Jana Kurrek

Lectures: Paquette, Elliot
Textbook: Dobrow, Robert

Stochastic Processes

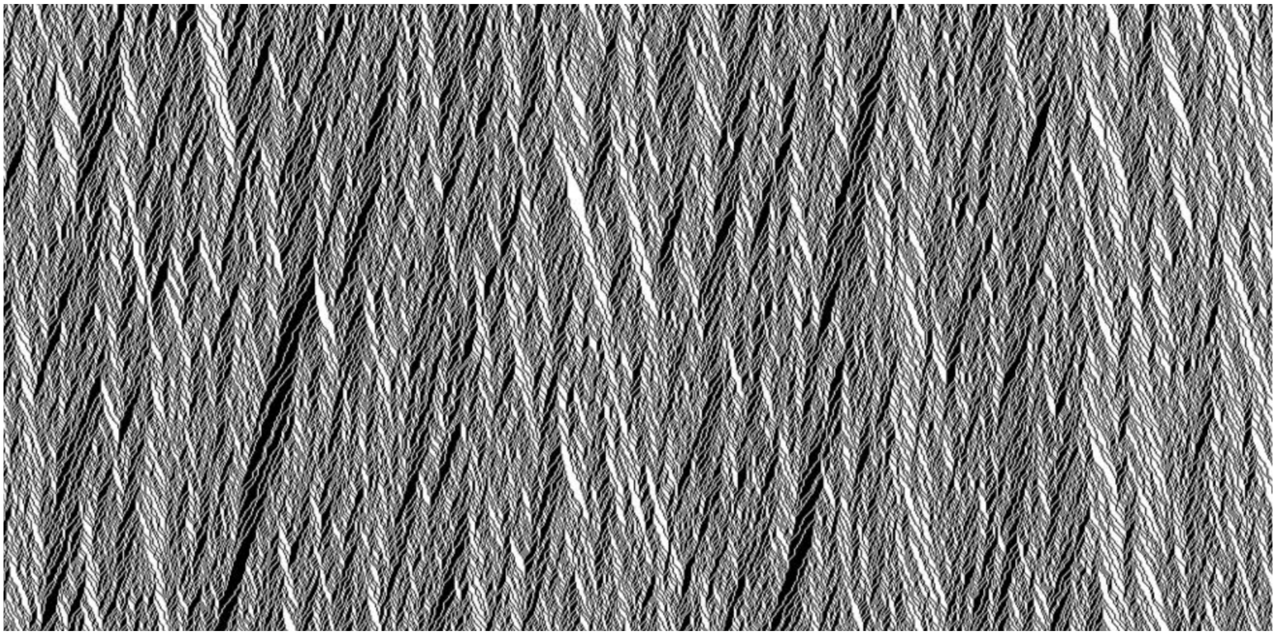


Figure 1. TASEP on $\mathbb{Z}/1000\mathbb{Z}$, with the initial distribution as the Bernoulli product measure.

Contents. Conditional expectation; Generating functions; Branching processes and random walk; Markov chains, transition matrices, classification of states, ergodic theorem; Birth and death processes; Queueing.

Contents

Axioms of Probability

Conditional Probability

Definition 1 (Conditional Probability). *The **conditional probability of** A **given** B, defined for $P(B) > 0$, is,*

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

Definition 2 (Independence). *Events A and B are **independent** if $P(A \mid B) = P(A)$. Equivalently, $P(A \cap B) = P(A)P(B)$.*

Definition 3 (Total Law of Probability). *Let $B_1, \cdots, B_k$ be a sequence of events that partition Ω. Then, for any event A,*

$$P(A) = P\left(\bigcup_{i \in [k]} (A \cap B_i)\right) = \sum_i^k P(A \cap B_i) = \sum_i^k P(A \mid B_i) P(B_i)$$

More generally, $P(A) = \mathbb{E}[P(A \mid X)]$.

Definition 4 (Bayes' Rule). *For events A and B*[1]*,*

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

Conditional Distributions

Definition 5 (Conditional Probability Mass Function). *The **conditional probability mass function of** Y **given** $X = x$ is,*

$$P(Y = y \mid X = x) = \frac{P(X = x, Y = y)}{P(X = x)}$$

Definition 6 (Conditional Probability Density Function). *The **conditional probability density function of** Y **given** $X = x$ is,*

$$f_{Y|X}(y|y) = \frac{f(x,y)}{f_X(x)}$$

Conditional Expectation

Definition 7 (Conditional Expectation). *The **conditional expectation of** Y **given** $X = x$, written $\mathbb{E}[Y \mid X = x](x)$, is a function of x,*

$$\mathbb{E}[Y \mid X = x](x) = \begin{cases} \sum_y y \cdot P(Y = y \mid X = x) & \Omega \textit{ is discrete} \\ \int_{-\infty}^{\infty} y \cdot f_{Y|X}(y \mid x) dy & \Omega \textit{ is continous} \end{cases}$$

Summary of Conditional Probability:

- Total Probability
$$P(A) = \mathbb{E}[P(A \mid X)]$$
- Total Expectation
$$\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y \mid X]]$$
- Total Conditional Expectation
$$P(Y \mid A) = \mathbb{E}[P(Y \mid X, A) \mid A]$$
- Total Conditional Probability
$$\mathbb{E}[Y \mid A] = \mathbb{E}[\mathbb{E}[Y \mid X, A] \mid A]$$

---

[1] With the Total Law of Probability,

$$P(B_i \mid A) = \frac{P(A \mid B_i) P(B_i)}{\sum_j P(A \mid B_j) P(B_j)}$$

Definition 8. *The* ***conditional expectation of*** *Y* ***given*** *A is,*

$$\begin{aligned} E(Y \mid A) &= \frac{1}{P(A)} \sum_y y P(\{Y = y\} \cap A) \\ &= \sum_y y P(Y = y \mid A) \end{aligned}$$

for the discrete case.

Definition 9 (Law of Total Expectation)**.** *If $A_1, \cdots, A_k$ partitions Ω and Y is a random variable, then the* ***law of total expectation*** *states that,*

$$\mathbb{E}[Y] = \sum_{i=1}^{k} \mathbb{E}[Y \mid A_i] P(A_i)$$

More generally, $\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y \mid X]]$.

Proof. For the discrete case,

$$\begin{aligned} \mathbb{E}[\mathbb{E}[Y \mid X]] &= \sum_x \mathbb{E}[Y \mid X = x] \cdot P(X = x) \\ &= \sum_x \left(\sum_y y P(Y = y \mid X = x) \right) P(X = x) \\ &= \sum_y y \sum_x P(Y = y \mid X = x) \cdot P(X = x) \\ &= \sum_y y \sum_x P(X = x, Y = y) \\ &= \sum_y y \cdot P(Y = y) \\ &= \mathbb{E}(Y) \end{aligned}$$

□

Time-Homogeneous Markov Chains

Finite State, Time-Homogeneous Chains

Definition 10 (Finite State Stochastic Process)**.** *A* ***finite state stochastic process*** *$(X_n)_{n \geq 0}$ has time steps in $\mathbb{N}$ and values in $S = [N - 1]$.*

Definition 11 (Markov Property)**.** *The* ***Markov property*** *claims that for every $n \in \mathbb{N}$ and every sequence of states $(i_0, i_1, \cdots)$ where $i_j \in S$, the behavior of a system depends only on the previous state,*

$$P(X_n = i_n \mid X_0 = i_0, \cdots, X_{n-1} = i_{n-1}) = P(X_n = i_n \mid X_{n-1} = i_{n-1})$$

Definition 12 (Time Homogeneity)**.** *A Markov chain is* ***time-homogeneous*** *if the probabilities in Definition 11 do not depend on n,*

$$P(X_n = i_n \mid X_{n-1} = i_{n-1}) = P(X_1 = i_1 \mid X_0 = i_0) \quad (n \in \mathbb{N})$$

Definition 13 (Transition Matrix)**.** *The **transition matrix** $\boldsymbol{P}$ for a time-homogeneous Markov chain is the $N \times N$ matrix whose (i,j)th entry $\boldsymbol{P}_{ij}$ is the one-step transition probability $p(i,j) = P(X_1 = j \mid X_0 = i)$.*

Remark 14. *The transition matrix $\boldsymbol{P}$ is stochastic, that is,*

- *(Non-Negative Entries)* $0 \leq \boldsymbol{P}_{ij} \leq 1$ *for* $1 \leq i,j \leq N$
- *(Row Sum Equal to 1)* $\sum_{j=1}^{N} \boldsymbol{P}_{ij} = 1$ *for* $1 \leq i \leq N$

Example 1: Biased Coin Flips

Let $(X_n)_{n\geq 0}$ denote a sequence of coin flips where,

$$P(X_{n+1} = H \mid X_n) = \begin{cases} 0.51 \text{ if } X_n = H \\ 0.49 \text{ if } X_n = T \end{cases}$$

and,

$$P(X_{n+1} = T \mid X_n) = \begin{cases} 0.51 \text{ if } X_n = T \\ 0.49 \text{ if } X_n = H \end{cases}$$

Then,

$$\boldsymbol{P} = \begin{pmatrix} 0.51 & 0.49 \\ 0.49 & 0.51 \end{pmatrix} = \begin{pmatrix} P_{HH} & P_{HT} \\ P_{TH} & P_{TT} \end{pmatrix}$$

Transition Probabilities

Definition 15 (Probability Distribution Vector)**.** *The **distribution** of a discrete random variable X is the vector $\vec{\phi}$ if,*

$$\phi_j = P(X = j) \quad \forall j \in \mathbb{N}$$

Definition 16 (Initial Distribution Vector)**.** *The **initial probability distribution** of a Markov chain $(X_n)_{n\geq 0}$ is the distribution $\vec{\phi}$ of X_0.*

Theorem 17 (Transition Probabilities)**.** *The n-step transition probability $p_n(i,j) = P(X_n = i \mid X_0 = j)$ is the (i,j)th entry in the matrix $\boldsymbol{P}^n$.*

Proof. The base case is trivially true for $n = 1$. Assume the statement holds for a given n. Using the properties of matrix multiplication and the Law of Total Conditional Expectation,

$$\begin{aligned} P(X_{n+1} = j \mid X_0 = i) &= \sum_{k\in S} P(X_n = k \mid X_0 = i) \cdot P(X_{n+1} = j \mid X_n = k) \\ &= \sum_{k\in S} p_n(i,k) \cdot p(k,j) \\ &= \boldsymbol{P}^n \boldsymbol{P} \\ &= \boldsymbol{P}^{n+1} \end{aligned}$$

□

Theorem 18 (Chapman-Kolmogorov). *Let $x, y, z \in S$ and $m, n \in (0, \infty)$.*

$$\begin{aligned} p_{m+n}(x,y) &= P(X_{m+n} = y \mid X_0 = x) \\ &= \sum_{z \in S} P(X_{m+n} = y, X_m = z \mid X_0 = x) \\ &= \sum_{z \in S} p_m(x,z) \cdot p_n(z,y) \end{aligned}$$

The probabilistic interpretation for **Chapman-Kolmogorov** is that transitioning from x to y in $m+n$ steps is equivalent to transitioning from i to k in m steps and then moving from k to j in the remaining n steps.

Definition 19 (Distribution of X_n). *The* ***distribution of*** $(X_n)_{n \geq 0}$ *is,*

$$\phi \cdot \boldsymbol{P}^n \quad \text{i.e., } P(X_n = j) = (\phi \cdot \boldsymbol{P}^n)_j \quad \forall j \in \mathbb{N}$$

where $\boldsymbol{P}$ is the transition matrix of (X_n) and ϕ is the initial distribution.

Stationary Distributions

Definition 20 (Limiting Distribution Vector). *A* ***limiting distribution*** *for a time-homogeneous Markov chain $(X_n)_{n \geq 0}$ is a distribution $\vec{\pi}$ so that,*

$$\lim_{n \to \infty} p_n(i,j) = \pi_j$$

The definition of a limiting distribution is equivalent to,

- $\lim_{n \to \infty} P(X_n = j) = \pi_j$
- $\lim_{n \to \infty} \vec{\phi} \cdot \boldsymbol{P}^n = \vec{\pi}$, *where $\vec{\phi}$ is the initial distribution*
- $\lim_{n \to \infty} \boldsymbol{P}^n = \boldsymbol{V}$, *where $\boldsymbol{V}$ is a stochastic matrix whose rows are $\vec{\pi}$*

Definition 21 (Occupation Time). *The* ***occupation time*** *for a time-homogeneous Markov chain $(X_n)_{n \geq 0}$ from an initial state i is,*

$$\lim_{n \to \infty} \mathbb{E}\left[\frac{1}{n} \sum_{k=0}^{n-1} \mathbb{1}_{X_k = i} \mid X_0 = i\right]$$

which represents the long-term expected proportion of time spent visiting j.

Remark 22. *The limiting distribution gives the occupation time of (X_n),*

$$\begin{aligned} \lim_{n \to \infty} \mathbb{E}\left[\frac{1}{n} \sum_{k=0}^{n-1} \mathbb{1}_{\{X_k = j\}} \mid X_0 = i\right] &= \lim_{n \to \infty} \sum_{k=0}^{n-1} \mathbb{E}[\mathbb{1}_{X_k = j} \mid X_0 = i] \\ &= \lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} P(X_k = j \mid X_0 = i) \\ &= \lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} p_n(i,j) \\ &= \lim_{n \to \infty} \boldsymbol{P}^n_{ij} \\ &= \pi_j \text{ by Cesaro's Average} \end{aligned}$$

Definition 23 (Stationary Distribution Vector). *A limiting probability vector $\vec{\pi}$ is a* ***stationary probability distribution*** *for* $\boldsymbol{P}$ *if,*

$$\vec{\pi} = \vec{\pi} \cdot \boldsymbol{P}$$

Lemma 24. *The limiting distribution $\vec{\pi}$ of (X_n) is stationary.*

Proof. Assume that $\vec{\pi}$ is the limiting distribution. We need to show that $\vec{\pi} = \vec{\pi} \cdot \boldsymbol{P}$. For any initial distribution $\vec{\phi}$,

$$\vec{\pi} = \lim_{n\to\infty} \vec{\phi} \cdot \boldsymbol{P}^n = \lim_{n\to\infty} \vec{\phi} \cdot (\boldsymbol{P}^{n-1} \cdot \boldsymbol{P}) = \left(\lim_{n\to\infty} \vec{\phi} \cdot \boldsymbol{P}^{n-1} \right) \cdot \boldsymbol{P} = \vec{\pi} \cdot \boldsymbol{P}$$

☐

Note on Stationary Distributions: The converse of Lemma 24 is not true. Stationary distributions are not necessarily limiting distributions. For example, the chain with,

$$\boldsymbol{P} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

has every probability vector as a stationary distribution since,

$$\vec{\pi} \cdot \boldsymbol{P} = \vec{\pi} \quad \forall \vec{\pi}$$

$\boldsymbol{P} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$	Unique Stationary
$\boldsymbol{P} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	Multiple Stationary
$\boldsymbol{P} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$	No Limiting
$p(x, x+1) = 1 \quad (S = \mathbb{N})$	No Stationary

Example 2: Random Walk on a Weighted Graph

Let G be a weighted graph with edge weight function $w(i, j)$. For a random walk on G, the stationary distribution $\vec{\pi}$ is proportial to the sum of the edge weights incident to each vertex,

$$\pi_v = \frac{w(v)}{\sum_z w(z)} \quad \forall v \in V(G) \quad \text{where} \quad w(v) = \sum_{z\sim v} w(v, z)$$

is the sum of the edge weights on all edges incident to v.

Consider the random walk with transition matrix,

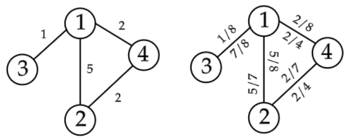
$$\boldsymbol{P} = \begin{pmatrix} 0 & \frac{5}{8} & \frac{1}{8} & \frac{2}{8} \\ \frac{5}{7} & 0 & 0 & \frac{2}{7} \\ 1 & 0 & 0 & 0 \\ \frac{2}{4} & \frac{2}{4} & 0 & 0 \end{pmatrix}$$

Let $w(i) = \sum_{j=1}^{4} w(i, j)$. Then,

$$\begin{aligned} \vec{\pi} &= \begin{pmatrix} \frac{w(1)}{\sum w(1)} & \frac{w(2)}{\sum w(2)} & \frac{w(3)}{\sum w(3)} & \frac{w(4)}{\sum w(4)} \end{pmatrix} \\ &= \begin{pmatrix} \frac{8}{20} & \frac{7}{20} & \frac{1}{20} & \frac{4}{20} \end{pmatrix} \end{aligned}$$

satisfies $\vec{\pi} \cdot \boldsymbol{P} = \vec{\pi}$.

