also be achieved by different means and in somewhat greater generality.

Accordingly, let E, F be locally compact Polish spaces, let $\mu \in \mathcal{M}(E)$ and let κ be a stochastic kernel from E to F with $\mu\kappa := \int \mu(dx)\kappa(x, \cdot) \in \mathcal{M}(F)$. Let $(Y_{x,t})_{x\in E, t\in[0,1]}$ be independent random variables with distributions $\mathbf{P}_{Y_{x,t}} = \kappa(x, \cdot)$ for $x \in E$ and $t \in [0, 1]$.

For $X \sim \mathrm{PPP}_\mu$, define the lifting $\tilde{X}$ as that PPP on $E \times [0, 1]$ with intensity measure $\mu \otimes \lambda\big|_{[0,1]}$, where λ is the Lebesgue measure. Clearly, $X \overset{\mathcal{D}}{=} \tilde{X}(\cdot \times [0, 1])$. The random measure $\tilde{X}$ can be understood as a realization of X in which the different points of X are assigned arbitrary $[0, 1]$-valued labels to distinguish them. Now let

$$X^\kappa(A) := \int \tilde{X}\big(d(x,t)\big)\mathbb{1}_A(Y_{x,t}) \quad \text{for } A \in \mathcal{B}(F).$$

Theorem 24.24 *X^κ is a random measure with* $\mathbf{P}_{X^\kappa} = \mathrm{PPP}_{\mu\kappa}$.

Proof Clearly, almost surely X^κ is a measure. For $A \in \mathcal{B}_b(F)$, we have by assumption

$$\mathbf{E}\big[X^\kappa(A)\big] = \mathbf{E}\bigg[\int \tilde{X}\big(d(x,t)\big)\kappa(x, A)\bigg] = (\mu\kappa)(A) < \infty.$$

Hence $X^\kappa(A) < \infty$ almost surely, and thus X^κ is a random measure. We compute the Laplace transform of X^κ. Let $g(x) := -\log \mathbf{E}[e^{-f(Y_{x,t})}]$. Then (since $\tilde{X}$ has no double points)

$$\begin{aligned}
\mathcal{L}_{X^\kappa}(f) &= \mathbf{E}\bigg[\exp\bigg(-\int \tilde{X}\big(d(x,t)\big) f(Y_{x,t})\bigg)\bigg] \\
&= \mathbf{E}\bigg[\prod_{(x,t):\tilde{X}(\{(x,t)\})=1} e^{-f(Y_{x,t})}\bigg] \\
&= \mathbf{E}\bigg[\prod_{(x,t):\tilde{X}(\{(x,t)\})=1} \mathbf{E}\big[e^{-f(Y_{x,t})}\big]\bigg]
\end{aligned}$$

$$= \mathbf{E}\Bigg[\prod_{(x,t):\tilde{X}(\{(x,t)\})=1} e^{-g(x)} \Bigg] = \mathcal{L}_X(g)$$

$$= \exp\bigg(\int \mu(dx) \big(\mathbf{E}\big[e^{-f(Y_{x,t})}\big] - 1 \big) \bigg)$$

$$= \exp\bigg(\int \mu(dx) \int \kappa(x, dy) \big(e^{-f(y)} - 1 \big) \bigg)$$

$$= \exp\bigg(\int \mu\kappa(dy) \big(e^{-f(y)} - 1 \big) \bigg). \qquad \square$$

Example 24.25 (PPP as invariant distribution) As an application of the previous theorem, consider a stochastic process on $E = \mathbb{Z}^d$ or $E = \mathbb{R}^d$ that consists of a system of independent random walks. Hence assume that we are given i.i.d. random variables Z_n^i, $i, n \in \mathbb{N}$ with distribution $\nu \in \mathcal{M}_1(E)$. Further, assume that, at time n, the position of the ith particle of our system of random walks is $S_n^i := S_0^i + \sum_{l=1}^n Z_l^i$, where S_0^i is an arbitrary, possibly random, starting point. Assume that the particles are indistinguishable. Hence we simply add the particles at each site:

$$X_n(A) := \sum_{i=1}^{\infty} \mathbb{1}_A\big(S_n^i\big) \quad \text{for } A \subset E.$$

