ECE 534: Random Processes

Lecture Notes: Galton–Watson Branching Processes and the Poisson Process

1 Markov processes

Informally, a random process is *Markov* if, given its present value, its future is independent of its past: the present state summarizes everything about the history that is relevant for predicting the future.

Definition 1.1 (Markov process)**.** A random process $X = (X_t : t \in \mathbb{T})$, with index set $\mathbb{T} = \mathbb{N}$ (discrete time) or $\mathbb{T} = [0, \infty)$ (continuous time), is a *Markov process* if for all $n \geq 1$, all times $t_1 < t_2 < \cdots < t_n < t_{n+1}$ in $\mathbb{T}$, and all (Borel) sets B,

$$\mathbb{P}\{X_{t_{n+1}} \in B \mid X_{t_1}, \ldots, X_{t_n}\} = \mathbb{P}\{X_{t_{n+1}} \in B \mid X_{t_n}\}.$$

Equivalently, for every t, the future $(X_u : u > t)$ and the past $(X_s : s < t)$ are conditionally independent given the present X_t.

When the state space $\mathcal{S}$ is countable, a Markov process is called a *Markov chain*, and its evolution is described by the *transition probabilities*

$$p_{ij}(s,t) = \mathbb{P}\{X_t = j \mid X_s = i\}, \qquad s \leq t, \quad i, j \in \mathcal{S}.$$

The process is *time-homogeneous* if $p_{ij}(s,t)$ depends only on $t - s$; we then write $p_{ij}(\tau)$, and in discrete time the one-step probabilities $p_{ij} = p_{ij}(1)$ form the *transition matrix* $P = (p_{ij})$. In discrete time the Markov property reduces to the one-step condition

$$\mathbb{P}\{X_{n+1} = j \mid X_0 = i_0, \ldots, X_{n-1} = i_{n-1}, X_n = i\} = p_{ij},$$

and the distribution of the whole process is determined by P together with the initial distribution of X_0. The transition probabilities satisfy the Chapman–Kolmogorov equations $p_{ij}(s,u) = \sum_k p_{ik}(s,t)\, p_{kj}(t,u)$ for $s \leq t \leq u$.

A convenient sufficient condition is the following: if $X_{n+1} = f(X_n, U_n)$ for a deterministic function f and a sequence $U_0, U_1, \ldots$ of independent random variables that is independent of X_0, then (X_n) is a Markov process. Similarly, any process with independent increments is Markov, since $X_t = X_s + (X_t - X_s)$ with the increment independent of $(X_u : u \leq s)$.

Example 1.2 (Processes that are Markov)**.**

(a) *I.i.d. sequences.* If $X_0, X_1, \ldots$ are independent, then trivially the future is independent of the past, so X is Markov.

(b) *Random walks.* $S_n = X_1 + \cdots + X_n$ with X_i i.i.d. satisfies $S_{n+1} = S_n + X_{n+1}$, so it is Markov. For the simple random walk on $\mathbb{Z}$, $p_{i,i+1} = p$ and $p_{i,i-1} = 1 - p$.

(c) *Autoregressive processes.* $X_{n+1} = aX_n + W_n$ with (W_n) i.i.d. and independent of X_0 is Markov by the sufficient condition above.

(d) *Galton–Watson processes* (Section 2). $Z_{n+1} = \sum_{i=1}^{Z_n} X_{n,i}$ is a function of Z_n and fresh independent offspring counts, so Z is a time-homogeneous Markov chain on $\mathbb{N}$ with $p_{kj} = \mathbb{P}\{X_1 + \cdots + X_k = j\}$.

(e) *The Poisson process* (Section 3). It has independent increments, hence is a continuous-time Markov chain with $p_{ij}(\tau) = e^{-\lambda\tau}(\lambda\tau)^{j-i}/(j-i)!$ for $j \geq i$.

(f) *Brownian motion.* It also has independent increments, so it is a Markov process with continuous state space; given $W_s = x$, the value W_t is $N(x, \sigma^2(t-s))$.

Remark 1.3. Not every process is Markov. For instance, the moving average $Y_n = U_n + U_{n-1}$ with (U_n) i.i.d. is generally not Markov: Y_{n-1} carries information about U_{n-1} beyond what Y_n reveals, and this helps predict $Y_{n+1} = U_{n+1} + U_n$. Also, a function of a Markov process need not be Markov.