Random Walk on a Weighted Graph:



Example 3: Simple Random Walk on a Graph

For a simple random walk on a non-weighted graph,

$$w(i, j) = 1 \quad \forall i, j \in V(G) \quad \text{and} \quad w(v) = \deg(v)$$

If $|E(G)|$ is the number of edges in the graph, this gives,

$$\pi_v = \frac{\deg(v)}{\sum_z \deg(z)} = \frac{\deg(v)}{2|E(G)|}$$

Regular Chains

Definition 25 (Regularity). *A transition matrix $\boldsymbol{P}$ is* ***regular*** *if and only if there exists $n \in \mathbb{N}$ so that every entry of $\boldsymbol{P}^n$ is positive.*

Theorem 26. *If the matrix $\boldsymbol{P}^n$ is regular, then $\boldsymbol{P}^m$ is regular ($m > n$).*

Proof. Linear combinations of positive numbers are positive, so the result follows by the definition of matrix multiplication,

$$\boldsymbol{P}_{ij}^{n+1} = \sum_{\ell \in S} \boldsymbol{P}_{i\ell}^{n} \cdot \boldsymbol{P}_{\ell j} > 0$$

since $\boldsymbol{P}_{i\ell}^{n} > 0$ and at least one $\boldsymbol{P}_{\ell j}$ is positive[2]. ☐

Theorem 27. *A stochastic matrix $\boldsymbol{P}$ has an eigenvalue $\lambda^* = 1$. All other eigenvalues λ of $\boldsymbol{P}$ satisfy $|\lambda| \leq 1$, with strict inequality if P is regular.*

Proof. Let $\boldsymbol{P}$ be a $k \times k$ stochstic matrix. The rows of $\boldsymbol{P}$ sum to 1 by definition, so $\boldsymbol{P} \cdot \vec{1} = \vec{1}$ and $\lambda^* = 1$ is a right eigenvalue of $\boldsymbol{P}$. Suppose that $\vec{z}$ is the eigenvector corresponding to any other eigenvalue λ of $\boldsymbol{P}$. Let $|z_m| = \max_{1 \leq i \leq k} |z_i|$ be the component of $\vec{z}$ of maximum absolute value. Then,

$$|\lambda| \cdot |z_m| = |\lambda z_m| = |(\boldsymbol{P} \cdot \vec{z})_m| = |\sum_{i=1}^{k} \boldsymbol{P}_{mi} z_i| \leq |z_m| \sum_{i=1}^{k} \boldsymbol{P}_{mi} = |z_m|$$

and consequently $|\lambda| \leq 1$.

Assume that $\boldsymbol{P}$ is regular. Then $\boldsymbol{P}^n > 0$ for some $n > 0$. $\boldsymbol{P}$ is a stochastic matrix, and it was shown above that $\boldsymbol{P}$ has an eigenvalue $\lambda^* = 1$. Moreover, all other eigenvalues λ of $\boldsymbol{P}$ satisfy $|\lambda| \leq 1$. We want to show that the inequality is strict. If λ is an eigenvalue of $\boldsymbol{P}$, then λ^n is an eigenvalue of $\boldsymbol{P}^n$. Let $\vec{x}$ be its corresponding eigenvalue, with $|x_m| = \max_{1 \leq i \leq k} |x_i|$. Then,

$$|\lambda|^n \cdot |x_m| = |(\boldsymbol{P}^n \cdot \vec{z})_m| = |\sum_{i=1}^{k} \boldsymbol{P}_{mi}^{n} x_i| \leq |x_m| \sum_{i=1}^{k} \boldsymbol{P}_{mi}^{n} = |x_m|$$

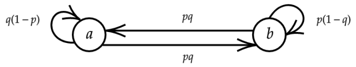
Since the entries of $\boldsymbol{P}^n$ are positive, the last inequality only holds if $|x_1| = \cdots = |x_k|$. Similarly, the first inequality only holds if $x_1 = \cdots = x_m$. But the constant vector whose components are the same is an eigenvector associated with the eigenvalue 1. Hence, if $\lambda \neq 1$, one of the inequalities must be strict. Thus, $|\lambda|^n < 1$. ☐

Theorem 28. *Every finite state time-homogeneous Markov chain $(X_n)_{n \geq 0}$ has a stationary distribution $\vec{\pi}$. Moreover, the stationary distribution(s) of (X_n) are in bijective correspondance with the left 1-eigenvectors of $\boldsymbol{P}$*[3].

The two-state Markov chain with,

$$\boldsymbol{P} = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$$

can be expressed as,



Therefore, its stationary distribution is,

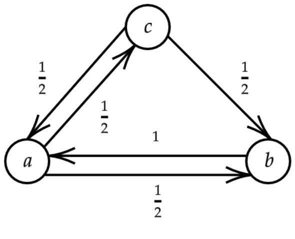
$$\pi = \left(\frac{q}{p+q}, \frac{p}{p+q} \right)$$

Example of Regularity:
The following matrix is regular,

$$\boldsymbol{P} = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1 & 0 & 0 \\ 1/2 & 1/2 & 0 \end{pmatrix}$$

because $\boldsymbol{P}^4$ is positive,

$$\boldsymbol{P}^4 = \begin{pmatrix} 9/16 & 5/16 & 1/8 \\ 1/4 & 3/8 & 3/8 \\ 1/2 & 5/16 & 3/16 \end{pmatrix}$$



---

[2] Otherwise $\boldsymbol{P}^n$ has a zero column.

[3] We require that the eigenvectors are non-negative with components that sum to 1.

Proof. If $\boldsymbol{P}$ is a stochastic matrix, then $\boldsymbol{P}$ has a right eigenvalue $\lambda^* = 1$ (see Theorem 27). Since $\det(\boldsymbol{P}^T) = \det(\boldsymbol{P})$, the left and right eigenvalues of $\boldsymbol{P}$ are equal. Hence, $\boldsymbol{P}$ has at least one left eigenvector $\vec{\pi}$ for the eigenvalue 1. Normalizing this eigenvector to sum to 1 gives[4],

$$\boldsymbol{P} \cdot \vec{\pi} = 1 \cdot \vec{\pi} \quad \text{i.e., } \boldsymbol{P} \cdot \vec{\pi} = \vec{\pi}$$

The proof that the stationary distribution(s) of (X_n) are in bijective correspondance with the left 1-eigenvectors of $\boldsymbol{P}$ is similar. □

Corollary 29. *If (X_n) has a unique stationary distribution, then the distribution is a left eigenvector of $\boldsymbol{P}$ corresponding to $\lambda^* = 1$.*

Theorem 30 (Perron-Frobenius)**.** *Let $\boldsymbol{M}$ be a $k \times k$ positive matrix. Then,*

- *There exists $\lambda^* \in \mathbb{R}^+$ (called the Perron-Frobenius eigenvalue) which is an eigenvalue of M. $|\lambda| < \lambda^*$ for all other eigenvalues λ of $\boldsymbol{M}$.*
- *The eigenspace of eigenvectors associated with λ^* is one-dimensional*

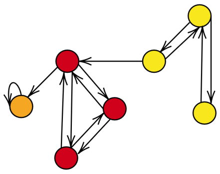
Proof. The proof of the Perron–Frobenius theorem can be found in many linear algebra textbooks, including Horn and Johnson (1990). □

Theorem 31 (Limit Theorem for Finite Regular Chains)**.** *If $(X_n)_{n\geq 0}$ is a finite state time-homogeneous Markov chain and $\boldsymbol{P}$ is regular, then there is a unique, positive, stationary distribution $\vec{\pi} > 0$ such that,*

$$\lim_{n\to\infty} P^n = \boldsymbol{V}$$

where $\boldsymbol{V}$ is a matrix with all rows equal to $\vec{\pi}$.

Example of Communication Classes: G has 3 communication classes,



Classification of States

Definition 32 (Communication)**.** *Two states $i, j \in S$ of a Markov chain **communicate**, written $i \leftrightarrow j$, if there exist $m, n \in \mathbb{N}$ such that,*

$$p_m(i,j) > 0 \quad \textit{and} \quad p_n(j,i) > 0$$

Equivalently, two states communicate if and only if each state has a positive probability of eventually being reached by a chain starting in the other state.

Theorem 33. *The relation $\leftrightarrow$ is an equivalence relation on the state space.*

Proof. The relation $\leftrightarrow$ is reflexive, symmetric, and transitive.

- (Reflexivity) $i \leftrightarrow i$ since $p_0(i,i) = 1 > 0$
- (Symmetry) $i \leftrightarrow j \implies j \leftrightarrow i$ by definition

---

[4] The vector with $|\pi|_i = |\pi_i|$ is still an eigenvector with eigenvalue 1.

- (Transitivity) $i \leftrightarrow j$ and $j \leftrightarrow i \implies i \leftrightarrow k$ since,

$$\begin{aligned} p_{m_1+m_2}(i,k) &= P(X_{m_1+m_2} = k \mid X_0 = i) \\ &\geq P(X_{m_1+m_2} = k, X_{m_1} = j \mid X_0 = i) \\ &= P(X_{m_1} = j \mid X_0 = i) \cdot P(X_{m_1+m_2} = k \mid X_{m_1} = j) \\ &= p_{m_1}(i,j) p_{m_2}(j,k) \\ &> 0 \end{aligned}$$

□

Definition 34 (Irreducibility). *The relation $\leftrightarrow$ partitions the state space into disjoint sets called **communication classes**. If there is only one communication class, then the chain is called irreducible.*

Definition 35 (Hitting Time). *Let $(X_n)_{n\geq 0}$ be a Markov chain with state space S. The **hitting time** ("first passage time") of $A \subseteq S$ when $X_0 = x$ is,*

$$\tau_A = \inf\{n \geq 0 \mid X_n \in A\}$$

Definition 36 (First Return Time). *Let $(X_n)_{n\geq 0}$ be a Markov chain with state space S. The **first return time** is a variant of hitting time,*

$$\tau_x^+ = \inf\{n \geq 1 \mid X_n = i\} \quad \textit{assuming} \quad X_0 = x$$

Definition 37 (Expected Number of Visits). *Let $(X_n)_{n\geq 0}$ be a Markov chain with state space S. The **expected number of visits** to $i \in S$ is,*

$$\sum_{n=0}^{\infty} p_n(i,i) \quad \textit{assuming} \quad X_0 = i$$

Proof. Let N_i be the random variable giving the total number of visits to $i \in S$, including the initial visit. We can write,

$$N_i = \sum_{n=0}^{\infty} \mathbb{1}_{\{X_n = i\}}$$

where the expectation of N_i when $X_0 = i$ is,

$$\mathbb{E}[N_i] = \mathbb{E}\left[\sum_{n=0}^{\infty} \mathbb{1}_{\{X_n=i\}}\right] = \sum_{n=0}^{\infty} P(X_n = i) = \sum_{n=0}^{\infty} p_n(i,i)$$

□

Theorem 38 (Limit Theorem for Finite Irreducible Chains). *If $(X_n)_{n\geq 0}$ is a finite state time-homogeneous Markov chain and $\boldsymbol{P}$ is irreducible, then,*

$$\mu_j = \mathbb{E}[\tau_j^+ \mid X_0 = j] < \infty \quad \textit{for all } i \in S$$

and there exists a unique, positive, stationary distribution $\vec{\pi} > 0$ such that,

$$\pi_j = \frac{1}{\mathbb{E}[\mu_j]}, \textit{ for all } j \in S$$

Furthermore,

$$\pi_j = \lim_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n-1} \boldsymbol{P}_{ij}^m \quad \textit{for all } j \in S$$

First Step Analysis

Definition 39 (First-Step Analysis). ***First-step analysis*** *is the process of conditioning on the first step of the chain and using the law of total expectation to find the expected return time,* $\mathbb{E}[\tau_j^+ \mid X_0 = j]$.

Remark 40. *If* $(X_n)_{n\geq 0}$ *is irreducible, then the expected return time can also be found by taking the reciprocal of the stationary probability* π.

Example 4: First-Step Analysis

Consider the Markov chain with transition matrix,

$$\begin{array}{cc} & \begin{array}{ccc} a & b & c \end{array} \\ \boldsymbol{P} = & \left(\begin{array}{ccc} 0 & 1 & 0 \\ 1/2 & 1 & 1/2 \\ 1/3 & 1/3 & 1/3 \end{array}\right) \begin{array}{c} a \\ b \\ c \end{array} \end{array}$$

Define $e_x := \mathbb{E}[\tau_j^+ \mid X_0 = x]$ for $x \in \{a, b, c\}$. Thus, e_a is the desired expected return time, and e_b, e_c are the first hitting times to a for the chain started in b and c.

$$\begin{aligned} e_a &= 1 + e_b \\ e_b &= \frac{1}{2} + \frac{1}{2}(1 + e_c) \\ e_c &= \frac{1}{3} + \frac{1}{3}(1 + e_b) + \frac{1}{3}(1 + e_c) \end{aligned}$$

Solving these equations gives,

$$e_c = \frac{8}{3} \quad e_b = \frac{7}{3} \quad e_a = \frac{10}{3}$$

Recurrence and Transience

Definition 41 (Recurrent). *A state* $i \in S$ *is* ***recurrent*** *if a Markov chain starting at* i *will return to* i *infinitely often, with probability 1.*

Definition 42 (Transient). *A state* $i \in S$ *is* ***transient*** *if a Markov chain starting at* i *will return to* i *only finitely often, with probability 1.*

Theorem 43. *Recurrence and transience are class properties,*

$$\begin{aligned} &\text{If } i \in S \text{ is recurrent and } j \leftrightarrow i \implies j \text{ is recurrent} \\ &\text{If } i \in S \text{ is transient and } j \leftrightarrow i \implies j \text{ is transient} \end{aligned}$$

Proof. It suffices to show that if $i \in S$ is transient and $j \leftrightarrow i$, then j is transient. Since $j \leftrightarrow i$, there exist $s, r \geq 0$ such that $p_s(i, j) > 0$ and

$p_r(j,i) > 0$. For all $n \in \mathbb{N}$, it holds that,

$$p_{n+r+s}(i,i) \geq p_s(i,j) \cdot p_n(j,j) \cdot p_r(j,i)$$

by Chapman-Kolmogorov. Therefore,

$$\begin{aligned}\sum_{n=1}^{\infty} p_n(j,j) &\leq \frac{1}{p_s(i,j) \cdot p_r(j,i)} \sum_{n=1}^{\infty} p_{n+r+s}(i,i) \text{ expanding the expectation} \\ &\leq \frac{1}{p_s(i,j) \cdot p_r(j,i)} \sum_{n=1}^{\infty} p_n(i,i) \\ &< \infty \quad \text{since } i \text{ is transient}\end{aligned}$$

It follows that j is transient. Hence, if one state of a communication class is transient, all states in that class are transient.

Conversely, if one state is recurrent, then the others must be recurrent. By contradiction, if the communication class contains a transient state then by what was just proven all the states are transient. □

Theorem 44. *Every state in a finite, irreducible Markov chain is recurrent.*

Proof. Every pair $i, j \in S$ belongs to the same communication class, and that class has finitely many elements. By definition, there is positive probability of reaching j from i since $i \leftrightarrow j$. If i is visited infinitely often then we get this chance of visiting j infinitely often. If an event has a positive probability of occuring, and we get an infinite number of trials, then it will occur an infinite number of times. □

Theorem 45. *Let $(X_n)_{n\geq 0}$ be an irreducible Markov chain. Then,*

$$\mathbb{E}[N_i] = \sum_{n=0}^{\infty} p_n(i,i) = \frac{1}{1 - P(\tau_i^+ < \infty)} = \frac{1}{P(\tau_i^+ = \infty)} = (I - \boldsymbol{P})_{ii}^{-1}$$

Moreover,

$$\mathbb{E}[N_i] = \begin{cases} \textit{finite} & \iff i \textit{ is transient} \\ \textit{infinite} & \iff i \textit{ is recurrent} \end{cases}$$

Recall:
Let $\boldsymbol{A}$ be a square matrix with the property that $\boldsymbol{A}^n \to 0$, as $n \to \infty$. Then, $\sum_{n=0}^{\infty} A^n = (I - A)^{-1}$. This gives the matrix analog of the sum of a geometric series of real numbers.