Each X_n is a measure on E and, if at the beginning the particles are not too concentrated locally, it is a locally finite measure and hence a random measure. Assume that $X_0 \sim \mathrm{PPP}_\mu$ for some $\mu \in \mathcal{M}(E)$. Define $\kappa(x, \cdot) = \delta_x * \nu$, and write κ^n for the n-fold application of κ; that is, $\kappa^n(x, \cdot) = \delta_x * \nu^{*n}$. We thus get $X_0^\kappa \overset{\mathcal{D}}{=} X_1$. Indeed, independence of the motions of the individual particles in the definition of X_0^κ is exactly independence of the random walks. As X_1 is also a PPP, we get inductively $X_n^\kappa \overset{\mathcal{D}}{=} X_{n+1}$ and thus $X_n \sim \mathrm{PPP}_{\mu\kappa^n} = \mathrm{PPP}_{\mu*\nu^{*n}}$. In particular, $X_0 \overset{\mathcal{D}}{=} X_n$ if and only if $\mu * \nu = \mu$. Clearly, this is true if we have $E = \mathbb{Z}^d$ and μ the counting measure or if $E = \mathbb{R}^d$ and μ is the Lebesgue measure. For example, if $E = \mathbb{Z}^d$, then under rather mild assumptions on ν one can show that the counting measure $\mu = \lambda$ is the *unique* (up to multiples) solution of $\mu * \nu = \mu$. In this case, every invariant measure is a convex combination of PPPs with different intensity measures $\theta\lambda$. $\Diamond$

Exercise 24.2.1 Use an approximation with simple functions in order to show the claim of Corollary 24.15 without using characteristic functions.

Exercise 24.2.2 Prove the coloring theorem (Theorem 24.23).

Exercise 24.2.3 Let $p_1, p_2 \in (0, 1]$ and $r_1, r_2 > 0$. Show the following statement about the stochastic order of negative binomial distributions: $b^-_{r_1,p_1} \le_{\mathrm{st}} b^-_{r_2,p_2}$ if and only if

$$p_1 \ge p_2 \quad \text{and} \quad p_1^{r_1} \ge p_2^{r_2}.$$

24.3 The Poisson–Dirichlet Distribution*

The goal of this section is to solve the following problem. Take a stick of length 1. Choose a point of the stick uniformly at random and break the stick at this point. Put the left part of the stick (with length, say, W_1) aside. With the remaining part of the stick proceed just as with the original stick. Break it in two and put the left part (of length W_2) aside. Successively, we thus collect fractions of the stick of lengths $W_1, W_2, W_3, \ldots$. What is the joint distribution of $(W_1, W_2, \ldots)$? Furthermore, if we order the numbers $W_1, W_2, \ldots$ in decreasing order $W_{(1)} \geq W_{(2)} \geq \ldots$, what is the distribution of $(W_{(1)}, W_{(2)}, \ldots)$? And finally, why do we ask these questions in a chapter on Poisson point processes?

Answering these questions requires some preparation. We saw that the Beta distribution occurs naturally in Pólya's urn model as the limiting distribution of the fraction of balls of a given color. Clearly, Pólya's urn model can be considered for any number $n \geq 2$ of colors. The limiting distribution is then the n-dimensional generalization of the Beta distribution, namely the so-called Dirichlet distribution.

Define the $(n-1)$-dimensional simplex

$$\Delta_n := \big\{(x_1, \ldots, x_n) \in [0,1]^n : x_1 + \ldots + x_n = 1\big\}.$$

Definition 24.26 Let $n \in \{2, 3, \ldots\}$ and $\theta_1, \ldots, \theta_n > 0$. The *Dirichlet distribution* $\mathrm{Dir}_{\theta_1,\ldots,\theta_n}$ is the distribution on Δ_n that is defined for measurable $A \subset \Delta_n$ by

$$\mathrm{Dir}_{\theta_1,\ldots,\theta_n}(A) = \int \mathbb{1}_A(x_1, \ldots, x_n)\, f_{\theta_1,\ldots,\theta_n}(x_1, \ldots, x_n)\, dx_1 \ldots dx_{n-1}.$$