2 Galton–Watson branching processes

2.1 Definition

A Galton–Watson process models the size of a population in which each individual, independently of all others, produces a random number of offspring and then dies. Let $(p_k)_{k\geq 0}$ be a probability mass function on $\mathbb{N} = \{0, 1, 2, \dots\}$, called the *offspring distribution*, and let X denote a generic random variable with $\mathbb{P}\{X = k\} = p_k$. We write

$$m = \mathbb{E}[X] = \sum_{k\geq 0} k\, p_k, \qquad \sigma^2 = \operatorname{Var}(X),$$

and assume throughout that $m < \infty$.

Definition 2.1 (Galton–Watson process)**.** Let $\{X_{n,i} : n \geq 0,\ i \geq 1\}$ be i.i.d. random variables with the offspring distribution (p_k). The *Galton–Watson process* $Z = (Z_n : n \geq 0)$ is defined by $Z_0 = 1$ and

$$Z_{n+1} = \sum_{i=1}^{Z_n} X_{n,i}, \qquad n \geq 0, \tag{1}$$

with the convention that an empty sum is zero. Here Z_n is the size of the n-th generation and $X_{n,i}$ is the number of children of the i-th individual of generation n.

Let $\mathcal{F}_n = \sigma(X_{\ell,i} : \ell < n,\ i \geq 1)$ denote the information contained in the first n generations. Then Z_n is $\mathcal{F}_n$-measurable, and the family $(X_{n,i})_{i\geq 1}$ is independent of $\mathcal{F}_n$. In particular, Z_n is independent of $(X_{n,i})_{i\geq 1}$. It follows that Z is a time-homogeneous Markov chain on $\mathbb{N}$, and state 0 is absorbing: once the population dies out it stays extinct.

To rule out trivial cases we assume $p_1 < 1$; if $p_1 = 1$ then $Z_n = 1$ for all n.

2.2 Wald's identity

Equation (1) expresses Z_{n+1} as a sum of i.i.d. terms in which the *number* of terms is itself random. The mean of such a sum is given by Wald's identity.

Theorem 2.2 (Wald's identity)**.** *Let $X_1, X_2, \dots$ be i.i.d. random variables with $\mathbb{E}|X_1| < \infty$, and let N be a random variable taking values in $\mathbb{N}$ with $\mathbb{E}[N] < \infty$. Suppose that for every $i \geq 1$ the event $\{N \geq i\}$ is independent of X_i. (This holds in particular if N is independent of the whole sequence (X_i).) Then, with $S_N = \sum_{i=1}^{N} X_i$,*

$$\mathbb{E}[S_N] = \mathbb{E}[N]\,\mathbb{E}[X_1].$$

Proof. The key step is to write the random sum as an infinite sum with random indicators:

$$S_N = \sum_{i=1}^{N} X_i = \sum_{i=1}^{\infty} X_i\,\mathbf{1}_{\{N \geq i\}}.$$

Step 1: nonnegative terms. Suppose first that $X_i \geq 0$. Then every term in the series is nonnegative, so by the monotone convergence theorem (Tonelli) expectation and summation may be interchanged:

$$\mathbb{E}[S_N] = \sum_{i=1}^{\infty} \mathbb{E}\big[X_i\,\mathbf{1}_{\{N \geq i\}}\big] = \sum_{i=1}^{\infty} \mathbb{E}[X_i]\,\mathbb{P}\{N \geq i\} = \mathbb{E}[X_1] \sum_{i=1}^{\infty} \mathbb{P}\{N \geq i\},$$

where the second equality uses the independence of X_i and $\{N \geq i\}$, and the third uses that the X_i are identically distributed. Finally, for any $\mathbb{N}$-valued random variable,

$$\sum_{i=1}^{\infty} \mathbb{P}\{N \geq i\} = \sum_{i=1}^{\infty} \sum_{k=i}^{\infty} \mathbb{P}\{N = k\} = \sum_{k=1}^{\infty} \sum_{i=1}^{k} \mathbb{P}\{N = k\} = \sum_{k=1}^{\infty} k\,\mathbb{P}\{N = k\} = \mathbb{E}[N].$$

Hence $\mathbb{E}[S_N] = \mathbb{E}[N]\,\mathbb{E}[X_1]$.