Proof. Assume that $\mathbb{E}[N_i] < \infty$. Let R_i be the number of returns to a state i. Define a sequence $(\tau_i^{(n)})_{n\geq 0}$ by,

$$\tau_i^{(n)} = \begin{cases} \inf\{n \geq 1 \mid X_n^* = i\} & R_i \geq n \\ \infty & \text{otherwise} \end{cases}$$

where (X_n^*) is the process (X_n) started at $\tau_i^{(n-1)}$. $N_i = \sum_{n=1}^{\infty} \mathbb{1}_{\{X_n = i\}}$ is 1 more than the number of returns R_i, so,

$$N_i = 1 + R_i = 1 + \sum_{n=1}^{\infty} \mathbb{1}_{\{\tau_i^{(n)} < \infty\}}$$

and $\tau_i^{(n)} = \infty$ if and only if X_n visits i fewer than n times. Now, $P(\tau_i^{(n)} < \infty) = [P(\tau_i^{(1)} = \infty)]^n$ by time homogeneity $(\star)$. Therefore,

$$\begin{aligned}
\mathbb{E}[N_i] &= \mathbb{E}\left[1 + \sum_{n=1}^{\infty} \mathbb{1}_{\{\tau_i^{(n)} < \infty\}}\right] \\
&= \mathbb{E}\left[\sum_{n=0}^{\infty} \mathbb{1}_{\{\tau_i^{(n)} < \infty\}}\right] \\
&= \sum_{n=0}^{\infty} \mathbb{E}\left[\mathbb{1}_{\{\tau_i^{(n)} < \infty\}}\right] \quad \text{by Linearity of Expectation} \\
&= \sum_{n=0}^{\infty} P(\tau_i^{(n)} < \infty) \\
&= \sum_{n=0}^{\infty} P(\tau_i^{(1)} < \infty) \\
&= \sum_{n=0}^{\infty} [P(\tau_i^{(1)} < \infty)]^n \quad (\star) \\
&= \frac{1}{1 - P(\tau_i^{+} < \infty)} \quad \text{by definition of a geometric series}
\end{aligned}$$

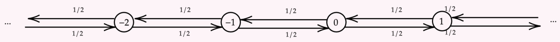
Thus, $\mathbb{E}[N_i] = \frac{1}{1-P(\tau_i^{+} < \infty)} = \begin{cases} \text{finite} & \iff i \text{ is transient} \\ \text{infinite} & \iff i \text{ is recurrent} \end{cases}$ □

Stirling's Formula states that as $n \to \infty$,

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$$

Example 5: Simple Symmetric Random Walk on $\mathbb{Z}$

The **simple symmetric random walk on $\mathbb{Z}$** is recurrent,



$$\begin{aligned}
\mathbb{E}[N_i] &= \sum_{n \geq 0} p_n(0,0) \\
&= \sum_{n \geq 0} p_{2n}(0,0) \\
&= \sum_{n \geq 0} \binom{2n}{n} \cdot \frac{1}{2^{2n}} \\
&\geq \sum_{n \geq 1} \frac{1}{\sqrt{4n}} \quad \text{by Stirling's Formula} \\
&= \infty
\end{aligned}$$

Definition 46 (Canonical Decomposition)**.** *The **canonical decomposition** of the state space S of a finite Markov chain is a separation of S,*

$$S = T \cup R_1 \cup \cdots \cup R_m$$

where $R_1, \cdots, R_m$ are the communication classes of recurrent states and T is the set of all transient states. $\boldsymbol{P}$ has the block matrix form,

$$\boldsymbol{P} = \begin{array}{cc} \begin{array}{ccccc} T & R_1 & R_2 & \cdots & R_m \end{array} & \\ \begin{pmatrix} * & * & * & \cdots & * \\ 0 & \boldsymbol{P}_1 & 0 & \cdots & 0 \\ 0 & 0 & \boldsymbol{P}_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \boldsymbol{P}_m \end{pmatrix} & \begin{array}{c} T \\ R_1 \\ R_2 \\ \vdots \\ R_m \end{array} \end{array}$$

where each square stochastic matrix $\boldsymbol{P}_i$ corresponds to a closed recurrent communication class which is irreducible with a restricted state space[5]*.*

Remark 47. *The block matrix form facilitates taking matrix powers,*

$$\lim_{n\to\infty} \boldsymbol{P}^n = \begin{array}{cc} \begin{array}{ccccc} T & R_1 & R_2 & \cdots & R_m \end{array} & \\ \begin{pmatrix} * & * & * & \cdots & * \\ 0 & \lim_{n\to\infty} \boldsymbol{P}_1^n & 0 & \cdots & 0 \\ 0 & 0 & \lim_{n\to\infty} \boldsymbol{P}_2^n & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \lim_{n\to\infty} \boldsymbol{P}_m^n \end{pmatrix} & \begin{array}{c} T \\ R_1 \\ R_2 \\ \vdots \\ R_m \end{array} \end{array}$$

Corollary 48. *For every recurrent class R_i, there is a stationary distribution $\vec{\pi}$ so that $\pi_i > 0$ if and only if $i \in R_i$.*

Corollary 49. *The dimension of the eigenspace of $\boldsymbol{P}$ for the eigenvalue 1 is the number of recurrent classes in the Markov chain.*

Periodicity

Definition 50 (Period)**.** *The* ***period*** *of a state i, $d = d(i)$ is,*

$$gcd(J_i) \text{ where } J_i = \{n \geq 0 \mid p_n(i,i) > 0\}$$

If $d(i) = 1$, state i is called aperiodic.

Theorem 51. *The states of a communication class all have the same period.*

Proof. Suppose that there exist states i, j such that $i \leftrightarrow j$ and $d(i) \neq d(j)$. Since i and j communicate, there exist $r, s \in \mathbb{N}$ such that,

$$p_m(i,j) > 0 \quad \text{and} \quad p_n(j,i) > 0$$

Then $m + n$ is a possible return time for i,

$$p_{m+n}(i,i) = \sum_{k\in S} p_m(i,k) \cdot p_n(k,i) \geq p_m(i,j) \cdot p_n(j,i) > 0$$

[5] A communication class is **closed** if it consists of all recurrent states.

and $d(i)$ is a divisor of $m+n$. Assume that $p_r(j,j) > 0$ for some integer r. Then $p_{r+m+n}(i,i) \geq p_m(i,j) \cdot p_r(j,j) \cdot p_n(j,i) > 0$ and $d(i)$ is a divisor of $r+m+n$. Since $d(i)$ divides both $m+n$ and $r+m+n$, it must also divide r. Thus, $d(i)$ is a common divisor of the set $\{r > 0 \mid p_r(j,j) > 0\}$. Since $d(j)$ is the largest such divisor, $d(i) \leq d(j)$. A symmetric argument gives that $d(j) \leq d(i)$. ☐

Example 6: Periodicity

A random walk on the n-cycle has no limiting distribution when n is even. The graph is regular, and the unique stationary distribution is uniform. However, the chain alternates between even and odd states, and its position after n states depends on the parity of the initial state.

Definition 52 (Periodic Chain)**.** *A Markov chain is* ***periodic*** *if it is irreducible and all states have period greater than 1.*

Definition 53 (Aperiodic Chain)**.** *A Markov chain is* ***aperiodic*** *if it is irreducible and all states have period equal to 1.*

Note on Periodicity:
Any state i satisfying that $\boldsymbol{P}_{ii} > 0$ is necessarily aperiodic. Thus, a suficient condition for an irreducible Markov chain to be aperiodic is that $\boldsymbol{P}_{ii} > 0$ for some i, i.e., at least one diagonal entry of the transition matrix is non-zero.

Theorem 54. *$\boldsymbol{P}$ is irreducible and aperiodic if and only if $\boldsymbol{P}$ is regular.*

Proof. Suppose that $\boldsymbol{P}$ is irreducible and aperiodic. Consider any states i,j. Since $\boldsymbol{P}$ is irreducible, there exists $m(i,j)$ so that $p_{m(i,j)}(i,j) > 0$. Since $\boldsymbol{P}$ is aperiodic, there exists $M(i)$ so that for all $n \geq M(i)$, $p_n(i,i) > 0$. Taken together, this means that for $n \geq M(i)$,

$$p_{n+m(i,j)}(i,j) \geq p_n(i,i)p_{m(i,j)}(i,j) > 0$$

Now, $p_n(i,j) > 0$ for all $n \geq \max\{M(i) + m(i,j) \mid (i,j) \in S \times S\}$.

Suppose that $\boldsymbol{P}$ is regular. By definition, there exists an $M > 0$ such that for all $n \geq M$, $\boldsymbol{P}^n$ has all entries strictly positive. This means that $p_n(i,j) > 0$ for all states i,j, and consequently that $\boldsymbol{P}$ is irreducible. If $\boldsymbol{P}^n$ has strictly positive entries, so too does $\boldsymbol{P}^{n+1}$. Thus, $P(X_n = i \mid X_0 = i) > 0$ and $P(X_{n+1} = i \mid X_0 = i) > 0$. Since $gcd(n, n+1) = 1$, $\boldsymbol{P}$ is aperiodic. ☐

Theorem 55 (Limit Theorem for Aperiodic Irreducible Chains)**.** *If $(X_n)_{n\geq 0}$ is a finite state, time-homogeneous, aperiodic, irreducible Markov chain and $\boldsymbol{P}$ is its transition matrix, then there is a unique, positive, stationary distribution $\vec{\pi} > 0$ such that,*

$$\lim_{n\to\infty} P^n = \boldsymbol{V}$$

where $\boldsymbol{V}$ is a matrix with all rows equal to $\vec{\pi}$.

Absorbing Chains

Definition 56 (Absorption)**.** *A state $i \in S$ is called **absorbing** if,*

$$p(i,i) = 1$$

An absorbing chain has at least one absorbing state[6]*.*

Definition 57 (Absorption Probability)**.** *Let $(X_n)_{n\geq 0}$ be a Markov chain with all states transient or absorbing. The **absorption probability** is the probability that the chain is absorbed in state j from transient state i.*

Definition 58 (Absorption Time)**.** *Let $(X_n)_{n\geq 0}$ be a Markov chain with all states transient or absorbing. The **absorption time** is the expected number of steps from transient state i to absorption in some absorbing state.*

Note on Absorption Time:
The problem of computing hitting time reduces to the problem of computing time to absorption because a state can be modified to be absorbing.

Theorem 59. *Let $(X_n)_{n\geq 0}$ be finite-state and irreducible with transition matrix $\boldsymbol{P}$. To find the expected hitting time, $\mathbb{E}[\tau_i]$,*

- *Consider a new chain in which i is an absorbing state*
- *Define $\tilde{\boldsymbol{P}}$ by deleting the ith row and setting $\tilde{\boldsymbol{P}}_{ii} = 1$*
- *Define $\boldsymbol{Q}$ by deleting the ith row and i-h column of $\boldsymbol{P}$*

Assume that the chain starts in state a. The first time that $\boldsymbol{P}$ hits i is,

$$\begin{cases} \sum_{b\neq i}(\boldsymbol{I} - \tilde{\boldsymbol{P}})^{-1}_{a,b} & \text{if } i \neq a \\ 1 + \sum_{j\neq i} \boldsymbol{P}_{ij} \cdot \sum_{b\neq i}(\boldsymbol{I} - \tilde{\boldsymbol{P}})^{-1}_{j,b} & \text{if } i = a \end{cases}$$

Proof. We need to find,

$$\begin{aligned} \mathbb{E}[\tau_a] &= \mathbb{E}[\inf\{n \geq 0 \mid X_n \in A\}] \\ &= \sum_{n=1}^{\infty} n \cdot P(\tau_a = n) \\ &= \sum_{n=1}^{\infty} P(\tau_a \geq n) \end{aligned}$$

For an absorbing state a,

$$P(\tau_a \geq n) = P(X_{n-1} \neq a) \quad (n \geq 1)$$

Considering every probable path that does not go through a,

$$\begin{aligned} \mathbb{E}[\tau_a] &= \sum_{i=1}^{k} \sum_{n=1}^{\infty} \sum_{b\neq a} \pi_i \cdot \tilde{\boldsymbol{P}}^{n-1}_{i,b} \\ &= \sum_{i=1}^{k} \sum_{b\neq 1} \pi_i (\boldsymbol{I} - \tilde{\boldsymbol{P}})^{-1}_{i,b} \end{aligned}$$

□

---

[6] Intuitively, this is a state *i* that the chain never leaves once it first visits *i*.

Positive and Null Recurrence

Definition 60 (Positive Recurrent). *A recurrent state j is* ***positive recurrent*** *if the expected return time $\mathbb{E}[\tau_j^+ \mid X_0 = j]$ is finite*[7].

Theorem 61 (Limit Theorem for Irreducible, Positive Recurrent Chains). *Let $(X_n)_{n\geq 0}$ be an infinite, irreducible, and positive recurrent Markov chain. There exists a unique, positive, stationary distribution π, which is the limiting distribution of the chain*[8]. *That is,*

$$\pi_j = \lim_{n\to\infty} \boldsymbol{P}_{ij}^n \quad \textit{for all } i, j$$

Moreover,

$$\pi_j = \frac{1}{\mathbb{E}[\tau_j^+]} \quad \textit{for all } j$$

Definition 62 (Null Recurrent). *A recurrent state j is* ***null recurrent*** *if the expected return time $\mathbb{E}[\tau_j^+ \mid X_0 = j]$ is infinite*[9].

Theorem 63. *Positive and null recurrence are class properties*[10].

Proof. Assume that i is a positive recurrent state. Let j be another state in the same communication class as i. Since $i \leftrightarrow j$, there exist $s, r \geq 0$ such that $p_r(j,i) > 0$ and $p_s(i,j) > 0$. Thus,

$$\begin{aligned}
\frac{1}{\mu_j} &= \lim_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n-1} p_m(j,i) \\
&\geq \lim_{n\to\infty} \frac{1}{n} \sum_{m=r+s}^{n-1} p_r(j,i) \cdot p_{m-r-s}(i,i) \cdot p_s(i,j) \\
&= \lim_{n\to\infty} \left(\frac{n-r-s}{n}\right) p_r(j,i) \left(\frac{1}{n-r-s} \sum_{m=r+s}^{n-1} p_{m-r-s}(i,i)\right) P_s(i,j) \\
&= p_r(j,i) \left(\frac{1}{\mu_i}\right) p_s(i,j) > 0.
\end{aligned}$$

Hence, $\mu_j < \infty$ and j is positive recurrent. ☐

Definition 64 (Ergodic). *A Markov chain is called* ***ergodic*** *if it is irreducible, aperiodic, and all states have finite expected return times*[11].

Theorem 65 (Limit Theorem for Ergodic Chains). *Let $(X_n)_{n\geq 0}$ be an ergodic Markov chain. There exists a unique, positive, stationary distribution π, which is the limiting distribution of the chain. That is,*

$$\pi_j = \lim_{n\to\infty} \boldsymbol{P}_{ij}^n \quad \textit{for all } i, j$$

Moreover,

$$\pi_j = \frac{1}{\mathbb{E}[\tau_j^+]} \quad \textit{for all } j$$

---

[7] Positive recurrence is the infinite analog of finite recurrent chains.

[8] For infinite irreducible chains that are null recurrent, no stationary distribution exists.

[9] Null recurrent chains do not have stationary distributions.

[10] In particular, all states in a recurrent communication class are either positive or null recurrent.

[11] Every finite chain is ergodic, and the condition that all states have finite expected return times is equivalent to all states being positive recurrent.

Reversibility

Definition 66 (Detailed Balance Equations)**.**

$$\pi_i \boldsymbol{P}_{ij} = \pi_j \boldsymbol{P}_{ji} \quad \textit{for all } i, j \in S$$

More generally, we can write,

$$P(X_0 = i_0, X_1 = i_1, \cdots, X_n = i_n) = P(X_0 = i_n, X_1 = i_{n-1}, X_n = i_0)$$

Theorem 67. *Let $(X_n)_{n\geq 0}$ be a Markov chain with transition matrix $\boldsymbol{P}$. If $\vec{\pi}$ satisfies the detailed balance equations, then $\vec{\pi}$ is stationary*[12] *for $\boldsymbol{P}$.*

Proof. $\sum_{i\in S} \pi_i \boldsymbol{P}_{ij} = \sum_{i\in S} \pi_j \boldsymbol{P}_{ji} = \pi_j$ since **P** is stochastic. ☐

Definition 68. *An irreducible Markov chain $(X_n)_{n\geq 0}$ with transition matrix $\boldsymbol{P}$ and stationary distribution $\vec{\pi}$ is **time reversible** if it satisfies the detailed balance equations. The time reversal of (X_n) is the chain with,*

$$\hat{\boldsymbol{P}}_{ij} := \frac{\pi_j \boldsymbol{P}_{ij}}{\pi_i}$$

as its transition matrix.