Here

$$f_{\theta_1,\ldots,\theta_n}(x_1, \ldots, x_n) = \frac{\Gamma(\theta_1 + \ldots + \theta_n)}{\Gamma(\theta_1) \ldots \Gamma(\theta_n)} x_1^{\theta_1 - 1} \ldots x_n^{\theta_n - 1}.$$

If the parameters $\theta_1, \ldots, \theta_n$ are integer-valued, they correspond to the numbers of balls of the different colors that are originally in the urn. Assume that the colors $n-1$ and n are light green and green and that in the dim light we cannot distinguish them. Then we should still end up with a Dirichlet distribution in the limit but with $n-1$ instead of n and with $\theta_{n-1} + \theta_n$ instead of θ_{n-1} and θ_n; that is, $\mathrm{Dir}_{\theta_1,\ldots,\theta_{n-2},\theta_{n-1}+\theta_n}$. Let $(M_t)_{t\geq 0}$ be the *Moran Gamma subordinator*, the stochastic process with right continuous, monotone increasing paths $t \mapsto M_t$ and independent, stationary, Gamma-distributed increments: $M_t - M_s \sim \Gamma_{1,t-s}$ for $t > s \geq 0$. An important connection between M and the Dirichlet distribution is revealed by the corollaries of the following theorem and by Theorem 24.32.

Theorem 24.27 *Let $n \in \mathbb{N}$, $\theta_1, \ldots, \theta_n > 0$ and $\Theta := \theta_1 + \ldots + \theta_n$. Let $X \sim \mathrm{Dir}_{\theta_1,\ldots,\theta_n}$ and let $Z \sim \Gamma_{1,\Theta}$ be independent random variables. Then the random variables $S_i := Z \cdot X_i$, $i = 1, \ldots, n$ are independent and $S_i \sim \Gamma_{1,\theta_i}$.*

Proof In the following, always let $x_n := 1 - \sum_{i=1}^{n-1} x_i$ and $s = \sum_{j=1}^{n} s_j$. Let

$$\Delta'_n := \left\{ (x_1, \ldots, x_{n-1}) \in (0,1)^{n-1} : \sum_{i=1}^{n-1} x_i < 1 \right\}.$$

For $x \in \Delta'_n$ and $z \geq 0$, the distribution of $(X_1, \ldots, X_{n-1}, Z)$ has the density

$$f(x_1, \ldots, x_{n-1}, z) = \prod_{j=1}^{n} \big(x_j^{\theta_j - 1} / \Gamma(\theta_j) \big) z^{\Theta - 1} e^{-z}.$$

Consider the map

$$F : \Delta'_n \times (0, \infty) \to (0, \infty)^n, \quad (x_1, \ldots, x_{n-1}, z) \mapsto (z x_1, \ldots, z x_n).$$

This map is invertible with inverse map

$$F^{-1} : (s_1, \ldots, s_n) \mapsto (s_1/s, \ldots, s_{n-1}/s, s).$$

The Jacobian determinant of F is $\det(F'(x_1, \ldots, x_{n-1}, z)) = z^{n-1}$. By the transformation formula for densities (Theorem 1.101), $(S_1, \ldots, S_n)$ has density

$$\begin{aligned} g(s_1, \ldots, s_n) &= \frac{f(F^{-1}(s_1, \ldots, s_n))}{|\det(F'(F^{-1}(s_1, \ldots, s_n)))|} \\ &= \frac{s^{\Theta - 1} e^{-s}}{s^{n-1}} \prod_{j=1}^{n} \big((s_j/s)^{\theta_j - 1} / \Gamma(\theta_j) \big) \\ &= \prod_{j=1}^{n} \big(s_j^{\theta_j - 1} e^{-s_j} / \Gamma(\theta_j) \big). \end{aligned}$$