Step 2: general case. Write $X_i = X_i^+ - X_i^-$ with $X_i^{\pm} \geq 0$. Each of the sequences (X_i^+) and (X_i^-) satisfies the hypotheses of Step 1 (independence of $\{N \geq i\}$ is preserved under functions of X_i). Applying Step 1 to $\sum_{i=1}^{\infty} |X_i| \mathbf{1}_{\{N \geq i\}}$ shows that this sum has finite mean $\mathbb{E}[N]\mathbb{E}|X_1|$, so the series defining S_N converges absolutely and the interchange of $\mathbb{E}$ and $\sum$ is justified by dominated convergence. Applying Step 1 separately to the positive and negative parts gives

$$\mathbb{E}[S_N] = \mathbb{E}[N]\,\mathbb{E}[X_1^+] - \mathbb{E}[N]\,\mathbb{E}[X_1^-] = \mathbb{E}[N]\,\mathbb{E}[X_1].$$ □

Remark 2.3. The same proof works when N is a *stopping time* for the sequence (X_i), i.e. when $\{N \leq i-1\}$ (equivalently $\{N \geq i\}$) is determined by $X_1, \dots, X_{i-1}$. Then $\{N \geq i\}$ is independent of X_i, which is the only property used above.

2.3 The mean of Z_n

Proposition 2.4. *For all $n \geq 0$, $\mathbb{E}[Z_n] = m^n$.*

Proof. Apply Wald's identity to (1) with $N = Z_n$ and the i.i.d. sequence $X_{n,1}, X_{n,2}, \dots$ in place of $X_1, X_2, \dots$. As noted above, Z_n is a function of the generations before n and is therefore independent of $(X_{n,i})_{i \geq 1}$, so the hypotheses of Theorem 2.2 hold (the finiteness of $\mathbb{E}[Z_n]$ follows inductively from the same computation). Therefore

$$\mathbb{E}[Z_{n+1}] = \mathbb{E}[Z_n]\,\mathbb{E}[X] = m\,\mathbb{E}[Z_n].$$

Since $\mathbb{E}[Z_0] = 1$, induction gives $\mathbb{E}[Z_n] = m^n$. □

The same conclusion can be phrased conditionally: given $Z_n = k$, the next generation is a sum of k i.i.d. copies of X, so

$$\mathbb{E}[Z_{n+1} \mid \mathcal{F}_n] = m\,Z_n. \tag{2}$$

Variance. If in addition $\sigma^2 < \infty$, the conditional variance formula gives

$$\operatorname{Var}(Z_{n+1}) = \mathbb{E}[\operatorname{Var}(Z_{n+1} \mid Z_n)] + \operatorname{Var}(\mathbb{E}[Z_{n+1} \mid Z_n]) = \sigma^2 \mathbb{E}[Z_n] + m^2 \operatorname{Var}(Z_n),$$

that is,

$$V_{n+1} = \sigma^2 m^n + m^2 V_n, \qquad V_0 = 0,$$

where $V_n = \operatorname{Var}(Z_n)$. Solving the recursion,

$$\operatorname{Var}(Z_n) = \begin{cases} \sigma^2 m^{n-1} \dfrac{m^n - 1}{m - 1}, & m \neq 1, \\ n\sigma^2, & m = 1. \end{cases} \tag{3}$$

2.4 Generating functions

The probability generating function of the offspring distribution is

$$G(s) = \mathbb{E}[s^X] = \sum_{k \geq 0} p_k s^k, \qquad 0 \leq s \leq 1.$$

On $[0, 1]$, G is nondecreasing and convex, with $G(0) = p_0$, $G(1) = 1$, and $G'(1^-) = m$. It is strictly convex whenever $p_0 + p_1 < 1$.

Let $G_n(s) = \mathbb{E}[s^{Z_n}]$. Conditioning on $\mathcal{F}_n$ and using independence of the offspring counts,

$$\mathbb{E}\big[s^{Z_{n+1}} \mid \mathcal{F}_n\big] = \mathbb{E}\Big[\prod_{i=1}^{Z_n} s^{X_{n,i}} \;\Big|\; \mathcal{F}_n\Big] = G(s)^{Z_n},$$

so $G_{n+1}(s) = \mathbb{E}[G(s)^{Z_n}] = G_n(G(s))$. Since $G_0(s) = s$, we obtain

$$G_n = \underbrace{G \circ G \circ \cdots \circ G}_{n \text{ times}}, \qquad \text{and also} \qquad G_{n+1} = G \circ G_n. \tag{4}$$

2.5 Extinction probability and the three regimes

Let $\{\text{extinction}\} = \{Z_n = 0 \text{ for some } n\}$. Because 0 is absorbing, $\{Z_n = 0\} \subseteq \{Z_{n+1} = 0\}$, so

$$q := \mathbb{P}\{\text{extinction}\} = \lim_{n \to \infty} \mathbb{P}\{Z_n = 0\} = \lim_{n \to \infty} G_n(0).$$

Theorem 2.5 (Extinction)**.** *Assume $p_1 < 1$. The extinction probability q is the smallest root of the equation $s = G(s)$ in $[0, 1]$. Moreover,*

- *if $m \leq 1$, then $q = 1$;*
- *if $m > 1$, then $q < 1$, and q is the unique root of $s = G(s)$ in $[0, 1)$.*

The value of m therefore separates three regimes.