Remark 69. *If a chain with transition matrix $\boldsymbol{P}$ is reversible, then $\hat{\boldsymbol{P}} = \boldsymbol{P}$.*

Proof. By the detailed balance equations,

$$\hat{\boldsymbol{P}}_{ij} = \frac{\pi_j \boldsymbol{P}_{ij}}{\pi_i} = \frac{\pi_j \boldsymbol{P}_{ji}}{\pi_i}$$

so $\boldsymbol{P}_{ji} = \boldsymbol{P}_{ij}$. ☐

Example 7: Reversibility

A simple random walk on a graph is time reversible.

$$\begin{aligned} \pi_i \boldsymbol{P}_{ij} &= \left(\frac{\deg(i)}{2|E(G)|}\right)\left(\frac{1}{\deg(i)}\right) \quad \text{for neighbors } i, j \\ &= \frac{1}{2|E(G)|} \\ &= \left(\frac{\deg(j)}{2|E(G)|}\right)\left(\frac{1}{\deg(j)}\right) \\ &= \pi_j \boldsymbol{P}_{ji} \end{aligned}$$

Markov Chain Monte Carlo

Markov Chain Coupling

Definition 70 (Total Variation Distance)**.** *The **total variation distance** between two distributions μ and ν on a state space S is defined by*[13]*,*

$$\|\mu - \nu\|_{TV} = \sup_{A\subseteq S} \|\mu(A) - \nu(A)\|$$

---

[12] Checking detailed balance is often the simplest way to verify that a particular distribution is stationary.

[13] This definition is explicitly probabilistic: the distance between μ and ν is the maximum difference between the probabilities assigned to a single event by the two distributions.

Example 8: Coupling Distance

Suppose, for illustration, that the total variation distance,

$$\|\pi_7 - \pi\|_{TV} = 0.17$$

This tells us that the probability of any event, for example, the probability of winning any specified card game using a deck shuffled 7 times, differs by at most 0.17 from the probability of the same event using a perfectly shuffled deck.

A **coupling of two probability distributions** μ and ν is a pair of random variables (X, Y) defined on a single probability space such that the marginal distribution of X is μ and the marginal distribution of Y is ν. That is, a coupling (X, Y) satisfies,

$$P(X = x) = \mu(x) \text{ and } P(Y = y) = \nu(y)$$

Definition 71 (Coupling of Markov Chains)**.** *A **coupling of Markov chains** is a process $(X_n, Y_n)_{n\geq 0}$ with the property that both $(X_n), (Y_n)$ are Markov chains with transition matrix **P**, although the chains may have different starting distributions.*

Definition 72 (Coupling Time)**.** *The **coupling time** of a process $(X_n, Y_n)_{n\geq 0}$ is defined to be the first time T in which X_n equals Y_n,*

$$T = \inf\{n \mid X_n = Y_n\}$$

Lemma 73. *Consider a coupling $(X_n, Y_n)_{n\geq 0}$ of Markov chains $(X_n), (Y_n)$.*

$$\|\vec{\pi}_0 \cdot \boldsymbol{P}^n - \vec{\pi}\|_{TV} \leq P(T > n) \text{ for all } n > 0$$

where $\vec{\pi}$ is the stationary distribution of (Y_n).

The coupling inequality reduces the problem of showing that,

$$\|\vec{\pi}_0 \cdot \boldsymbol{P}^n - \vec{\pi}\| \to 0$$

to that of showing,

$$P(T > n) \to 0 \iff P(T < \infty) = 1$$

Proof. Define the process (Y_n^*) by,

$$Y_n^* = \begin{cases} Y_n & \text{if } n < T \\ X_n & \text{if } n \geq T \end{cases}$$

(Y_n^*) is a Markov chain with the same probability transition matrix **P** as (X_n). This is because Y_n and X_n share **P** and Z_n and X_n share $\vec{\pi}_0$. Since we also have that $(Y_n^*) \sim \pi$ for all n,

$$\begin{aligned} \vec{\pi}_0 \cdot \boldsymbol{P}^n(A) - \vec{\pi} &= P(X_n \in A) - P(Y_n^* \in A) \\ &= P(X_n \in A, T \leq n) + P(X_n \in A, T > n) \\ &\quad - P(Y_n^* \in A, T \leq n) - P(Y_n^* \in A, T > n) \end{aligned}$$

However, on the event $\{T \leq n\}$, $Y_n^* = X_n$, so that $P(X_n \in A, T \leq n) = P(Y_n^* \in A, T \leq n)$. Simplifying gives that,

$$\vec{\pi}_0 \cdot \boldsymbol{P}^n(A) - \vec{\pi} = P(X_n \in A, T > n) - P(Y_n^* \in A, T > n) < P(T > n)$$

□

Remark 74 (Doeblin Coupling Argument). *Proving the Markov Convergence Theorem can be done by showing that* $P(T < \infty) = 1$.

Proof. The bivariate chain $Z = \{(X_n, Y_n) \mid n \geq 0\}$ is a Markov chain on the state space $S \times S$. Its transition matrix $\boldsymbol{P}^Z$ can be written,

$$\boldsymbol{P}^Z_{(i_1,i_2),(j_1,j_2)} = \boldsymbol{P}^X_{(i_1,j_1)} \cdot \boldsymbol{P}^Y_{(i_2,y_2)}$$

and the stationary distribution is,

$$\pi^Z(i,j) = \pi_i \pi(y)$$

$P(T < \infty) = 1$ occurs if the Z chain hits $\{(j,j) \mid j \in S\} \subseteq S \times S$. Since Z has a stationary distribution, it suffices to show that Z is irreducible[14]. This can be done using a numbertheoretic proof. ☐

Metropolis-Hastings Algorithm

Definition 75 (Markov Chain Monte Carlo). *Given a discrete or continuous probability distribution $\vec{\pi}$, the goal of* ***Markov Chain Monte Carlo*** *is to simulate a random variable $X \sim \vec{\pi}$.*

Remark 76 (Strong Law of Large Numbers for Markov Chains). *If $(X_n)_{n\geq 0}$ is a finite state, time-homogeneous, aperiodic, irreducible Markov chain and r is a bounded and real-valued function, then*[15],

$$\lim_{n\to\infty} \frac{r(X_1) + \cdots + r(X_n)}{n} = \mathbb{E}[r(X)] \quad a.s.$$

where $\mathbb{E}[r(X)] = \sum_j r(j)\pi_j$.

Definition 77 (Metropolis-Hastings Algorithm). *Let $\vec{\pi}$ be a discrete probability distribution. The* ***Metropolis-Hastings Algorithm*** *constructs a reversible Markov chain $(X_n)_{n\geq 0}$ whose stationary distribution is $\vec{\pi}$,*

1. *Let $\boldsymbol{T}$, the proposal chain, be a transition matrix*[16] *for any irreducible Markov chain with the same state space as $\vec{\pi}$*
2. *Assume that at time n, the chain is at state i*
3. *Choose a new state j, the proposal state, according to $\boldsymbol{T}_{ij}$*
4. *Let $U \sim Unif(0,1)$. If $X_n = i$, define an acceptance function,*

$$a(i,j) = \frac{\pi_j \boldsymbol{T}_{ji}}{\pi_i \boldsymbol{T}_{ij}} \quad \text{and let } X_{n+1} := \begin{cases} j & \text{if } U \leq a(i,j) \\ i & \text{otherwise} \end{cases}$$

Proof. The sequence $(X_n)_{n\geq 1}$ constructed by the Metropolis-Hastings Algorithm is a Markov chain, as each X_{n+1} only depends on X_n. Let

---

[14] If an irreducible Markov chain has a stationary distribution, then the chain is recurrent.

[15] The chain is not i.i.d, but successive excursions between visits to the same state are independent.

[16] It is assumed that the user knows how to sample from **T**.

P be its transition matrix. We need to show that (X_n) is reversible with stationary distribution is $\vec{\pi}$. Given $X_0 = i$, then,

$$\begin{aligned} P(U \leq a(i,j)) &= \begin{cases} a(i,j) & \text{if } a(i,j) \leq 1 \\ 1 & \text{otherwise} \end{cases} \\ &= \begin{cases} a(i,j) & \text{if } \pi_j \boldsymbol{T}_{ji} \leq \pi_i \boldsymbol{T}_{ij} \\ 1 & \text{otherwise} \end{cases} \end{aligned}$$

and for $i \neq j$,

$$\boldsymbol{P}_{ij} = \begin{cases} \boldsymbol{T}_{ij} \cdot a(i,j) & \text{if } \pi_j \boldsymbol{T}_{ji} \leq \pi_i \boldsymbol{T}_{ij} \\ \boldsymbol{T}_{ij} & \text{otherwise} \end{cases}$$

The diagonal entries of **P** are determined by the fact that the rows of **P** sum to 1. There are two cases,

- If $\pi_j \boldsymbol{T}_{ji} \leq \pi_i \boldsymbol{T}_{ij}$,

$$\pi_i \boldsymbol{P}_{ij} = \pi_i \boldsymbol{T}_{ij} a(i,j) = \pi_i \boldsymbol{T}_{ij} \left(\frac{\pi_j \boldsymbol{T}_{ij}}{\pi_i \boldsymbol{T}_{ij}} \right) = \pi_j \boldsymbol{T}_{ji} = \pi_j \boldsymbol{P}_{ji}$$

- If $\pi_j \boldsymbol{T}_{ji} < \pi_i \boldsymbol{T}_{ij}$,

$$\pi_i \boldsymbol{P}_{ij} = \pi_i \boldsymbol{T}_{ij} a(i,j) = \pi_j \boldsymbol{T}_{ji} \left(\frac{\pi_i \boldsymbol{T}_{ij}}{\pi_j \boldsymbol{T}_{ji}} \right) = \pi_j \boldsymbol{T}_{ji} a(j,i) = \pi_j \boldsymbol{P}_{ji}$$

Hence, the detailed balance equations are satisfied. ☐

Classical Markov Chains

Electrical Networks

We can represent Markov chains as electrical networks, which we typically denote by undirected weighted graphs $G = (V, E)$.

Definition 78 (Electrical Network)**.** *A **network** is a pair (G, c), where $G = (V, E)$ is a countable graph and $c : E \to (0, \infty)$ is a function assigning a positive **conductance** to each edge such that,*

$$C_x := \sum_{y \sim x} C_{xy} < \infty \quad \forall x \in V$$

and $(X_n)_{n \geq 0}$ is a random walk with transition matrix,

$$\boldsymbol{P}_{xy} := \frac{C_{xy}}{C_x}$$

Definition 79 (Resistance). *The **resistance** R_{xy} of an edge $e = (x, y)$ is,*

$$R_{xy} = \frac{1}{C_{xy}}$$

That is, it is the reciprocal of its conductance.

Definition 80 (Effective Conductance). *The **effective conductance** between two disjoint, non-empty finite sets of vertices A and B is defined,*

$$c^{eff}_{A,B} := \sum_{v \in A} c_a \cdot P(\tau_v < \tau_b^+)$$

where τ_v is the first hitting time and τ_a^+ is the first return time[17]*. When $A = \{a\}$ and $B = \{b\}$, the effective conductance is,*

$$C^{eff}_{A,B} := \sum_{v \in A} C_a \cdot P(\tau_v < \tau_b^+ \mid X_0 = a)$$

When A and B are not disjoint, we define $C^{eff}_{A,B} = \infty$.

Definition 81 (Effective Resistance). *The **effective resistance** between two disjoint, non-empty finite sets of vertices A and B is defined,*

$$R^{eff}_{A,B} = \frac{1}{C^{eff}_{A,B}}$$

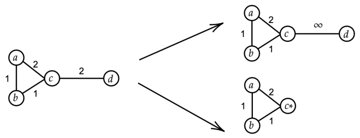
Remark 82 (Series Additivity). *Vertices connected in series can be replaced by one vertex whose resistance is the sum of the resistances.*

Remark 83 (Parallel Additivity). *Vertices connected in parallel can be replaced by one vertex whose conductance is the sum of the conductances.*

Remark 84. *Given a network G and two disjoint, finite, non-empty sets of vertices A and B, we can form a network G' by contracting ("shorting") each of the sets A and B into single vertices $[A]$ and $[B]$. Then,*

$$c^{eff}_{A,B}(G) = c^{eff}_{[A],[B]}(G')$$

Shorting two vertices adds infinite conductance between them (or, merges them while preserving edges). Thus, shorting two sets of nodes can only decrease the effective resistence of the network between two nodes.



Cutting an edge can only increase the effective resistance between the two vertices that it is adjacent to.

Remark 85 (Rayleigh's Monotonicity Principle). *Let $G = (V, E)$ be a finite graph, and let A and B be disjoint, non-empty subsets of V. Then,*

$$c^{eff}_{A,B}(G)$$

*is a **monotone increasing** function of $c \in (0, \infty)^E$.*

Remark 86. *An irreducible Markov chain (X_n) is recurrent if and only if the conductance to infinity, $\lim_{n\to\infty} C_{0,[n]}$, is finite,*

$$P(\tau_0 < \infty) = \lim_{n\to\infty} C_{0,[n]} < \infty$$

[17] Recall that we write τ_z and τ_a^+ for the first time that the random walk visits z and the first positive time that the random walk visits a respectively.

Pólya's Theorem

Theorem 87 (Pólya's Theorem)**.** *The d-dimensional hybercubic lattice $\mathbb{Z}^d$ is recurrent if $d \leq 2$ and transient if $d \leq 3$.*

Corollary 88. *A simple random walk on any subgraph of $\mathbb{Z}^2$ is recurrent*[18]*.*

Pólya's Urn

Definition 89 (Pólya Urn Model)**.** ***Pólya's Urn*** *is the process,*

- *An urn contains two balls, one black and one white*
- *Proceed by choosing a ball at random from those already in the urn*
- *Return the chosen ball to the urn and add another ball of the same color*

The sequence of ordered pairs listing the numbers of black and white balls is a Markov chain. A configuration (a, b) with a black balls and b white balls evolves according to,

$$(a,b) \to \begin{cases} (a+1,b) & \textit{with probability } \frac{a}{a+b} \\ (b,b+1) & \textit{with probability } \frac{b}{a+b} \end{cases}$$

Example 9: Pólya's Urn

We want to find the likelihood that the system reaches,

$$(B,W) = (m,n) \quad \text{starting from} \quad (B,W) = (b,w)$$

Consider the transition,

$$(1,1) \to (3,3)$$

where one possible path is,

$$(1,1) \to (1,2) \to (1,3) \to (2,3) \to (3,3)$$

Its likelihood is,

$$\frac{1}{2} \times \frac{2}{3} \times \frac{1}{4} \times \frac{2}{5} = \frac{(1 \cdot 2) \cdot (1 \cdot 2)}{2 \cdot 3 \cdot 4 \cdot 5}$$

there are $\binom{4}{2} = 6$ distinct routes, each with this same probability. In general, for a path $(b, w) \to (m, n)$,

$$\frac{[b(b+1)\cdots(m-1)] \cdot [w(w+1)\cdots(n-1)]}{(b+w)(b+w+1)\cdots(m+n-1)}$$

Rewriting this probability using factorials,

$$\frac{(m-1)!}{(b-1)!} \times \frac{(n-1)!}{(w-1)!} \times \frac{(b+w-1)!}{(m+n-1)!}$$

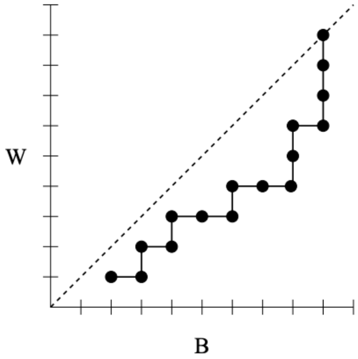


Figure 1: The urn process as a trajectory on a two-dimensional lattice, where bullets indicate intermediate stages.

---

[18] Adding a finite number of edges to a transient graph preserves transience.

The total number of distinct paths from (b, w) to (m, n) is,

$$\binom{m+n-b-w}{m-b}$$

so the transition probability P that, starting from configuration (b, w), the system reaches (m, n) is:

$$P = \binom{m-1}{b-1}\binom{n-1}{w-1}\binom{m+n-1}{b+w-1}^{-1}$$

In particular, for $(b, w) = (1, 1)$,

$$P = \frac{1}{m+n-1} \quad \text{or, if } U = m+n, \quad P = \frac{1}{U}$$

Branching Processes

Mean Generation Size

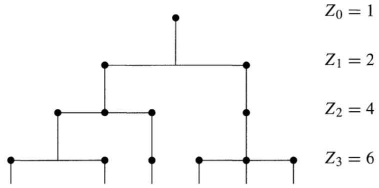


Figure 2: Family tree.

Definition 90 (Offspring Distribution)**.** *The **offspring distribution** X gives the probability x_k that an individual gives birth to k children,*

$$X = (x_1, x_2, \cdots)$$

independently of other individuals.

Definition 91 (Branching Process)**.** *Let Z_n be a random variable that denotes the size of a population of living species. The Markov chain $(Z_n)_{n \geq 1}$ with values in $\mathbb{N}_0$ is a **branching process** if,*

$$Z_{n+1} = \sum_{j=1}^{Z_n} X_j$$

where X_j denotes the number of children born to the jth person in the nth generation. $(X_j)_{j \geq 1}$ is an i.i.d sequence with common distribution X. Furthermore, Z_n is independent of (X_j).