However, this is the density for independent Gamma distributions. □

Corollary 24.28 *If $t_i := \sum_{j=1}^{i} \theta_j$ for $i = 0, \ldots, n$, then the random variables $X = ((M_{t_i} - M_{t_{i-1}})/M_{t_n}, i = 1, \ldots, n)$ and $S := M_{t_n}$ are independent with distributions $X \sim \mathrm{Dir}_{\theta_1, \ldots, \theta_n}$ and $S \sim \Gamma_{1, t_n}$.*

Corollary 24.29 *Let $(X_1, \ldots, X_n) \sim \mathrm{Dir}_{\theta_1, \ldots, \theta_n}$. Then $X_1 \sim \beta_{\theta_1, \sum_{i=2}^{n} \theta_i}$ and $(X_2/(1 - X_1), \ldots, X_n/(1 - X_1)) \sim \mathrm{Dir}_{\theta_2, \ldots, \theta_n}$ are independent.*

Proof Let M be as in Corollary 24.28. Then $X_1 = M_{t_1}/M_{t_n} \sim \beta_{\theta_1, t_n - \theta_1}$. Since $X_1 = (\frac{M_{t_n} - M_{t_1}}{M_{t_1}} + 1)^{-1}$, we see that X_1 depend only on M_{t_1} and $M_{t_n} - M_{t_1}$. On the other hand,

$$\left(\frac{X_2}{1 - X_1}, \ldots, \frac{X_n}{1 - X_1} \right) = \left(\frac{M_{t_2} - M_{t_1}}{M_{t_n} - M_{t_1}}, \ldots, \frac{M_{t_n} - M_{t_{n-1}}}{M_{t_n} - M_{t_1}} \right)$$

is independent of M_{t_1}. By Corollary 24.28, it is also independent of $M_{t_n} - M_{t_1}$ and is $\mathrm{Dir}_{\theta_2,\ldots,\theta_n}$-distributed. □

Corollary 24.30 *Let $V_1, \ldots, V_{n-1}$ be independent, $V_i \sim \beta_{\theta_i, \theta_{i+1}+\ldots+\theta_n}$ and $V_n = 1$. Then*

$$\left(V_1, (1-V_1)V_2, (1-V_1)(1-V_2)V_3, \ldots, \left(\prod_{i=1}^{n-1}(1-V_i)\right)V_n\right) \sim \mathrm{Dir}_{\theta_1,\ldots,\theta_n}.$$

Proof This follows by iterating the claim of Corollary 24.29. □

It is natural to ask what happens if we distinguish more and more colors (instead of pooling them). For simplicity, consider a symmetric situation where we have $\theta_1 = \ldots = \theta_n = \theta/n$ for some $\theta > 0$. Hence we consider

$$\mathrm{Dir}_{\theta;n} := \mathrm{Dir}_{\theta,\ldots,\theta} \quad \text{for } \theta > 0.$$

If $X^n = (X_1^n, \ldots, X_n^n) \sim \mathrm{Dir}_{\theta/n;n}$, then, by symmetry, we have $\mathbf{E}[X_i^n] = 1/n$ for every $n \in \mathbb{N}$ and $i = 1, \ldots, n$. Hence, clearly $(X_1^n, \ldots, X_k^n) \overset{n\to\infty}{\Longrightarrow} 0$ for any $k \in \mathbb{N}$. In order to obtain a nontrivial limit, one possibility is to reorder the values by decreasing size: $X_{(1)}^n \ge X_{(2)}^n \ge \ldots$.

Definition 24.31 Let $\theta > 0$ and let $(M_t)_{t\in[0,\theta]}$ be a Moran Gamma subordinator. Let $m_1 \ge m_2 \ge \ldots \ge 0$ be the jump sizes of M in decreasing order and let $\tilde{m}_i = m_i/M_\theta$, $i = 1, 2, \ldots$. The distribution of the random variables $(\tilde{m}_1, \tilde{m}_2, \ldots)$ on $S := \{(x_1 \ge x_2 \ge \ldots \ge 0) : x_1 + x_2 + \ldots = 1\}$ is called the *Poisson–Dirichlet distribution* PD_θ with parameter $\theta > 0$.