Subcritical, $m < 1$. Extinction is certain, and it happens fast. By Markov's inequality and Proposition 2.4,

$$\mathbb{P}\{Z_n > 0\} = \mathbb{P}\{Z_n \geq 1\} \leq \mathbb{E}[Z_n] = m^n \to 0$$

geometrically. The expected total progeny is finite: $\mathbb{E}\big[\sum_{n \geq 0} Z_n\big] = \sum_{n \geq 0} m^n = 1/(1 - m)$.

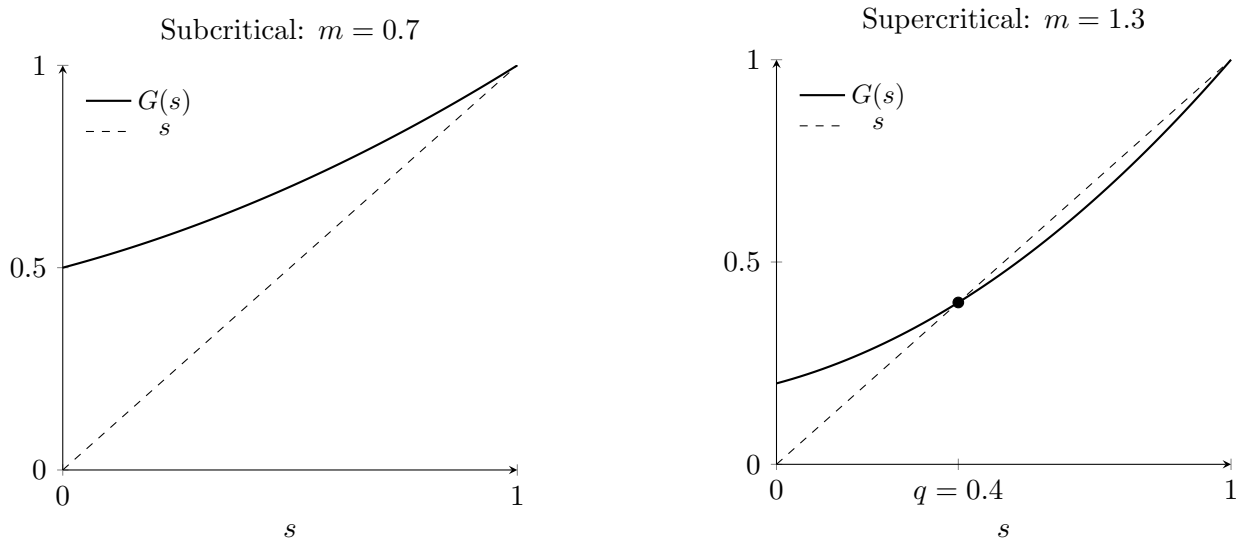


Figure 1: Offspring generating functions. Left: $p_0, p_1, p_2 = 0.5, 0.3, 0.2$; $G(s) > s$ on $[0, 1)$ and $q = 1$. Right: $p_0, p_1, p_2 = 0.2, 0.3, 0.5$; the graph crosses the diagonal at $q = 0.4$.

Critical, $m = 1$ (with $p_1 < 1$). Extinction is still certain, $q = 1$, even though $\mathbb{E}[Z_n] = 1$ for every n and, by (3), $\mathrm{Var}(Z_n) = n\sigma^2 \to \infty$. The population dies out with probability one, but on the rare event that it survives to generation n it tends to be large. Extinction is slow: if $\sigma^2 < \infty$, a classical result of Kolmogorov states that $\mathbb{P}\{Z_n > 0\} \sim 2/(n\sigma^2)$ as $n \to \infty$ (stated here without proof).