Definition 92 (Extinction Time)**.** *The **extinction time** T_0 of a branching process is the hitting time to zero, that is, $T_0 := \tau_0$.*

Definition 93 (Mean Generation Size)**.** *The **mean generation size** μ_n is the mean size of the nth generation, that is, $\mu_k := \mathbb{E}[Z_n]$.*

Definition 94 (Mean Offspring Size)**.** *The **mean offspring size** μ is the mean of the offspring distribution, that is, $\mu_n := \mathbb{E}[X]$.*

Theorem 95. *Let $\mu = \sum_{k=0}^{\infty} k \cdot x_k$ be the mean of the offspring distribution.*

$$\mathbb{E}[Z_n] = \mu^n$$

Proof. By the Total Law of Expectation,

$$\begin{aligned}
\mathbb{E}[Z_n] &= \sum_{k=0}^{\infty} \mathbb{E}[Z_n \mid Z_{n-1} = k] \cdot P\left(Z_{n-1} = k\right) \\
&= \sum_{k=0}^{\infty} \mathbb{E}\left[\sum_{i=1}^{Z_{n-1}} X_i \;\middle|\; Z_{n-1} = k \right] \cdot P\left(Z_{n-1} = k\right) \\
&= \sum_{k=0}^{\infty} \mathbb{E}\left[\sum_{i=1}^{k} X_i \mid Z_{n-1} = k \right] \cdot P\left(Z_{n-1} = k\right) \\
&= \sum_{k=0}^{\infty} \mathbb{E}\left[\sum_{i=1}^{k} X_i \right] \cdot P\left(Z_{n-1} = k\right) \text{ since } X_i \text{ and } Z_{n-1} \text{ independent} \\
&= \sum_{k=0}^{\infty} \mu \cdot kP\left(Z_{n-1} = k\right) \\
&= \mu \cdot \mathbb{E}[Z_{n-1}]
\end{aligned}$$

Iterating the recurrence for $n \geq 0$ gives that,

$$\mathbb{E}[Z_n] = \mu \mathbb{E}[Z_{n-1}] = \mu^2 \mathbb{E}[Z_{n-2}] = \cdots = \mu^n \mathbb{E}[Z_0] = \mu^n$$

since $Z_0 = 1$ ☐

Theorem 96. *If $\mu < 1$, then $P(T_0 > n) \leq \mathbb{E}[Z_0] = \mu^n$. In particular, the branching process goes extinct with probability 1, $(T_0 < \infty) = 1$.*

Proof. By Markov's Inequality,

$$P(T_0 > n) = P(Z_n \geq 1) \leq \mathbb{E}[Z_n] = \mu^n \mathbb{E}[Z_0]$$

☐

Definition 97 (Criticality). *A branching process is* ***subcritical*** *if $\mu < 1$,* ***critical*** *if $\mu = 1$, and* ***supercritical*** *if $\mu > 1$. Moreover,*

$$\lim_{n \to \infty} \mathbb{E}[Z_n] = \lim_{n \to \infty} \mu^n = \begin{cases} 0, & \text{if } \mu < 1 \\ 1, & \text{if } \mu = 1 \\ \infty, & \text{if } \mu > 1 \end{cases}$$

Note on Criticality:
For a subcritical process, mean generation size declines exponentially to zero. For a supercritical process, it exhibits exponential growth.

Generating Functions

Definition 98 (Generating Function). *Let X be a discrete random variables with values in $\mathbb{N}_0$. The* ***probability generating function*** *of X is,*

$$\begin{aligned}
G(s) = \mathbb{E}[s^X] &= \sum_{k=0}^{\infty} s^k \cdot P(X = k) = \sum_{k=0}^{\infty} s^k \cdot \pi_k \\
&= P(X = 0) + sP(X = 1) + s^2 P(X = 2) + \cdots
\end{aligned}$$

where $\vec{\pi} = (\pi_k)_{k \geq 0}$ is the law of X.

Example 10: Generating Functions

Let $X \sim \text{Unif}(\{0,1,2\})$. Then,

$$G(s) = \frac{1}{3} + s\left(\frac{1}{3}\right) + s^2\left(\frac{1}{3}\right) = \frac{1}{3}\left(1 + s + s^2\right)$$

Let $X \sim \text{Geom}(p)$. For $|s| < 1$,

$$G(s) = \sum_{k=1}^{\infty} s^k p(1-p)^{k-1} = sp\sum_{k=1}^{\infty}(s(1-p))^{k-1} = \frac{sp}{1-s(1-p)}$$

Let $X \sim \text{Po}(\mu)$. For $\mu > 0$,

$$G(s) = \sum_{k=0}^{\infty} \frac{e^{-\mu}\mu^k}{k!} \cdot s^k = e^{-\mu} \cdot \sum_{k=0}^{\infty} \frac{(\mu s)^k}{k!} = e^{-\mu}e^{\mu s} = e^{\mu(s-1)}$$

Remark 99. *The series $G(s)$ converges absolutely for $|s| \leq 1$.*

Proof. Let $\vec{\pi}$ be the law of X. Then,

$$|G(s)| = \left|\sum_{k=0}^{\infty} s^k \cdot \pi_k\right| \leq \sum_{k=0}^{\infty} |s|^k \cdot \pi_k \leq 1$$

so $G(s)$ exists and is well-defined for $|s| \leq 1$. ☐

Theorem 100. *Let (X_n) be an i.i.d sequence of random variables. Define $Z := \sum_{i=1}^{n} X_n$. The probability generating function of Z is $[G_X(s)]^n$.*

Proof. Expanding the definition,

$$\begin{aligned} G_Z(s) &= \mathbb{E}[s^Z] \\ &= \mathbb{E}[s^{\sum_{i=1}^{n} X_n}] \\ &= \mathbb{E}\left[\prod_{k=1}^{n} s^{X_k}\right] \\ &= \prod_{k=1}^{n} \mathbb{E}[s^{X_k}] \text{ by independence} \\ &= G_{X_1}(s) \cdots G_{X_n}(s) \end{aligned}$$

If X_n is i.i.d., then,

$$G_Z(s) = G_{X_1}(s) \cdots G_{X_n}(s) = [G_X(s)]^n$$

where X has the same distribution as X_i. ☐

Theorem 101. *Probabilities for X can be obtained from the generating function by successive differentiation. If $G^{(j)}$ is the jth derivative of G,*

$$G^{(j)}(s) = \sum_{k=j}^{\infty} k(k-1)\cdots(k-j+1)s^{k-j}P(X=j)$$

Properties of Generating Functions:

1. If X and Y satisfy,

$$G_X(s) = G_Y(s) \quad \forall s$$

then $X \overset{law}{=} Y$.

2. If X and Y are independent, then,

$$G_{X+Y}(s) = G_X(s) \cdot G_Y(s)$$

Proof. Observe that,

$$\begin{aligned}
G(0) &= P(X=0) \\
G'(0) &= \sum_{k=1}^{\infty} ks^{k-1}P(X=k)\Big|_{s=0} = P(X=1) \\
G''(0) &= \sum_{k=2}^{\infty} k(k-1)s^{k-2}P(X=k)\Big|_{s=0} = 2P(X=2)
\end{aligned}$$

and so on. In general,

$$G^{(j)}(0) = \sum_{k=j}^{\infty} k(k-1)\cdots(k-j+1)s^{k-j}P(X=j)\Big|_{s=0} = j!P(X=j)$$

and thus

$$P(X=j) = \frac{G^{(j)}(0)}{j!}, \quad \text{for } j = 0, 1, \ldots$$

$\square$

Theorem 102. *The generating function of the nth generation size Z_n is the n-fold composition of the offspring distribution generating function,*

$$G_n(s) = \mathbb{E}[s^{Z_n}] = \underbrace{G \circ G \circ \cdots \circ G}_{n \text{ times}}(\mathbb{E}[s^{Z_0}])$$

Proof. The generating function of the nth generation size Z_n is,

$$G_n(s) = \mathbb{E}[s^{Z_n}] = \mathbb{E}\left[s^{\sum_{k=1}^{Z_{n-1}} X_k}\right] = \mathbb{E}\left[\mathbb{E}\left[s^{\sum_{k=1}^{Z_{n-1}} X_k} \mid Z_{n-1}\right]\right]$$

where the last inequality is by the Total Law of Expectation.

$$\begin{aligned}
\mathbb{E}\left[s^{\sum_{k=1}^{Z_{n-1}} X_k} \mid Z_{n-1} = z\right] &= \mathbb{E}\left[s^{\sum_{k=1}^{z} X_k} \mid Z_{n-1} = z\right] \text{ by conditioning} \\
&= \mathbb{E}\left[s^{\sum_{k=1}^{z} X_k}\right] \text{ by independence} \\
&= \mathbb{E}\left[\prod_{k=1}^{z} s^{X_k}\right] \\
&= \prod_{k=1}^{z} \mathbb{E}[s^{X_k}] \text{ by independence} \\
&= [G(s)]^z \text{ for all } z
\end{aligned}$$

$G(s)$ is the generating function of the offspring distribution,

$$G(s) = \sum_{k=0}^{\infty} s^k \pi_k$$

so this gives that,

$$\mathbb{E}[s^{\sum_{k=1}^{Z_{n-1}} X_k} \mid Z_{n-1}] = [G(s)]^{Z_{n-1}}$$

Taking expectations,

$$G_n(s) = \mathbb{E}[G(s)^{Z_{n-1}}] = G_{n-1}(G(s))$$

The result follows by induction on n. □

Corollary 103. *The extinction probability of a branching process is,*

$$P(T_0 < \infty) = \lim_{k\to\infty} \lim_{s\to 0} \underbrace{G \circ G \circ \cdots \circ G}_{k \text{ times}}(\mathbb{E}[s^{Z_0}])$$

Proof. The generating function for the nth generation size Z_n is,

$$G_n(s) = \sum_{k=0}^{\infty} s^k P\left(Z_n = k\right)$$

Since $P(T_0 \le k) = P(Z_k = 0)$,

$$P(Z_k = 0) = \lim_{s\to 0} \mathbb{E}[s^{Z_k}] = \lim_{s\to 0} G_k(s) = \lim_{s\to 0} \underbrace{G \circ G \circ \cdots \circ G}_{k \text{ times}}(\mathbb{E}[s^{Z_0}])$$

Now, $P(T_0 < \infty) = P(\bigcup_{k=1}^{\infty}\{T_0 \le k\}) = P(\bigcup_{k=1}^{\infty}\{Z_k = 0\})$. Therefore,

$$\begin{aligned} P(T_0 < \infty) &= P(\bigcup_{k=1}^{\infty}\{Z_k = 0\}) \\ &= \lim_{k\to\infty} P(\{Z_k = 0\}) \text{ since } \{Z_k = 0\} \subseteq \{Z_{k+1} = 0\} \\ &= \lim_{k\to\infty} \lim_{s\to 0} \underbrace{G \circ G \circ \cdots \circ G}_{k \text{ times}}(\mathbb{E}[s^{Z_0}]) \end{aligned}$$

by the continuity of measure over increasing unions. □

Theorem 104. *The probability generating function $G(s)$ of a discrete random variable X is convex and non-decreasing on $[0, 1]$ with $G(1) = 1$.*

Proof. We have that,

1. $G(1) = 1$ since,
   $$\mathbb{E}[1^X] = \sum_{k=0}^{\infty} P(X = k) = 1$$
2. $G(s)$ is strictly increasing since,
   $$G'(s) = \sum_{k=1}^{\infty} ks^{k-1}\pi_k > 0 \quad (s > 0 \text{ and } \pi_k \neq 0 \text{ for all } k \ge 1)$$
3. $G(s)$ is strictly convex since,
   $$G''(s) = \sum_{k=2}^{\infty} k(k-1)s^{k-2}\pi_k > 0 \quad (s > 0 \text{ and } \pi_k \neq 0 \text{ for all } k \ge 2)$$

□

Theorem 105. *$G'(1)$ can be used to find the mean of X.*

Proof. $G'(1) = \sum_{k=1}^{\infty} k \cdot \pi_k = \sum_{k=1}^{\infty} k \cdot P(X = k) = \mathbb{E}[X]$. □

Theorem 106. *Let $G(s)$ be the probability generating function of a discrete random variable X. The smallest positive root of the equation $G(s) = s$ is the probability of eventual extinction, that is, $P(T_0 < \infty)$.*

Proof. The extinction probability $P(T_0 < \infty)$ of a branching process is a root of the equation $s = G(s)$. To see this,

$$\begin{aligned} P(T_0 \le k) &= P(Z_k = 0) \\ &= G_k(0) \\ &= G(G_{k-1}(0)) \\ &= G(P(Z_{k-1} = 0)) \\ &= G(P(T_0 \le k-1)) \end{aligned}$$

Taking the limits on both sides and using the continuity of G,

$$P(T_0 < \infty) = G(P(T_0 < \infty))$$

Let x be a positive solution of $s = G(s)$. We need to show that,

$$P(T_0 < \infty) = \lim_{k\to\infty} P(Z_k = 0) \le x$$

The proof is by induction on k. Since $G(s)$ is increasing on $(0, 1]$ and $x > 0$, $P(Z_1 = 0) = G_1(0) = G(0) \le G(x) = x$. Assume that $P(Z_k = 0) \le x$ for $k < n$. Then, $P(Z_n = 0) = G_n(0) = G(G_{n-1}(0)) = G(P(T_0 \le k-1)) \le G(x) = x$. Taking limits as $n \to \infty$ gives,

$$P(T_0 < \infty) \le x$$

□

Theorem 107. *Let $G(s)$ be the probability generating function of a discrete random variable X. Exactly one of the following holds,*

1. *$G(s) = s$ for infinitely many $s \in [0, 1]$ and,*

$$\lim_{k\to\infty} \underbrace{G \circ G \circ \cdots \circ G}_{k \text{ times}}(s) = s \quad \forall s \in [0,1]$$

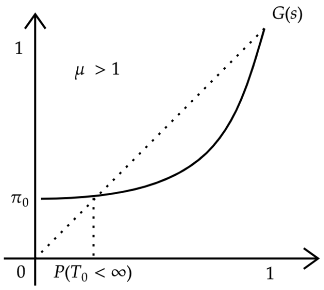
2. *$G(s) = s$ for two points $s_1, s_2 \in [0, 1]$ and,*

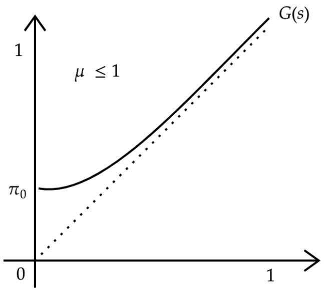
$$\lim_{k\to\infty} \underbrace{G \circ G \circ \cdots \circ G}_{k \text{ times}}(s) = s_2 \quad \forall s \in [0,1] \text{ and } s_2 \neq 1$$

3. *$G(s) = s$ for a unique point $s \in [0, 1]$ and,*

$$\lim_{k\to\infty} \underbrace{G \circ G \circ \cdots \circ G}_{k \text{ times}}(s) = 1 \quad \forall s \in [0,1]$$

Proof. If $\pi_k = \delta_{1k}$ for all $k \geq 0$, then $G(s) = s$ for all $s \in [0,1]$ and $\mu = 1$. This implies that $G(s)$ has infinitely many fixed points in the interval $[0,1]$. Assume that $\pi_k \neq \delta_{1k}$. Since G is convex, the two curves $y = G(s)$ and $y = s$ can intersect at either one or two points. The derivative of $G(s)$ at $s = 1$ distinguishes these two cases.





1. If $\mu \leq 1$, then $G'(1) \leq 1$. Since G' is strictly increasing in s, $G'(s) < G'(1) = 1$ for $0 < s < 1$. Let $h(s) = s - G(s)$. Then $h'(s) = 1 - G'(s) > 0$ for $0 < s < 1$. But h is increasing and $h(1) = 0$, so $h(s) < 0$ and $s < G(s)$ for $0 < s < 1$. Then, $y = G(s)$ lies above $y = s$ for $0 < s < 1$, and $s = 1$ is the only point of intersection. Thus, $P(T_0 < \infty) = 1$.
2. If $\mu > 1$, then $G'(1) > 1$, then $h(0) = 0 - G(0) = -\pi_0 = -P(X = 0) < 0$ since $P(X = 0) = 0$ contradicts convexity. Also, $h'(1) = 1 - G'(1) = 1 - \mu < 0$. Thus, $h(s)$ is decreasing at $s = 1$. Since $h(1) = 0$, there exists $0 < t < 1$, such that $h(t) > 0$. It follows that there exists a fixed point $s_2 = G(s_2)$ satisfying $0 < s_2 < 1$.