To be honest, we still have to show that $\sum_{i=1}^\infty \tilde{m}_i = 1$. To this end, let Y be a PPP on $(0,\infty) \times (0,\theta]$ with intensity measure $\nu \otimes \lambda$, where λ is the Lebesgue measure and $\nu(dx) = e^{-x}x^{-1}\,dx$ is the Lévy measure of the $\Gamma_{1,1}$ distribution. We can define M by $M_t := \sum_{(x,s):Y(\{x,s\})=1, s\le t} x$. Now we have $m_1 = \sup\{x \in (0,\infty) : Y(\{x\}\times(0,\theta]) = 1\}$. Inductively, we get $m_n = \sup\{x < m_{n-1} : Y(\{x\}\times(0,\theta]) = 1\}$ for $n \ge 2$. Interchanging the order of summations, we obtain $M_\theta = \sum_{n=1}^\infty m_n$.

Theorem 24.32 *If $X^n \sim \mathrm{Dir}_{\theta/n;n}$ for $n \in \mathbb{N}$, then $\mathbf{P}_{(X_{(1)}^n, X_{(2)}^n, \ldots)} \overset{n\to\infty}{\longrightarrow} \mathrm{PD}_\theta$.*

Proof The idea is to express the random variables X^n, $n \in \mathbb{N}$, in terms of the increments of the Moran Gamma subordinator $(M_t)_{t\in[0,\theta]}$ in such a way that convergence of distributions implies almost sure convergence. Hence, let $X_i^n = (M_{\theta i/n} - M_{\theta(i-1)/n})/M_\theta$. By Corollary 24.28, we have $X^n \sim \mathrm{Dir}_{\theta/n;n}$. Let $t_1, t_2, \ldots \in (0,\theta]$ be the positions of the jumps $m_1 \ge m_2 \ge \ldots$. Evidently, $X_{(1)}^n \ge \tilde{m}_1$ for every n. If n is large enough that $|t_1 - t_2| > \theta/n$, then $X_{(2)}^n \ge \tilde{m}_2$. Inductively, we get

$\liminf_{n\to\infty} X^n_{(i)} \ge \tilde{m}_i$ almost surely. Using the convention $X^n_{(i)} = 0$ for $i > n$, we have $\sum_{i=1}^\infty X^n_{(i)} = 1$ for every $n \in \mathbb{N}$. By Fatou's lemma, we thus get

$$1 = \sum_{i=1}^\infty \tilde{m}_i \le \sum_{i=1}^\infty \liminf_{n\to\infty} X^n_{(i)} \le \liminf_{n\to\infty} \sum_{i=1}^\infty X^n_{(i)} = 1.$$

Therefore, $\lim_{n\to\infty} X^n_{(i)} = \tilde{m}_i$ almost surely. □

Instead of *ordering* the values of X^n by their sizes, there is a different way of arranging the terms so that the distributions converge. Think of a biological population in which a certain phenotypical property can be measured with different levels of precision. If we distinguish n different values of this property, then we write X^n_i for the proportion of the population that has type $i \in \{1, \ldots, n\}$.

Now successively choose individuals from the population at random. Let I^n_1 be the type of the first individual. Denote by I^n_2 the type of the first individual that is not of type I^n_1. That is, I^n_2 is the second *type* that we see. Now inductively define I^n_k as the kth type that we see; that is, the type of the first individual that has none of the types $I^n_1, \ldots, I^n_{k-1}$. Consider the vector $\hat{X}^n = (\hat{X}^n_1, \ldots, \hat{X}^n_n)$, where $\hat{X}^n_k = X^n_{I^n_k}$. Since the probability of the event $\{I^n_1 = i\}$ is proportional to the size of the subpopulation of type i, we say that $\hat{X}^n$ is the successively size-biased vector.

The distribution of $\hat{X}^n$ does not change if we change the order of the $X^n_1, \ldots, X^n_n$. For example, instead of $X^n_1, \ldots, X^n_n$, we can use the order statistics $(X^n_{(1)}, \ldots, X^n_{(n)})$ and again end up with $\hat{X}^n$ as the successively size-biased vector. Hence we can define the successively size-biased vector $\hat{X}$ for the infinite vector $X \sim \mathrm{PD}_\theta$. If $X^n \sim \mathrm{Dir}_{\theta/n;n}$, then by Theorem 24.32, we have $\hat{X}^n \overset{n\to\infty}{\Longrightarrow} \hat{X}$. Hence we can compute the distribution of $\hat{X}$.