Supercritical, $m > 1$. The population survives forever with positive probability $1 - q > 0$. In fact (for $p_0 > 0$) every state $k \geq 1$ is transient, since from k the chain jumps to 0 in one step with probability $p_0^k > 0$. A Markov chain spends only finitely many steps in each transient state, so with probability one either $Z_n \to 0$ (extinction) or $Z_n \to \infty$. The next subsection shows that on survival the growth is exactly geometric at rate m.

2.6 Normalizing to obtain a martingale

Equation (2) says the population is expected to grow by the factor m in each generation. Dividing out this deterministic growth gives a martingale.

Theorem 2.6. *Let $W_n = Z_n/m^n$. Then (W_n) is a nonnegative martingale with respect to $(\mathcal{F}_n)$, with $\mathbb{E}[W_n] = 1$. Consequently, $W_n \to W$ almost surely for some random variable W with $0 \leq W < \infty$.*

Proof. W_n is $\mathcal{F}_n$-measurable and integrable with $\mathbb{E}[W_n] = m^{-n}\mathbb{E}[Z_n] = 1$. By (2),

$$\mathbb{E}[W_{n+1} \mid \mathcal{F}_n] = \frac{1}{m^{n+1}}\,\mathbb{E}[Z_{n+1} \mid \mathcal{F}_n] = \frac{mZ_n}{m^{n+1}} = W_n.$$

Finally, a nonnegative martingale converges almost surely to a finite limit (martingale convergence theorem). □

How much information the limit W carries depends on the regime.

- $m \le 1$**.** Since $Z_n = 0$ eventually, $W_n \to 0$ a.s., so $W = 0$ even though $\mathbb{E}[W_n] = 1$ for every n. The martingale converges almost surely but *not* in L^1; it is not uniformly integrable.

- $m > 1$ **and** $\sigma^2 < \infty$**.** By (3),

$$\mathbb{E}[W_n^2] = 1 + \frac{\mathrm{Var}(Z_n)}{m^{2n}} = 1 + \frac{\sigma^2\,(1 - m^{-n})}{m(m-1)} \;\le\; 1 + \frac{\sigma^2}{m(m-1)}.$$

  The martingale is bounded in L^2, so $W_n \to W$ in L^2 as well as almost surely. Hence

$$\mathbb{E}[W] = 1, \qquad \mathrm{Var}(W) = \frac{\sigma^2}{m(m-1)} = \frac{\sigma^2}{m^2 - m}.$$

  In particular $\mathbb{P}\{W > 0\} > 0$, and one can show that $\{W = 0\} = \{\text{extinction}\}$ up to a null set, so that $\mathbb{P}\{W = 0\} = q$. On the survival event,

$$Z_n \approx W m^n \quad \text{with } W > 0,$$

  i.e. the population grows geometrically at rate exactly m, with a random multiplicative constant W determined by the early generations.

3 The Poisson process

3.1 Counting processes

Definition 3.1 (Counting process)**.** A random process $N = (N_t : t \ge 0)$ is a *counting process* if $N_0 = 0$, each N_t takes values in $\mathbb{N}$, and every sample path $t \mapsto N_t$ is nondecreasing, right-continuous, and increases only by jumps of size 1.

Here N_t counts the number of events (arrivals, customers, particle detections, ...) in $(0, t]$, and $N_t - N_s$ is the number in $(s, t]$. Equivalently, a counting process is described by its *arrival times*

$$S_n = \min\{t \ge 0 : N_t \ge n\}, \qquad n \ge 1,$$

with $S_0 = 0$, and its *interarrival* (or *intercount*) *times* $T_n = S_n - S_{n-1}$. The two descriptions are related by

$$\{N_t \ge n\} = \{S_n \le t\}, \qquad N_t = \max\{n \ge 0 : S_n \le t\}. \tag{5}$$

A counting process has *independent increments* if for any $0 \le t_0 < t_1 < \cdots < t_k$ the increments $N_{t_1} - N_{t_0}, \ldots, N_{t_k} - N_{t_{k-1}}$ are independent, and *stationary increments* if the distribution of $N_{t+\epsilon} - N_t$ does not depend on t.

3.2 Three definitions

Fix a rate $\lambda > 0$. Recall that a random variable X is Poisson with mean m if $\mathbb{P}\{X = n\} = e^{-m} m^n / n!$ for $n \in \mathbb{N}$.

Definition 3.2 (Poisson process, version I: Poisson increments)**.** A counting process N is a Poisson process with rate λ if

(i) N has independent increments, and

(ii) for $0 \le s < t$, $N_t - N_s$ is Poisson with mean $\lambda(t - s)$.