□

Example 11: Computing Extinction Probabilities

Consider a branching process with,

$$Z_0 = 1 \quad \text{and} \quad \vec{\pi} = \begin{pmatrix} 1/3 & 1/3 & 1/3 \end{pmatrix}$$

where $\vec{\pi}$ is the offspring distribution. The curves,

$$y = s \quad \text{and} \quad y = G(s) = \frac{1}{3}(1 + s + s^2)$$

intersect at $s = 1$. Therefore, $\mu \leq 1$, and the extinction probability is $P(T_0 < \infty) = 1$. We can also compute μ explicitly,

$$\mu = \frac{1}{3}(0 + 1 \cdot s + 2 \cdot s)\Big|_{s=1} = 1$$

Example 12: Computing Extinction Probabilities

Consider a branching process with,

$$Z_0 = 1 \quad \text{and} \quad \vec{\pi} \sim \text{Po}(\mu)$$

where $\vec{\pi}$ is the offspring distribution. Recall that,

$$G(s) = e^{\mu(s-1)}$$

Solving $s = e^{\mu(s-1)}$ numerically by iteration,

$G_1(0)$	$G_2(0)$	$G_3(0)$	$G_4(0)$	$G_10(0)$	$G_{15}(0)$
0.135335	0.177403	0.192975	0.199079	0.203169	0.203187

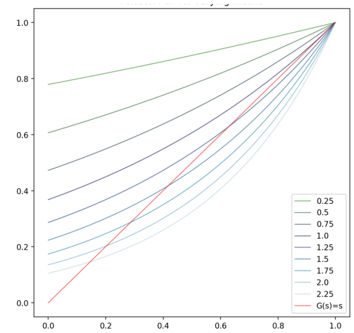


Figure 3: A Poisson probability generating function with various means μ.

Poisson Processes

Definition 1

Definition 108 (Counting Function). *A* ***counting process*** $(N_t)_{t\geq 0}$ *is a collection of random variables with values in* $\mathbb{N}_0$ *such that,*

$$N_t \geq N_s \quad \forall t \geq s \geq 0$$

and $\lim_{t\to s^+} N_t = N_s$ *for all* $s \in \mathbb{R}$ *(right-continuity*[19]*).*

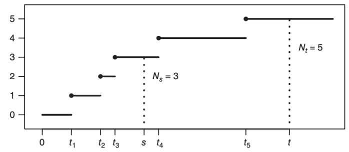


Figure 4: Counting process.

There are three equivalent definitions of a Poisson process, each of which gives special insights into the stochastic model.

Definition 109 (Poisson Process – 1a). *A* ***Poisson process*** *with parameter* $\lambda \geq 0$ *is a counting process* $(N_t)_{t\geq 0}$ *satisfying,*

1. $N_0 = 0$
2. $N_t - N_s \sim Po(\lambda \cdot (t-s))$ *for all* $t > s > 0$
3. $N_t - N_s$ *is independent of* N_r *for all* $t > s > 0$ *and* $0 \leq r \leq s$

where $N_t - N_s$ *is the number of events that have occured in* $(s, t]$.

The parameter λ is called the **rate** because $\mathbb{E}[N_t] = t \cdot \lambda$,

$$\mathbb{E}[N_t] = \mathbb{E}[N_t - N_0] = \lambda \cdot t$$

Example 13: Poisson Process

Jana sends Ioan 10 text messages per hour after 10am. We want to find the probability that Ioan has exactly 18 texts by noon and 70 texts at 5pm. This problem can be modelled by a Poisson process with rate 10, where the desired probability is,

$$P(N_2 = 18, N_7 = 70)$$

---

[19] If $0 \leq s < t$, then $N_t - N_s$ is the number of events in the interval $(s, t]$

By definition,

$$\{N_2 = 18, N_7 = 70\} = \{N_2 = 18, N_7 - N_2 = 52\}$$

Since $[0, 2]$ and $(2, 7]$ are disjoint,

$$\begin{aligned} P\left(N_2 = 18, N_7 = 70\right) &= P\left(N_2 = 18, N_7 - N_2 = 52\right) \\ &= P\left(N_2 = 18\right) \cdot P\left(N_7 - N_2 = 52\right) \\ &= P\left(N_2 = 18\right) \cdot P\left(N_5 = 52\right) \\ &= \left(\frac{e^{-20} \cdot 20^{18}}{18!}\right) \cdot \left(\frac{e^{-50} \cdot 50^{52}}{52!}\right) \\ &= 0.0045 \end{aligned}$$

Definition 110 (Poisson Process – 1b). *A **Poisson process** with parameter $\lambda \geq 0$ is a counting process $(N_t)_{t\geq 0}$ with,*

$$N_0 = 0 \quad \textit{and} \quad N_t - N_s \sim Po(\lambda \cdot (t - s))$$

for all $t > s > 0$, conditionally on N_r for $0 \leq r \leq s$.

Lemma 111. $N_t - N_s \sim N_{t-s}$[20]

Proof. $N_{t-s} - N_0 \sim \mathrm{Po}(\lambda \cdot (t - s - 0)) = \mathrm{Po}(\lambda \cdot (t - s))$. ☐

Theorem 112. *Let $(N_t)_{t\geq 0}$ be a Poisson process with parameter λ. Then,*

$$(N_{t+s} - N_s)_{t\geq 0} \sim Po(\lambda) \quad \textit{for } s > 0$$

and $(N_{t+s} - N_s)_{t\geq 0}$ is a Poisson process.

Proof. We need to show that the translated process is probabilistically equivalent to the original process. It suffices to show that,

$$Y_t := N_{t+s} - N_s$$

is a Poisson Process (Definition 1b). Clearly, $Y_0 = 0$. Now,

$$\begin{aligned} Y_t - Y_r &= N_{t+s} - N_s - (N_{r+s} - N_s) \\ &= N_{t+s} - N_{r+s} \end{aligned}$$

so $Y_t - Y_r \sim \mathrm{Po}(\lambda \cdot (t - r))$ given (N_q) for $0 \leq q \leq r$. Furthermore, (N_q) contains all information in $(Y_q) = (N_{q+s} - N_s)$ for $0 \leq q \leq r$, ☐

Definition 2

Definition 113 (Interarrival Time). *The **interarrival** times $(X_k)_{k\geq 0}$ are the times between consecutive jumps of a counting process.*

---

[20] Be careful when thinking about this conditionally.

Definition 114 (Arrival Time). *The* ***arrival*** *times* $(S_k)_{k\geq 0}$ *are the times,*

$$\lim_{t\to s^+} N_t \neq \lim_{t\to s^-} N_t$$

at which $(N_t)_{t\geq 0}$ *increases.*

Definition 115 (Poisson Process – 2). *Let* $(X_n)_{n\geq 0}$ *be a sequence of i.i.d exponential random variables with parameter* λ*. For* $t > 0$,

$$N_t = \max\{n \mid X_1 + \cdots X_n \leq t\}$$

with $N_0 = 0$*. Then,* $(N_t)_{t\geq 0}$ *is a* ***Poisson process*** *with parameter* λ*.*

$$S_n = \sum_{i=1}^{n} X_i \quad n \in \mathbb{N}$$

defines a sequence (S_n) *of arrival times of the process, where* S_k *is the time of the kth arrival. The interrarival time between* $k-1$ *and* k *is,*

$$X_k = S_k - S_{k-1} \quad k \in \mathbb{N} \quad \textit{with } S_0 = 0$$

Lemma 116. *If* $(N_t)_{t\geq 0}$ *is a rate* $\lambda \cdot t$ *Poisson Process, as in the interarrival definition, then* $N_t \sim Po(\lambda t)$ *for all* $t \geq 0$*.*

Proof. Let (N_t) be a rate $\lambda \cdot t$ Poisson Process. Then,

$$P(N_t = k) = P(\underbrace{\{S_k \leq t\} \cap \{S_{k+1} > t\}}_{:=A}) \quad (k \in \mathbb{N})$$

By the Total Probability Rule,

$$\begin{aligned} P(N_t = k) &= \mathbb{E}[P(A \mid S_k)] \\ &= \mathbb{E}\left[\mathbb{1}_{\{S_k \leq t\}} \cdot P(A \mid S_k) + \mathbb{1}_{\{S_k > t\}} \cdot P(A \mid S_k)\right] \\ &= \mathbb{E}\left[\mathbb{1}_{\{S_k \leq t\}} \cdot P(\{S_k \leq t\} \cap \{S_{k+1} > t\} \mid S_k)\right] \\ &= \mathbb{E}\left[\mathbb{1}_{\{S_k \leq t\}} \cdot P(S_{k+1} > t \mid S_k)\right] \\ &= \mathbb{E}\left[\mathbb{1}_{\{S_k \leq t\}} \cdot P(S_{k+1} - S_k > t - S_k \mid S_k)\right] \\ &= \mathbb{E}\left[\mathbb{1}_{\{S_k \leq t\}} \cdot P(X_{k+1} > t - S_k \mid S_k)\right] \end{aligned}$$

Applying Property $(\star)$ of the exponential distribution,

$$\begin{aligned} \mathbb{E}\left[\mathbb{1}_{\{S_k \leq t\}} \cdot P(X_{k+1} > t - S_k \mid S_k)\right] &= \int_0^t e^{-\lambda(t-x)} f(x)dx \\ &= \int_0^t e^{-\lambda(t-x)} \frac{\lambda^k x^{k-1}}{(k-1)!} e^{-\lambda x} dx \\ &= \frac{\lambda^k \cdot t^k \cdot e^{-\lambda t}}{k!} \end{aligned}$$

where the density $f(x)$ of S_k can be found using the PGF[21]. $\square$

Note on Arrival Time:
$S_K - S_{k-1} = X_k \quad$ (where $S_0 = 0$)

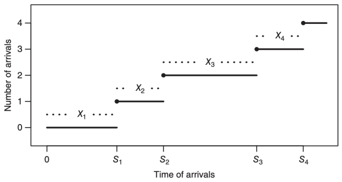


Figure 5: Arrival times $S_1, S_2, \cdots$ and interarrival times $X_1, X_2, \cdots$.

Note on Exponential Distribution.

Let $X_i \sim \text{Exp}(\lambda_i)$. Then,

$P(X \geq t) = e^{-\lambda \cdot t}$

$\mathbb{E}[f(X)] = \int_0^\infty \lambda e^{-\lambda t} f(t) dt (\star)$

$\mathbb{E}[X] = \frac{1}{\lambda}$

$G(s) = \mathbb{E}[r^X] = \int_0^\infty \lambda e^{-\lambda t} r^t dt = \frac{\lambda}{\lambda - \log r}$

$P\left(\min\{X_1, \cdots, X_n\} > t\right) = e^{-(\lambda_1 + \ldots + \lambda_n)t}$

$P\left(M = X_i\right) = \frac{\lambda_i}{\lambda_1 + \cdots + \lambda_n}$ where $M = \min_i\{X_i\}$

$P(X > s + t \mid X > s) = P(X > t)$ for $s, t > 0$

---

[21] S_k is Gamma(k, λ) distributed, so

$$f(x) = \frac{\lambda^k x^{k-1}}{(k-1)!} e^{-\lambda x}$$

Definition 117 (Memoryless). *A random variable X is* ***memoryless*** *if,*

$$P(X > s + t \mid X > s) = P(X > t)$$

Lemma 118. *The exponential distribution is memoryless*[22]*,*

Proof. Let $X \sim \text{Exp}(\lambda)$. Then for all $t > s > 0$,

$$P(X > t + s \mid X > s) = \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}} = e^{-\lambda t}$$

Therefore, $X - s \sim \text{Exp}(\lambda)$ for $X > s$. $\square$

Lemma 119. *The sum of independent exponentials is exponential.*

Proof. Let $X_i \sim \text{Exp}(\lambda_i)$. Then,

$$\begin{aligned} P\left(\min\left(X_1, \ldots, X_n\right) > t\right) &= P\left(X_1 > t, \ldots, X_n > t\right) \\ &= P\left(X_1 > t\right) \ldots P\left(X_n > t\right) \\ &= e^{-\lambda_1 t} \ldots e^{-\lambda_n t} \\ &= e^{-(\lambda_1 + \ldots + \lambda_n)t} \end{aligned}$$

$\square$

Advantages of Definition 1:

- Ease of construction and calculation
- Explicit law for interarrival times

Advantages of Definition 2:

- Independence of increments
- Explicit law for statistics of N_t

Example 14: Poisson Process

Buses arrive at a bus stop according to a Poisson process with parameter $\lambda = 6$. Suppose that you arrive at 1pm. Then,

1. Probability of waiting at least 15 minutes

$$P\left(\underbrace{S_1}_{=X_1} > \frac{1}{4}\right) = \text{Exp}\left(-\frac{3}{2}\right)$$

2. Probability that exactly 3 buses arrive in the next hour

$$P(N_1 = 3) = \frac{6^3}{e^6 \cdot 3!}$$

3. Expected time to wait for the bus

$$\mathbb{E}\left[\underbrace{S_1}_{=X_1}\right] = \frac{1}{6} = 10 \text{ min}$$

4. 18 buses arrive between 12:50pm and 1:00pm. The expected time to wait for the bus does not depend on the past

$$\mathbb{E}\left[\underbrace{S_1}_{=X_1}\right] = \frac{1}{6} = 10 \text{ min}$$

---

[22] In fact, it is the only continuous distribution that is memoryless.

Definition 3

Definition 120 (Poisson Process – Definition 3). *A **Poisson process** with parameter λ is a counting process $(N_t)_{t\geq 0}$ satisfying*[23]*,*

1. $N_0 = 0$
2. (N_t) *has stationary and independent increment*
3. $P(N_h = 0) = 1 - \lambda h + o(h)$
4. $P(N_h = 1) = \lambda h + o(h)$
5. $P(N_h > 1) = o(h)$

where $f(h) = o(g(h))$ *means that,*

$$\lim_{h\to 0} \frac{f(h)}{g(h)} = 0$$

Applications of Poisson Processes

Thinning and Superposition

Definition 121 (Thinning Poisson Process). *Let $(N_t)_{t\geq 0}$ be a Poisson process with parameter λ. Assume that each arrival, independent of other arrivals, is marked as a "Type k" ($k \in [n]$) event with probability p_k (where $\sum p_k$ = 1). Let $N_t^{(k)}$ be the number of "Type k" events in $[0, t]$. Then,*

$$\left(N_t^{(k)}\right)_{t\geq 0} \text{ is a Poisson process with rate } \lambda p_k$$
$$\left(N_t^{(i)}\right)_{t\geq 0} \text{ and } \left(N_t^{(j)}\right)_{t\geq 0} \text{ are independent } (0 \leq i \neq j \leq n)$$

*Each process is called a **thinned Poisson process**.*

Note on Thinning:
Let Z be a Poisson process with parameter λ. Suppose that $(X_i)_{i\geq 0}$ is a sequence of i.i.d Bernoulli trials with success parameter p. If (X_i) is independent of (N_t), then $Y = \sum_{i=1}^{Z} X_i \sim \text{Po}(\lambda p)$. To see this, compute the probability generating function of Y.

Definition 122 (Superposition Process Poisson). *Assume that,*

$$\left(N_t^{(1)}\right)_{t\geq 0}, \cdots, \left(N_t^{(n)}\right)_{t\geq 0}$$

are n independent Poisson processes with parameters $\lambda_1, \cdots, \lambda_n$. Let,

$$N = N_t^{(1)} + \cdots + N_t^{(n)}$$

for $t \geq 0$. Then, $(N_t)_{t\geq 0} \sim Po(t \cdot (\lambda_1 + \cdots + \lambda_n))$.

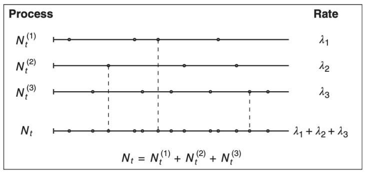


Figure 6.7 The N_t process is the superposition of $N_t^{(1)}$, $N_t^{(2)}$, and $N_t^{(3)}$.

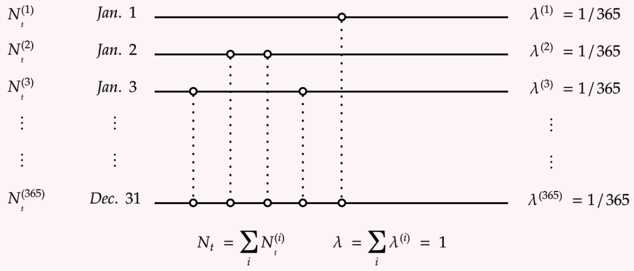
Example 15: Birthday Problem

The **Birthday Problem** asks: "If people enter a room one by one, how many people are in the room the first time two people share a birthday, ignoring year and leap days?"