Theorem 24.33 *Let $\theta > 0$ and $X^n \sim \mathrm{Dir}_{\theta/n;n}$, $n \in \mathbb{N}$. Let $X \sim \mathrm{PD}_\theta$. Further, let $V_1, V_2, \ldots$ be i.i.d. random variables on* $[0, 1]$ *with density $x \mapsto \theta(1-x)^{\theta-1}$. Define $Z_1 = V_1$ and $Z_k = (\prod_{i=1}^{k-1}(1 - V_i))V_k$ for $k \ge 2$. Then*:

(i) $\hat{X}^n \overset{n\to\infty}{\Longrightarrow} \hat{X}$.
(ii) $\hat{X} \overset{\mathcal{D}}{=} Z$.

The distribution of Z is called the GEM_θ distribution (*Griffiths–Engen–McCloskey*).

Proof Statement (i) was shown in the discussion preceding the theorem. In order to show (ii), we compute the distribution of $\hat{X}^n$ and show that it converges to the distribution of Z.

Let $\hat{X}^{n,1}$ be the vector $X^{n,1} = (X^n_{I^n_1}, X_2, \ldots, X^n_{I^n_1-1}, X^n_{I^n_1+1}, \ldots, X^n_n)$, in which only the first coordinate is sampled size-biasedly. We show that

$$\hat{X}^{n,1} \sim \mathrm{Dir}_{(\theta/n)+1,\theta/n,\ldots,\theta/n}\,. \tag{24.6}$$

Let $f(x) = (\Gamma(\theta)/\Gamma(\theta/n)^n) \cdot \prod_{k=1}^n x_k^{(\theta/n)-1}$ be the density of $\mathrm{Dir}_{\theta/n;n}$. We compute the density $f^{n,1}$ of $X^{n,1}$ by decomposing according to the value i of I_1^n:

$$\begin{aligned} f^{n,1}(x) &= \sum_{i=1}^n x_1 f(x_2, \ldots, x_i, x_1, x_{i+1}, \ldots, x_n) = n x_1 f(x) \\ &= \frac{n\Gamma(\theta)}{\Gamma(\theta/n)^n} x_1^{\theta/n} \prod_{i=2}^n x_i^{(\theta/n)-1} \\ &= \frac{\Gamma(\theta+1)}{\Gamma((\theta/n)+1)\Gamma(\theta/n)^{n-1}} x_1^{\theta/n} \prod_{i=2}^n x_i^{(\theta/n)-1}. \end{aligned}$$

However, this is the density of $\mathrm{Dir}_{(\theta/n)+1,\theta/n,\ldots,\theta/n}$. By Corollary 24.29, we have

$$\hat{X}^{n,1} \stackrel{\mathcal{D}}{=} \big(V_1^n, \big(1 - V_1^n\big)Y_1, \ldots, \big(1 - V_1^n\big)Y_{n-1}\big),$$

where

$$V_1^n \sim \beta_{(\theta/n)+1,\theta(n-1)/n} \quad \text{and} \quad Y = (Y_1, \ldots, Y_{n-1}) \sim \mathrm{Dir}_{\theta/n;n-1}$$

are independent. Applying this to Y, we get inductively

$$\hat{X}^n \stackrel{\mathcal{D}}{=} Z^n, \tag{24.7}$$

where

$$Z_1^n = V_1^n \quad \text{and} \quad Z_k^n = \left(\prod_{i=1}^{k-1} \big(1 - V_i^n\big)\right) V_k^n \quad \text{for } k \geq 2$$

and where $V_1^n, \ldots, V_{n-1}^n$ are independent and $V_i^n \sim \beta_{(\theta/n)+1,\theta(n-i)/n}$. Now it is easy to check that $\beta_{(\theta/n)+1,\theta(n-i)/n} \stackrel{n\to\infty}{\longrightarrow} \beta_{1,\theta}$ for every $i \in \mathbb{N}$. Recall that $\beta_{1,\theta}$ has the density $x \mapsto \theta(1-x)^{\theta-1}$. Hence $V_i^n \stackrel{n\to\infty}{\Longrightarrow} V_i$ for every i and thus $Z^n \stackrel{n\to\infty}{\Longrightarrow} Z$ and $\hat{X}^n \stackrel{n\to\infty}{\Longrightarrow} Z$. Together with (i), this proves claim (ii). □