Definition 3.3 (Poisson process, version II: i.i.d. exponential intercount times)**.** Let $T_1, T_2, \dots$ (the intercount times) be i.i.d. exponential random variables with parameter λ, i.e. $\mathbb{P}\{T_i > t\} = e^{-\lambda t}$. Put $S_0 = 0$, $S_n = T_1 + \cdots + T_n$, and $N_t = \max\{n : S_n \le t\}$. The resulting counting process N is a Poisson process with rate λ.

Definition 3.4 (Poisson process, version III: conditionally uniform arrivals)**.** A counting process N is a Poisson process with rate λ if for every $t > 0$,

(i) N_t is Poisson with mean λt, and

(ii) conditioned on $\{N_t = n\}$, $n \ge 1$, the arrival times $S_1 < S_2 < \cdots < S_n$ in $[0, t]$ are distributed as the order statistics of n i.i.d. random variables uniformly distributed on $[0, t]$. That is, given $N_t = n$, the n arrival points, taken in random order, are i.i.d. Uniform$[0, t]$, and $(S_1, \dots, S_n)$ has conditional density

$$f_{S_1,\dots,S_n | N_t = n}(s_1, \dots, s_n) = \frac{n!}{t^n}, \qquad 0 < s_1 < \cdots < s_n < t.$$

Version I is the most convenient for computation, version II is the most convenient for simulation and for constructing the process on $[0, \infty)$, and version III gives a second simulation method on a fixed window $[0, t]$ (draw N_t, then scatter that many uniform points), and shows that, given how many events occurred, the events carry no information about where in the window they occurred.

3.3 Equivalence of the definitions

Theorem 3.5. *Definitions 3.2, 3.3, and 3.4 define the same process: a counting process satisfies one of them if and only if it satisfies all three.*

We prove I $\Leftrightarrow$ II and I $\Leftrightarrow$ III.

I $\Rightarrow$ II. We find the joint density of the arrival times $S_1, \dots, S_n$. Fix $0 < t_1 < t_2 < \cdots < t_n$ and small $\epsilon > 0$ such that the intervals $(t_i - \epsilon, t_i]$ are disjoint. The event $\{S_i \in (t_i - \epsilon, t_i] \text{ for } i = 1, \dots, n\}$ is, up to events of probability $o(\epsilon^n)$, the event that there are no arrivals in the gaps $(0, t_1 - \epsilon]$, $(t_1, t_2 - \epsilon]$, $\dots$, $(t_{n-1}, t_n - \epsilon]$ and exactly one arrival in each $(t_i - \epsilon, t_i]$. By independent Poisson increments, the probability of the latter event is

$$e^{-\lambda(t_n - n\epsilon)} \prod_{i=1}^{n} \lambda\epsilon\, e^{-\lambda\epsilon} = \lambda^n e^{-\lambda t_n}\, \epsilon^n.$$

Dividing by ϵ^n and letting $\epsilon \downarrow 0$, the joint density of $(S_1, \dots, S_n)$ is

$$f_{S_1,\dots,S_n}(t_1, \dots, t_n) = \lambda^n e^{-\lambda t_n}, \qquad 0 < t_1 < \cdots < t_n.$$

The map $(u_1, \dots, u_n) \mapsto (t_1, \dots, t_n)$ with $t_k = u_1 + \cdots + u_k$ is linear with Jacobian determinant 1 and maps $(0, \infty)^n$ onto the region $\{0 < t_1 < \cdots < t_n\}$. Since $t_n = u_1 + \cdots + u_n$,

$$f_{T_1,\dots,T_n}(u_1, \dots, u_n) = \lambda^n e^{-\lambda(u_1 + \cdots + u_n)} = \prod_{i=1}^{n} \lambda e^{-\lambda u_i}, \qquad u_i > 0.$$

So $T_1, \dots, T_n$ are i.i.d. exponential with parameter λ, and by (5) N is recovered from the S_n exactly as in Definition 3.3.

II $\Rightarrow$ I. Reversing the change of variables above, the i.i.d. exponential interarrival times give the joint density $\lambda^m e^{-\lambda t_m}$ on $\{0 < t_1 < \cdots < t_m\}$ for $(S_1, \ldots, S_m)$, for every m.