---

[23] There cannot be infinitely many arrivals in a finite interval, and in an infinitesimal interval there may occur at most one event.

This problem can be embedded in a superposition of a Poisson process. People enter a room according to a Poisson process $(N_t)_{t\geq 1}$ with rate $\lambda = 1$. Each person is independently and uniformly marked with one of 365 birthdays.



Let $(X_i)_{i\geq 0}$ be the interarrival sequence for the process of people entering the room. The X_i are i.i.d exponential with mean 1. Let T be the first time when two people in the room share the same birthday. If K people are in the room at that time,

$$T = \sum_{i=1}^{K} X_i$$

(X_i) is independent of K. Taking the expectation,

$$\mathbb{E}[T] = \mathbb{E}[K] \cdot \underbrace{\mathbb{E}[X_1]}_{=1} = \mathbb{E}[K]$$

Let Z_k be the time when the second person marked with birthday k enters the room. Then, the first time that two people in the room have the same birthday is,

$$T = \min_{1\leq k\leq 365} Z_k$$

Equivalently, Z_k is the arrival time of the second event of a Poisson process. Moreover, Z_k has a gamma distribution with parameters and density,

$$n = 2 \qquad \lambda = \frac{1}{365} \qquad f(t) = \frac{te^{-t/365}}{365^2}(t > 0)$$

The cumulative distribution function is,

$$P\left(Z_1 \leq t\right) = \int_0^t \frac{se^{-s/365}}{365^2} ds = 1 - \frac{e^{-t/365}(365+t)}{365}$$

This gives,

$$\begin{aligned} P(T > t) &= P\left(\min_{1 \le k \le 365} Z_k > t\right) \\ &= P(Z_1 > t, \cdots, Z_{365} > t) \\ &= P\left(Z_1 > t\right)^{365} \\ &= \left(1 + \frac{t}{265}\right)^{365} e^{-t} \quad (t > 0) \end{aligned}$$

Therefore, the desired birthday expectation is,

$$E(K) = E(T) = \int_0^\infty P(T > t) dt = \int_0^\infty \left(1 + \frac{t}{365}\right)^{365} e^{-t} dt$$

Poissonization and Depoissonization

Poisson processes can be used to prove theorems for discrete time processes. Given $(X_k)_{k \ge 0}$, do the following:

1. **(Poissonize)** Let $X_t^* = X_{N_t}$ for a Poisson process $(N_t)_{t \ge 0}$ to embed the discrete time process (X_k) in continuous time.
2. **(Analyze)** Show that X_t^* has the desired property.
3. **(Depoissonize)** Transfer the chain to the discrete time process.

Theorem 123 (Recurrence via Poissonization)**.** *Let $(X_k)_{k \ge 0}$ be a time-homogeneous Markov chain. If $(N_t)_{t \ge 0}$ is a Poisson process with $\lambda = 1$, then a state x is **recurrent** if and only if,*

$$\int_0^\infty P\left(X_{N_t} = x\right) dt = \infty$$

Proof. It suffices to show the following,

$$\int_0^\infty P\left(X_{N_t} = x\right) dt = \sum_{k=0}^\infty p_k(x, x)$$

Since $P\left(X_k = x\right) = P\left(X_{N_t} = x \mid N_t = k\right)$,

$$\begin{aligned}
\int_0^\infty P\left(X_{N_t} = x\right) dt &= \int_0^\infty \sum_{k=0}^\infty P(N_t = x) \cdot P\left(X_{N_t} = x \mid N_t = k\right) dt \\
&= \sum_{k=0}^\infty P\left(X_{N_t} = x \mid N_t = k\right) \cdot \int_0^\infty P(N_t = x) dt \\
&= \sum_{k=0}^\infty P\left(X_k = x\right) \cdot \int_0^\infty P(N_t = x) dt \\
&= \sum_{k=0}^\infty p_k(x,x) \cdot \underbrace{\int_0^\infty \frac{e^{-t}t^k}{k!} dt}_{=1} \\
&= \sum_{k=0}^\infty p_k(x,x)
\end{aligned}$$

☐

Example 16: Poissonized Simple Random Walk on $\mathbb{Z}$

Let $(X_k)_{k\geq 0}$ be a simple random walk on $\mathbb{Z}^d$.

1. **(Poissonize)** Let (N_t) be a Poisson process with $\lambda = 1$. Assume that each arrival is marked "L_t" if the chain moves to the left, and "R_t" if the chain moves to the right. Then,

$$\begin{aligned}
(N_t^{(L_t)})_{t\geq 0} &\sim \text{Po}\left(\frac{\lambda \cdot t}{2}\right) = \text{Po}\left(\frac{t}{2}\right) \\
(N_t^{(R_t)})_{t\geq 0} &\sim \text{Po}\left(\frac{\lambda \cdot t}{2}\right) = \text{Po}\left(\frac{t}{2}\right)
\end{aligned}$$

Moreover, $(N_t^{(L_t)})$ and $(N_t^{(R_t)})$ are independent. Thus,

$$X_t^* := X_{N_t} = (N_t^{(R_t)})_{t\geq 0} - (N_t^{(L_t)})_{t\geq 0} \sim \underbrace{\text{Po}\left(\frac{t}{2}\right)}_{+1(\text{"}R_t\text{"})} - \underbrace{\text{Po}\left(\frac{t}{2}\right)}_{-1(\text{"}L_t\text{"})}$$

2. **(Analyze)** We want to determine if X_t^* is recurrent,

$$\begin{aligned}
P\left(X_{N_t} = 0\right) &= P(L_t = R_t) \\
&= \sum_k^\infty P(L_t = k \mid R_t = k) \cdot P(R_t = k) \\
&= \sum_{k=0}^\infty \left(\frac{e^{-t/2}(t/2)^k}{(k!)}\right)^2 \quad \text{by independence} \\
&= e^{-t} \cdot \mathcal{I}_0(t) \quad \text{where } \mathcal{I}_0(t) \text{ is the Bessel function}
\end{aligned}$$

Therefore, $P(X_{N_t} = 0) \cdot \sqrt{2\pi t} \to 1$ as $t \to \infty$ using the Stirling Formula or Bessel function properties. Consequently,

$$\int_0^\infty P\left(X_{N_t} = 0\right) dt = \infty \quad \text{because} \quad \int_1^\infty \frac{dt}{\sqrt{2\pi t}} = \infty$$

3. **(Depoissonize)** Apply "Recurrence via Depoissonization".

Example 17: Poissonized Simple Random Walk on $\mathbb{Z}^d$

Let $(X_k)_{k\geq 0}$ be a simple random walk on $\mathbb{Z}^d$.

1. **(Poissonize)** Let (N_t) be a Poisson process with $\lambda = 1$. Applying Thinning as in the example on $\mathbb{Z}$,

$$X_k^* := X_{N_t} = \left(X_{N_t}^{(1)}, \cdots, X_{N_t}^{(d)}\right)$$

where $(X_{N_t}^{(i)})$ and $(X_{N_t}^{(j)})$ are independent if $i \neq j$, and,

$$(X_{N_t}^{(i)}) \sim \text{Po}\left(\frac{t}{d}\right) \qquad \text{for all } i \in [n]$$

2. **(Analyze)** We want to determine if X_t^* is recurrent,

$$\begin{aligned} P\left(X_{N_t} = \vec{0}\right) &= P\left(X_{N_t}^{(1)} = 0\right)^d \text{ by independence} \\ &= \left(\sum_{k=0}^{\infty}\left(\frac{e^{-t/2d}(t/2d)^k}{(k!)}\right)^2\right)^d \end{aligned}$$

Therefore, $\int_1^\infty \left(\frac{1}{\sqrt{2\pi t/d}}\right)^d dt < \infty$ if and only if $d \geq 3$ as

$$\left(\sqrt{\frac{2\pi t}{d}}\right)^d \cdot P\left(X_{N_t} = \vec{0}\right) \to 1$$

3. **(Depoissonize)** Apply "Recurrence via Poissonization".

Order Statistics

If a Poisson process contains exactly n events in $[0, t]$, then the unordered times of those events are uniformly distributed on $[0, t]$.

Remark 124 (Conditional on 1 Event). $P\left(S_1 \leq s \mid N_t = 1\right) = \frac{s}{t}$.

Proof. Using the definition of Conditional Probability,

$$\begin{aligned} P\left(S_1 \leq s \mid N_t = 1\right) &= \frac{P\left(S_1 \leq s, N_t = 1\right)}{P\left(N_t = 1\right)} \\ &= \frac{P\left(N_s = 1, N_t = 1\right)}{P\left(N_t = 1\right)} \\ &= \frac{P\left(N_s = 1, N_t - N_s = 0\right)}{P\left(N_t = 1\right)} \\ &= \frac{P\left(N_s = 1\right) \cdot P\left(N_{t-s} = 0\right)}{P\left(N_t = 1\right)} \\ &= \frac{e^{-\lambda s} \lambda s e^{-\lambda(t-s)}}{e^{-\lambda t} \lambda t} \\ &= \frac{s}{t} \end{aligned}$$

since $N_{t-s} \sim \mathrm{Po}(\lambda \cdot (t-s))$. ☐

Definition 125 (Order Statistic). *Let $U_1, \cdots, U_n$ be an i.i.d sequence of Unif$([0,t])$ random variables. Their joint density function is,*

$$f_{U_1,\ldots,U_n}\left(u_1, \ldots, u_n\right) = \frac{1}{t^n}$$

for $0 \leq u_1, \ldots, u_n \leq t$. Arrange U_i in increasing order,

$$U_{(1)} \leq U_{(2)} \leq \cdots \leq U_{(n)}$$

$U_(k)$ is the kth smallest of the U_i. The ordered sequence,

$$\left(U_{(1)}, \ldots, U_{(n)}\right)$$

*is the **order statistics** of the original sequence. Its joint density is,*

$$f_{U_{(1)},\ldots,U_{(n)}}\left(u_1, \ldots, u_n\right) = \frac{n!}{t^n}$$

$0 \leq u_1 < \cdots < u_n \leq t$.

Theorem 126 (Order Statistics via Poissonization). *Let $S_1, S_2, \cdots$ be the arrival times of a Poisson process with parameter λ. Conditional on $N_t = n$, the joint distribution of $(S_1, \cdots, S_n)$ is the distribution of the order statistics of n i.i.d uniform random variables on $[0,t]$.*

$$f\left(s_1, \ldots, s_n\right) = \frac{n!}{t^n}$$

for $0 < s_1 < \cdots < s_n < t$. Equivalently, let $U_1, \cdots, U_n$ be an i.i.d sequence of Unif$([0,t])$ random variables. Then, conditional on $N_t = n$,

$$\left(S_1, \ldots, S_n\right) \text{ and } \left(U_{(1)}, \ldots, U_{(n)}\right)$$

have the same distribution.

Proof. See Dobrow 6.5 (p.245). ☐

Corollary 127. *Results for arrival times offer a new method for simulating a Poisson process with parameter λ on an interval $[0, t]$:*

1. *Simulate the number of arrivals N in $[0, t]$ from $Po(\lambda \cdot t)$*
2. *Generate N i.i.d random variables uniformly distributed on $(0, t)$*
3. *Sort the variables in increasing order to give the Poisson arrival times*

Spatial Poisson Processes

The spatial Poisson process is a model for the distribution of events in a two- or higher-dimensional space[24]. For $d \geq 1$ and $A \subseteq \mathbb{R}^d$, let N_A denote the number of points in the set A. We write $|A|$ for the size of A (i.e., area in $\mathbb{R}^2$ and volume in $\mathbb{R}^3$).

Definition 128 (Spatial Poisson Process)**.** *A collection of random variables $(N_A)_{A \subseteq \mathbb{R}^d}$ is a* ***spatial Poisson process*** *with parameter λ if,*

1. *$N_A \sim Po(\lambda \cdot |A|)$ for each bounded set $A \subseteq \mathbb{R}^d$*
2. *N_A and N_B are independent random variables if A and B are disjoint*

The definition of a spatial Poisson process can be generalized to a Poisson random measure as follows,

Definition 129 (Poisson Random Measure)**.** *The* ***Poisson random measure*** *is the unique function $N : \mathcal{B} \to \mathbb{N}_0$ so that for any $A \in \mathcal{B}$,*

1. *If N_A is the number of points in A, then,*

$$N_A \sim Po\left(\int_A f(x)dx\right)$$

2. *If $A_1, \cdots, A_k \in \mathcal{B}$ are disjoint, compact sets,*

$$(N(A_1), N(A_2), \cdots, N(A_k))$$

are independent, Poisson distributed random variables with parameters,

$$\int_{A_j} f(x)dx \quad 1 \leq j \leq k$$

where $f : \mathbb{R}^d \to \mathbb{R}$ is called the "intensity" of the process.

Remark 130. *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a continuous, non-negative function.*

1. ***(Number of Points in** I**)** For any closed rectangle $I \subset \mathbb{R}^d$,*

$$N_I \sim Po\left(\int_I f(x)dx\right)$$

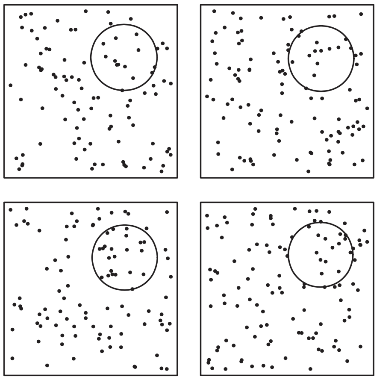


Figure 6: Samples of a spatial Poisson process with parameter $\lambda = 100$ on the square $[0, 1] \times [0, 1]$.

[24] The uniform distribution arises for the spatial process in a similar way to how it does for the one-dimensional Poisson process. Given a bounded set $A \subseteq \mathbb{R}^d$, conditional on there being n points in A, the location of the points are uniformly distributed on A.

2. ***(Location of Points in** I**)** Define an i.i.d collection X_I of points,*

$$(X_j)_{j\in[1,N_I]} \quad \textit{with} \quad X_j \sim P(X_j \in B) = \frac{\int_B f(x)dx}{\int_I f(x)dx}$$

3. ***(Spread of Points in** $\mathbb{R}^d$**)** Repeat over a partition into rectangles,*

$$X = \cup_I X_I \quad \textit{and} \quad N_A := |\{X \cap A\}| \quad \textit{for each } A \in \mathcal{B}$$

N is a Poisson random measure with intensity f[25].

Definition 131 (Non-Homogeneous Poisson Process). *A counting process $(N_t)_{t\geq 0}$ is **non-homogeneous Poisson** with intensity $\lambda(t)$ if,*

1. $N_0 = 0$
2. *For all $t > 0$, N_t has a Poisson distribution with mean,*

$$E\left(N_t\right) = \int_0^t \lambda(x)dx$$

3. *For $0 \leq q < r \leq s < t$. $N_r - N_q$ and $N_t - N_s$ are independent*

Continuous-Time Markov Chains

Holding Times

Definition 132 (Continuous-Time Markov Property). *The **Markov property** for continuous-time chains $(X_t)_{t\geq 0}$ with discrete state space S is,*

$$P\left(X_t = j \mid (X_u)_{0\leq u<s} \text{ and } X_s = i\right) = P\left(X_t = j \mid X_s = i\right)$$

for all $t \geq s$ and states $i, j \in S$.

Definition 133 (Time-Homogeneous). *A continuous-time Markov chain $(X_t)_{t\geq 0}$ with discrete state space S is **time-homogeneous** if,*

$$P\left(X_{t+s} = j \mid X_s = i\right) = P\left(X_t = j \mid X_0 = i\right)$$

for $s \geq 0$. The transition probabilities can be arranged in a function,

$$\mathbf{P}_{ij}(t) = P\left(X_t = j \mid X_0 = i\right)$$

which is the analog of $p_t(i, j)$ in the discrete setting[26].

Definition 134 (Chapman-Kolmogorov). *A continuous-time Markov chain $(X_t)_{t\geq 0}$ with transition function $\mathbf{P}(t)$ satisfies that,*

$$\mathbf{P}(s+t) = \mathbf{P}(s) \cdot \mathbf{P}(t) \textit{ for } s, t \geq 0.$$

---

[25] **Notes on Spatial Processes:**

1. If $d = 1$ and f is constant at $\lambda > 0$, then $N_{[0,t]} \leftrightarrow N_t$ is the Poisson process with rate λ.
2. If $d > 1$ and f is constant at $\lambda > 0$, then this is the homogeneous spatial Poisson process.

[26] $\mathbf{P}(t)$ is not the transition matrix $\tilde{\mathbf{P}}$ of the embedded chain, since $t \in \mathbb{R}$.