At the beginning of this chapter, we raised the question of how the sizes $W_1, W_2, \ldots$ of the stick lengths are distributed if at each step, we break the remaining part of the stick at a point chosen uniformly at random. The preceding theorem gives the answer: The vector $(W_{(1)}, W_{(2)}, \ldots)$ has distribution PD_1, and $(W_1, W_2, \ldots)$ has distribution GEM_1.

The Chinese Restaurant Process

We will study a further situation in which the Poisson–Dirichlet distribution arises naturally. As the technical details get a bit tricky, we content ourselves with the description of the problem and with stating (but not proving) two fundamental theorems. An excellent reference for this type of problem is [130].

Consider a Chinese restaurant with countably many enumerated round tables. At each table, there is enough space for arbitrarily many guests. Initially, the restaurant is empty. One by one an infinite number of guests arrive. The first guest sits down at table number one. If there are already n guests sitting at k tables, then the $(n+1)$th guest can choose between sitting down at any of the k occupied tables or at the free table with the smallest number (that is, $k+1$). Assume that the guest makes his choice at random (and independently of the previous choices of the other guests). For $l \le k$, denote by N_l^n the number of guests at the lth table and assume that the probability of choosing the lth table is $(N_l^n - \alpha)/(n+\theta)$. Then the probability of choosing the first free table is $(\theta + k\alpha)/(n+\theta)$. Here $\alpha \in [0,1]$ and $\theta > -\alpha$ are parameters. We say that $(N^n)_{n\in\mathbb{N}} = (N_1^n, N_2^n, \ldots)_{n\in\mathbb{N}}$ is the *Chinese restaurant process* with parameters (α, θ).

In the special case $\alpha = 0$, there is a nice interpretation: Assume that the new guest can also choose his seating position at the table (that is, his neighbor to the right). Then, for any of the present guests, the probability of being chosen as a right neighbor is $1/(n+\theta)$. The probability of starting a new table is $\theta/(n+\theta)$.

In order to study the large n behavior of $N^n/n = (N_1^n/n, N_2^n/n, \ldots)$, we extend the Poisson–Dirichlet distribution and the GEM distribution by a further parameter.

Definition 24.34 Let $\alpha \in [0,1)$ and $\theta > -\alpha$. Let $V_1, V_2, \ldots$ be independent and $V_i \sim \beta_{1-\alpha,\theta+i\alpha}$. Define $Z = (Z_1, Z_2, \ldots)$ by $Z_1 = V_1$ and

$$Z_k = V_k \prod_{i=1}^{k-1} (1 - V_i) \quad \text{for } k \ge 2.$$

Then $\mathrm{GEM}_{\alpha,\theta} := \mathbf{P}_Z$ is called the *GEM distribution* with parameters (α, θ). The distribution of the size-biased vector $(Z_{(1)}, Z_{(2)}, \ldots)$ is called the *Poisson–Dirichlet distribution* with parameters (α, θ), or briefly $\mathrm{PD}_{\alpha,\theta}$.

Explicit formulas for the densities of the finite-dimensional marginals of $\mathrm{PD}_{\alpha,\theta}$ can be found in [132]. Note that, for $\alpha = 0$, we recover the classical distributions $\mathrm{GEM}_\theta = \mathrm{GEM}_{0,\theta}$ and $\mathrm{PD}_\theta = \mathrm{PD}_{0,\theta}$.

Theorem 24.35 *Let $\alpha \in [0,1)$, $\theta > -\alpha$ and let $(N^n)_{n\in\mathbb{N}}$ be the Chinese restaurant process with parameters (α, θ). Then* $\mathbf{P}_{N^n/n} \xrightarrow{n\to\infty} \mathrm{PD}_{\alpha,\theta}$.