Fix $0 = \tau_0 < \tau_1 < \cdots < \tau_k$ and integers $n_1, \ldots, n_k \geq 0$, and let $n = n_1 + \cdots + n_k$. The event

$$A = \{N_{\tau_j} - N_{\tau_{j-1}} = n_j,\ j = 1, \ldots, k\}$$

occurs exactly when, among $S_1, \ldots, S_n$, the first n_1 lie in $(\tau_0, \tau_1]$, the next n_2 lie in $(\tau_1, \tau_2]$, and so on, and $S_{n+1} > \tau_k$. Let $R \subset \mathbb{R}^n$ denote the set of ordered vectors $t_1 < \cdots < t_n$ with this interval pattern. Integrating the density of $(S_1, \ldots, S_{n+1})$, first over $t_{n+1} \in (\tau_k, \infty)$,

$$\mathbb{P}(A) = \int_R \int_{\tau_k}^{\infty} \lambda^{n+1} e^{-\lambda t_{n+1}}\, dt_{n+1}\, dt_1 \cdots dt_n = \lambda^n e^{-\lambda \tau_k}\ \mathrm{vol}(R).$$

The region R is a product over j of the sets of ordered n_j-tuples in $(\tau_{j-1}, \tau_j]$, and the set of ordered m-tuples in an interval of length ℓ has volume $\ell^m/m!$ (it is one of the $m!$ equal-volume orderings of the cube). Hence, writing $e^{-\lambda\tau_k} = \prod_j e^{-\lambda(\tau_j - \tau_{j-1})}$,

$$\mathbb{P}(A) = \prod_{j=1}^{k} e^{-\lambda(\tau_j - \tau_{j-1})}\, \frac{\big(\lambda(\tau_j - \tau_{j-1})\big)^{n_j}}{n_j!}.$$

The joint probability mass function factors into a product of Poisson pmfs with means $\lambda(\tau_j - \tau_{j-1})$. This proves at once that the increments are independent and that each is Poisson with the correct mean, which is Definition 3.2.

Remark 3.6. The case $k = 1$ can also be checked directly from (5): S_n has the Gamma (Erlang) density $\lambda^n s^{n-1} e^{-\lambda s}/(n-1)!$, and $\mathbb{P}\{N_t = n\} = \mathbb{P}\{S_n \leq t\} - \mathbb{P}\{S_{n+1} \leq t\} = e^{-\lambda t}(\lambda t)^n/n!$.

I $\Rightarrow$ III. Part (i) of Definition 3.4 is part of Definition 3.2. For part (ii), we have already shown (I $\Rightarrow$ II and the computation that follows it) that under Definition 3.2, for every m, the vector $(S_1, \ldots, S_m)$ has density $\lambda^m e^{-\lambda s_m}$ on $\{0 < s_1 < \cdots < s_m\}$. Fix $t > 0$, $n \geq 1$, and a set B of ordered vectors in $(0, t)^n$. Since $\{N_t = n\} = \{S_n \leq t < S_{n+1}\}$, integrating the density of $(S_1, \ldots, S_{n+1})$ over $s_{n+1} \in (t, \infty)$ gives

$$\mathbb{P}\{(S_1, \ldots, S_n) \in B,\ N_t = n\} = \int_B \int_t^{\infty} \lambda^{n+1} e^{-\lambda s_{n+1}}\, ds_{n+1}\, ds_1 \cdots ds_n = \int_B \lambda^n e^{-\lambda t}\, ds_1 \cdots ds_n.$$

Dividing by $\mathbb{P}\{N_t = n\} = e^{-\lambda t}(\lambda t)^n/n!$,

$$\mathbb{P}\{(S_1, \ldots, S_n) \in B \mid N_t = n\} = \int_B \frac{n!}{t^n}\, ds_1 \cdots ds_n,$$

which is the density of the order statistics of n i.i.d. Uniform$[0, t]$ random variables (the joint density t^{-n} of the unordered points, times the $n!$ orderings that map to the same ordered vector). Note that the conditional density does not depend on λ.