Proof See [129] or [130, Theorem 25]. □

As for the one-parameter Poisson–Dirichlet distribution, there is a representation of $\mathrm{PD}_{\alpha,\theta}$ in terms of the size-ordered jumps of a certain subordinator. In the following, let $\alpha \in (0,1)$ and let $(M_t)_{t\in[0,1]}$ be an α-stable subordinator; that is, a subordinator with Lévy measure $\nu(dx) = x^{-\alpha-1}\,dx$. Further, let $m_1 \ge m_2 \ge \ldots \ge 0$ be the jumps of M, $\tilde{m}_i = m_i/M_1$ for $i \in \mathbb{N}$, and $\tilde{m} = (\tilde{m}_1, \tilde{m}_2, \ldots)$. We quote the following theorem from [130, Section 4.2].

Theorem 24.36 *Let* $\alpha \in (0, 1)$.

(i) $\tilde{m} \sim \mathrm{PD}_{\alpha,0}$.
(ii) *If* $\theta > -\alpha$, *then* $\mathrm{PD}_{\alpha,\theta} \ll \mathrm{PD}_{\alpha,0} = \mathbf{P}[\tilde{m} \in \cdot\,]$ *and the density is*

$$\mathrm{PD}_{\alpha,\theta}(dx) = \frac{M_1^{-\theta}}{\mathbf{E}[M_1^{-\theta}]}\mathbf{P}[\tilde{m} \in dx].$$

Exercise 24.3.1 Let $(X, 1-X) \sim \mathrm{Dir}_{\theta_1,\theta_2}$. Show that $X \sim \beta_{\theta_1,\theta_2}$ is Beta-distributed.

Exercise 24.3.2 Let $X = (X_1, \ldots, X_n) \sim \mathrm{Dir}_{\theta_1,\ldots,\theta_n}$. Show the following.

(i) For any permutation σ on $\{1, \ldots, n\}$, we have

$$(X_{\sigma(1)}, \ldots, X_{\sigma(n)}) \sim \mathrm{Dir}_{\theta_{\sigma(1)},\ldots,\theta_{\sigma(n)}}.$$

(ii) $(X_1, \ldots, X_{n-2}, X_{n-1} + X_n) \sim \mathrm{Dir}_{\theta_1,\ldots,\theta_{n-2},\theta_{n-1}+\theta_n}$.

Exercise 24.3.3 Let $(N^n)_{n\in\mathbb{N}}$ be the Chinese restaurant process with parameters $(0, \theta)$.

(i) Let $\theta = 1$.

  (a) Show that $\mathbf{P}[N_1^n = k] = 1/n$ for any $k = 1, \ldots, n$,
  (b) Show that, for $k_l = 1, \ldots, n - (k_1 + \ldots + k_{l-1})$,

$$\mathbf{P}\big[N_l^n = k_l \,\big|\, N_1^n = k_1, \ldots, N_{l-1}^n = k_{l-1}\big] = \frac{1}{n - (k_1 + \ldots + k_{l-1})}.$$

  (c) Infer the claim of Theorem 24.35 in the case $\alpha = 0$ and $\theta = 1$.

(ii) Let $\theta > 0$.

  (a) Show that $n\mathbf{P}[N_1^n = \lfloor nx \rfloor] \xrightarrow{n\to\infty} \theta(1-x)^{\theta-1}$ for $x \in (0, 1)$.
  (b) Show that

$$n\mathbf{P}\big[N_l^n = \lfloor nx_l \rfloor \,\big|\, N_1^n = \lfloor nx_1 \rfloor, \ldots, N_{l-1}^n = \lfloor nx_{l-1} \rfloor\big] \xrightarrow{n\to\infty} (\theta/y_l)(1 - x_l/y_l)^{\theta-1}$$

  for $x_1, \ldots, x_l \in (0, 1)$ with $y_l = 1 - (x_1 + \ldots + x_{l-1}) > x_l$.
  (c) As in (i), infer the claim of Theorem 24.35 for $\alpha = 0$ and $\theta > 0$.