III $\Rightarrow$ I. Fix $0 = t_0 < t_1 < \cdots < t_k$ and integers $n_1, \ldots, n_k \geq 0$, let $n = n_1 + \cdots + n_k$, and set $t = t_k$ and $\Delta_j = t_j - t_{j-1}$. The increments $N_{t_j} - N_{t_{j-1}}$ count the arrivals in the subintervals $(t_{j-1}, t_j]$ of $[0, t]$, so they are determined by N_t and the arrival points in $[0, t]$. Conditioned on $N_t = n$, these

n points are i.i.d. uniform on $[0,t]$, so each falls in $(t_{j-1},t_j]$ independently with probability Δ_j/t, and the counts per subinterval are multinomial:

$$\mathbb{P}\{N_{t_j}-N_{t_{j-1}}=n_j,\ j=1,\dots,k \mid N_t=n\}=\frac{n!}{n_1!\cdots n_k!}\prod_{j=1}^{k}\Big(\frac{\Delta_j}{t}\Big)^{n_j}.$$

Multiplying by $\mathbb{P}\{N_t=n\}=e^{-\lambda t}(\lambda t)^n/n!$ and using $t=\sum_j\Delta_j$ and $n=\sum_j n_j$,

$$\mathbb{P}\{N_{t_j}-N_{t_{j-1}}=n_j,\ j=1,\dots,k\}=e^{-\lambda t}\frac{\lambda^n}{n_1!\cdots n_k!}\prod_{j=1}^{k}\Delta_j^{n_j}=\prod_{j=1}^{k}e^{-\lambda\Delta_j}\,\frac{(\lambda\Delta_j)^{n_j}}{n_j!}.$$

The joint pmf factors into Poisson pmfs with means $\lambda\Delta_j$, so the increments are independent and N_t-N_s is Poisson with mean $\lambda(t-s)$. This is Definition 3.2. □

3.4 Mean, correlation, and covariance functions

Proposition 3.7. *Let N be a Poisson process with rate λ. For $s,t\ge 0$,*

$$\begin{aligned}\mu_N(t)&=\mathbb{E}[N_t]=\lambda t, \qquad \operatorname{Var}(N_t)=\lambda t,\\ R_N(s,t)&=\mathbb{E}[N_sN_t]=\lambda\min(s,t)+\lambda^2 st,\\ C_N(s,t)&=\operatorname{Cov}(N_s,N_t)=\lambda\min(s,t).\end{aligned}$$

Proof. Mean and variance. If X is Poisson with mean m, then

$$\mathbb{E}[X]=\sum_{n\ge1}n\,\frac{e^{-m}m^n}{n!}=m\sum_{n\ge1}\frac{e^{-m}m^{n-1}}{(n-1)!}=m,\qquad \mathbb{E}[X(X-1)]=m^2\sum_{n\ge2}\frac{e^{-m}m^{n-2}}{(n-2)!}=m^2,$$

so $\mathbb{E}[X^2]=m^2+m$ and $\operatorname{Var}(X)=m$. With $m=\lambda t$ this gives $\mathbb{E}[N_t]=\operatorname{Var}(N_t)=\lambda t$ and $\mathbb{E}[N_t^2]=\lambda t+\lambda^2t^2$.

Correlation. Suppose $s\le t$ and write $N_t=N_s+(N_t-N_s)$. The increment N_t-N_s is independent of $N_s=N_s-N_0$ and has mean $\lambda(t-s)$, so

$$\mathbb{E}[N_sN_t]=\mathbb{E}[N_s^2]+\mathbb{E}[N_s]\,\mathbb{E}[N_t-N_s]=\lambda s+\lambda^2s^2+\lambda s\cdot\lambda(t-s)=\lambda s+\lambda^2 st.$$

By symmetry, for general s,t we get $R_N(s,t)=\lambda\min(s,t)+\lambda^2st$.

Covariance. $C_N(s,t)=R_N(s,t)-\mu_N(s)\mu_N(t)=\lambda\min(s,t)+\lambda^2st-\lambda^2st=\lambda\min(s,t)$. □

Remark 3.8. (a) The Poisson process is *not* wide-sense stationary: its mean λt grows with time and $C_N(s,t)$ is not a function of $t-s$. Its *increments* are stationary, however.

(b) The covariance function $\lambda\min(s,t)$ is the same as that of a Brownian motion with variance parameter $\sigma^2=\lambda$. Any process with independent increments, $X_0=0$, and $\operatorname{Var}(X_t-X_s)=\lambda(t-s)$ has this covariance, by the same computation. Second-order statistics alone cannot distinguish the (integer-valued, jumping) Poisson process from the (Gaussian, continuous) Brownian motion.

(c) The *compensated* process $M_t=N_t-\lambda t$ has zero mean, covariance $\lambda\min(s,t)$, and is a martingale: $\mathbb{E}[M_t\mid N_u,u\le s]=M_s+\mathbb{E}[N_t-N_s]-\lambda(t-s)=M_s$. This parallels the normalization Z_n/m^n of the Galton–Watson process: subtracting (or dividing out) the deterministic trend leaves a martingale.