Proof. By conditioning on X_s,

$$\begin{aligned}
\boldsymbol{P}_{ij}(s+t) &= P\left(X_{s+t} = j \mid X_0 = i\right) \\
&= \sum_k P\left(X_{s+t} = j \mid X_s = k, X_0 = i\right) \cdot P\left(X_s = k \mid X_0 = i\right) \\
&= \sum_k P\left(X_{s+t} = j \mid X_s = k\right) \cdot P\left(X_s = k \mid X_0 = i\right) \\
&= \sum_k P\left(X_t = j \mid X_0 = k\right) \cdot P\left(X_s = k \mid X_0 = i\right) \\
&= \sum_k P_{ik}(s) \cdot P_{kj}(t) \\
&= [\mathbf{P}(s) \cdot \mathbf{P}(t)]_{ij}
\end{aligned}$$

□

Definition 135 (Holding Time)**.** *The* ***holding time*** *T_i at a state i is the length of time that a continuous-time Markov chain started in i stays in i before transitioning to a new state.*

Theorem 136. *Let T_i be the holding time at state i. Then $T_i \sim Exp(\lambda_i)$.*

Proof. The exponential distribution is the only continuous distribution that is memoryless, so it suffices to prove that T_i is memoryless. Let $s, t \geq 0$. Suppose that the chain starts in i. Then,

$$\{T_i > s\} = \{X_u = i \text{ for } u \in [0, s]\}$$

Moreover, $\{T_i > s + t\}$ implies that $\{T_i > s\}$, so,

$$P(T_i > s + t \mid X_0 = i) = P(\underbrace{\{T_i > s + t\}}_{A} \cap \underbrace{\{T_i > s\}}_{B} \mid X_0 = i)$$

Applying Bayes' Law with the conditional probability measure,

$$P(A \cap B \mid X_0 = i) = P(A \mid B \cap \{X_0 = i\}) \cdot P(B \mid X_0 = i)$$

and using homogeneity and the Markov property,

$$\begin{aligned}
&= P(T_i > s + t \mid \{T_i > s\} \cap \{X_0 = i\}) \cdot P(T_i > s \mid X_0 = i) \\
&= P\left(T_i > s + t \mid X_u = i \text{ for } u \in [0, s]\right) \cdot P\left(T_i > s \mid X_0 = i\right) \\
&= P\left(T_i > s + t \mid X_s = i\right) \cdot P\left(T_i > s \mid X_0 = i\right) \\
&= P\left(T_i > t \mid X_0 = i\right) \cdot P\left(T_i > s \mid X_0 = i\right)
\end{aligned}$$

shows that T_i satisfies the definition of memoryless,

$$P(T_i > s + t \mid X_0 = i) = P\left(T_i > t \mid X_0 = i\right) \cdot P\left(T_i > s \mid X_0 = i\right)$$

□

Definition 137 (Absorbing State). *A state i is* ***absorbing*** *if the parameter of the exponential distribution for the holding time T_i is zero,*

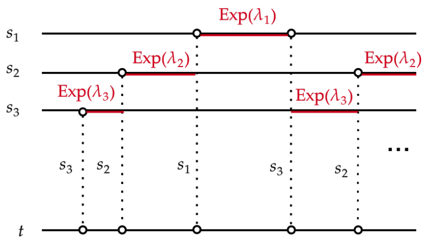
$$\mathbb{E}[T_i] = \frac{1}{0} = \infty$$

Definition 138. *A continuous-time Markov chain is* ***explosive*** *at a time $t \in \mathbb{R}^+$ if there are infinitely-many transitions in all neighborhoods*[27]*,*

$$(t - \epsilon, t) \quad \text{for } \epsilon > 0 \text{ arbitrary}$$

The evolution of a continuous-time Markov chain which is neither absorbing nor explosive can be described as follows,

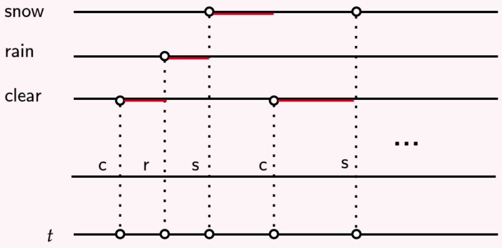
1. Starting from i, the process stays in i for an exponentially distributed length of time, which is $1/q_i$ on average
2. The chain hits a new state $j \neq i$, with probability p_{ij}
3. The process stays in j for an exponentially distributed length of time, which is $1/q_j$ on average. It then hits a new state $l \neq j$



Example 18: Continuous-Time Weather Chain

Define the state space $S = \{\texttt{rain}, \ \texttt{snow}, \ \texttt{clear}\}$. Assume that,

1. Rain lasts, on average, 3 hours
2. Snow lasts, on average, 6 hours
3. Clear weather lasts, on average, 12 hours



---

[27] In Dobrow, a state $i \in S$ is called explosive if the holding time parameter λ_i of T_i is infinite, i.e.,

$$\mathbb{E}[T_i] = \frac{1}{\infty} = 0$$

Changes in weather states are described by the matrix,

$$\tilde{P} = \begin{array}{c} \\ \\ \\ \end{array}\begin{array}{ccc|c} \texttt{rain} & \texttt{snow} & \texttt{clear} & \\ 0 & 1/2 & 1/2 & \texttt{rain} \\ 3/4 & 0 & 1/4 & \texttt{snow} \\ 1/4 & 3/4 & 0 & \texttt{clear} \end{array}$$

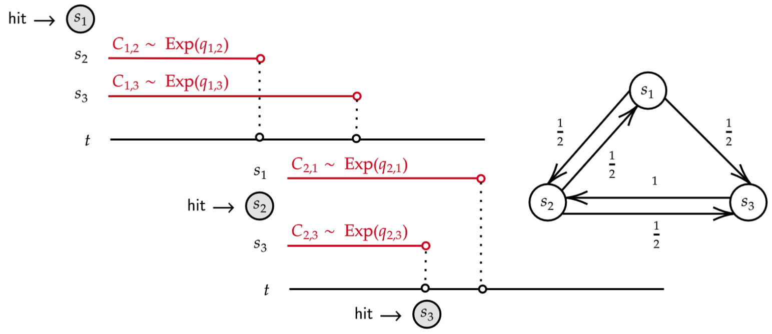
Let X_t be the weather at time t. Then, $(X_t)_{t \geq 0}$ is a continuous-time Markov chain. Moreover, $\tilde{\mathbf{P}}$, as well as the exponential parameters $(\lambda_r, \lambda_s, \lambda_c) = (1/3, 1/6, 1/12)$, specify $\mathbf{P}(t)$, i.e., $P\left(X_{t_1} = i_1, \ldots, X_{t_n} = i_n\right)$ for $n \geq 1$, states s_i and times $t_i \geq 0$.

Definition 139 (Embedded Chain). *A sequence $(Y_k)_{k \geq 0}$ is the* ***embedded chain*** *of a continuous-time process (X_k) if Y_k is the kth state visited.*

Remark 140. *The transition matrix $\tilde{\mathbf{P}}$ of the embedded chain of a continuous-time process is a stochastic matrix whose diagonal entries are zero*[28].

A continuous-time Markov chain can also be described by specifying transition rates between pairs of states. Suppose that for for each state i, there are independent alarm clocks associated with each of the states that the process can visit after i. Then,

1. If j can be hit from i, the alarm C_{ij} associated with (i, j) will ring after an exponentially distributed length of time with parameter q_{ij}
2. The minimum time for one C_{ij} $(i \neq j)$ to finish ringing is the minimum of independent exponentials, which was proven to be exponential with parameter $q_i = \sum_k q_{ik}$
3. When the process first hits i, the clocks start simulataneously, where the first alarm that rings determines the next state
4. If the (i, j) clock rings first and the process moves to j, a new set of exponential alarm clocks are started with rates $q_{j1}, q_{j2}, \cdots$



---

[28] The process can never transition from state i to state i. It remains in i for an exponential amount of time, after which it transitions to another state j $(j \neq i)$.

Infinitesimal Generator

Definition 141 (Rate Matrix)**.** *The **rate matrix** $(\mathbf{Q})$ for a continuous-time Markov chain is defined as follows,*

$$(\mathbf{Q})_{ij} := \begin{cases} q_i \cdot \tilde{\mathbf{P}}_{ij} & i \neq j \\ 0 & i = j \end{cases}$$
$$= \begin{cases} q_{ij} & i \neq j \\ 0 & i = j \end{cases}$$

Note on Generator Matrix:
A continuous-time Markov chain $(X_t)_{t\geq 0}$ with transition function $\mathbf{P}(t)$ and generator $\mathbf{Q}$ satisfies that,

1. $\mathbf{P}'(t) = \mathbf{P}(t) \cdot \mathbf{Q}$
2. $\mathbf{P}'(t) = \mathbf{Q} \cdot \mathbf{P}(t)$

These are called the "Kolmogorov Forward, Backward Equations".

Definition 142 (Generator Matrix)**.** *The **generator matrix** $\mathbf{Q}$ for a continuous-time Markov chain is defined as follows*[29]*,*

$$\mathbf{Q}_{ij} := \begin{cases} q_{ij} & i \neq j \\ \underbrace{-q_i}_{-\sum_k q_{ik}} & i = j \end{cases}$$

Corollary 143. *The generator is not a stochastic matrix. Diagonal entries are negative, entries can be greater than 1, and rows sum to 0.*

Corollary 144. *A continuous-time Markov chain $(X_t)_{t\geq 0}$ satisfies that,*

$$\tilde{\boldsymbol{P}}_{ij} := \begin{cases} P\left(C_{i,j} = \hat{C}\right) = \frac{q_{ij}}{q_i} = \frac{q_{ij}}{\sum_k q_{ik}} & i \neq j \\ 0 & i = j \end{cases}$$

where $\tilde{\boldsymbol{P}}$ is the embedded chain, and $\{C_{ij} = \hat{C}\}$ is the event that C_{ij} is the first alarm that rings and determines the next state[30]*.*

Corollary 145. *A continuous-time Markov chain $(X_t)_{t\geq 0}$ with transition function $\mathbf{P}(t)$ and generator $\mathbf{Q}$ satisfies that,*

$$\boldsymbol{P}(t) = e^{t\mathbf{Q}} = \sum_{n=0}^{\infty} \frac{1}{n!}(t\boldsymbol{Q})^n = \boldsymbol{I} + t\boldsymbol{Q} + \frac{t^2}{2}Q^2 + \frac{t^3}{6}Q^3 + \cdots$$

Note on Matrix Exponential:
Let $\mathbf{A}$ be a $k \times k$ matrix. Then,

$$e^A = \sum_{n=0}^{\infty} \frac{1}{n!} A^n$$
$$= I + A + \frac{1}{2}A^2 + \cdots$$

Classification of States

For characterizing the states of a continuous-time Markov chain, the definitions of accessibility, communication, and irreducibility are defined as in the discrete case. For example,

Definition 146 (Irreducible)**.** *A continuous-time Markov chain with transition function $\mathbf{P}(t)$ is **irreducible** if for all $i, j \in S$,*

$$\mathbf{P}_{ij}(t) > 0 \text{ for some } t > 0$$

Lemma 147. *If $\mathbf{P}_{ij}(t) > 0$ for some $t > 0$, then $\mathbf{P}_{ij}(t) > 0$ for all $t > 0$.*

---

[29] If the smallest signed value of a generator matrix $\mathbf{Q}$ is finite, then we say the chain is "bounded rate". Moreover, any finite chain is bounded rate.

[30] The holding time parameters λ_i are equal to q_i.

Proof. Suppose that $\mathbf{P}_{ij}(t) > 0$ for some $t > 0$. Then there exists a path from i to j in the embedded chain, and, since the exponential distribution is continuous, for any time s there is positive probability of reaching j from i in s time units[31]. ☐

Corollary 148. *All states of a continuous-time Markov chain are aperiodic.*

Stationary Distributions

Definition 149 (Stationary Distribution). *A probability distribution $\vec{\pi}$ is a* ***stationary probability distribution*** *for a continuous-time chain if*[32],

$$\vec{\pi} = \vec{\pi} \cdot \mathbf{P}(t)$$

Corollary 150. *The stationary distribution $\vec{\pi}$ is not the same as the stationary distribution of the embedded chain, $\vec{\psi}$*[33]*. However, for all j,*

$$\psi(j) = \frac{\pi_j q_j}{\sum_k \pi(k) q_k} \quad \pi_j = \frac{\psi(j)/q_j}{\sum_k \psi(k)/q_k}$$

Theorem 151 (Fundamental Limit Theorem). *Let $(X_t)_{t \geq 0}$ be a finite, irreducible, continuous-time Markov chain with transition function $\boldsymbol{P}(t)$. Then, there exists a unique stationary distribution $\vec{\pi}$, which is the limiting distribution of the chain. That is, for all $j \in S$,*

$$\lim_{t \to \infty} \mathbf{P}_{ij}(t) = \pi_j \textit{ for all } i$$

Equivalently,

$$\lim_{t \to \infty} \boldsymbol{P}(t) = \mathbf{\Pi}$$

where $\mathbf{\Pi}$ is a matrix all of whose rows are equal to $\vec{\pi}$.

Proof. The proof of the Fundamental Limit Theorem is omitted. ☐

Theorem 152. *A probability distribution $\vec{\pi}$ is a stationary distribution of a continuous-time Markov chain with generator $\mathbf{Q}$ if and only if,*

$$\vec{\pi}\mathbf{Q} = \vec{0} \quad \iff \quad \sum_i \pi_i Q_{ij} = 0 \textit{ for all } j \in S$$

Proof. Assume that $\vec{\pi} = \vec{\pi}\mathbf{P}(t)$ for all $t \geq 0$. Differentiating at $t = 0$,

$$\vec{0} = \vec{\pi}\mathbf{P}'(0) = \vec{\pi}\mathbf{Q}$$

Conversely, assume that $\vec{\pi}\mathbf{Q} = \vec{0}$. Right-multiplying by $\mathbf{P}(t)$,

$$\vec{0} = \vec{\pi}\mathbf{Q}\mathbf{P}(t) = \vec{\pi}\mathbf{P}'(t) \text{ for } t \geq 0$$

by the Kolmogorov backward equation. This implies that $\vec{\pi}\mathbf{P}(t)$ is constant, that is, $\vec{\pi}\mathbf{P}(t) = \vec{\pi}\mathbf{P}(0)$ for all t. But $\mathbf{P}(0) = I$[34], so,

$$\vec{\pi}\mathbf{P}(t) = \vec{\pi}\mathbf{P}(0) = \vec{\pi}I = \vec{\pi}$$

☐

---

[31] Formally, for $s \geq 0$, this means that,

1. $\mathbf{P}_{ij}(t+s) > 0$
2. $\mathbf{P}_{ij}(t-s) > 0$

[32] As in the discrete case, the limiting distribution, if it exists, is a stationary distribution. However, the converse is not necessarily true and depends on the class structure of the chain.

[33] π_i is the long-term proportion of **time** that the process spends in state i. Conversely, $\psi(i)$ is the long-term proportion of **transitions** that the process makes into state i.

[34] Observe that,

$$\mathbf{P}_{ij}(0) = P(X_t = j \mid X_0 = i) = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases}$$

Poisson Subordination

Definition 153 (Subordination)**.** *Let $(N_t)_{t\geq 0}$ be a Poisson process with parameter λ. Let $(Y_t)_{t\geq 0}$ be a finite-state, irreducible, discrete-time process with transition matrix* $\mathbf{R}$*. Define a continuous-time process $(X_t)_{t\geq 0}$ by $X_t := Y_{N_t}$*[35]*. The process (X_t) is* ***subordinated to a Poisson process****.*

Remark 154. *Let $\mathbf{P}(t)$ be the transition function of (X_t). Then,*

$$\begin{aligned}\mathbf{P}_{ij}(t) &= P(X_t = j \mid X_0 = i)\\ &= \sum_{k=0}^{\infty} P\left(X_t = j \mid \{N_t = k\} \cap \{X_0 = i\}\right) \cdot P\left(N_t = k \mid X_0 = i\right)\\ &= \sum_{k=0}^{\infty} P\left(Y_k = j \mid \{N_t = k\} \cap \{X_0 = i\}\right) \cdot P\left(N_t = k\right)\\ &= \sum_{k=0}^{\infty} P\left(Y_k = j \mid Y_0 = i\right) \cdot P\left(N_t = k\right)\\ &= \sum_{k=0}^{\infty} \mathbf{R}_{ij}^k \cdot \frac{e^{-\lambda t}(\lambda t)^k}{k!}\end{aligned}$$

We have just seen how to construct a continuous-time Markov chain from a discrete-time chain and a Poisson process. We will now see that many continuous-time chains can be represented as chains subordinated to a Poisson process.

Remark 155. *Let $\mathbf{Q}$ be the generator of a continuous-time Markov chain with holding time parameters $\{q_i\}$. If $q_i \leq \lambda$ for all i, then we can define,*

$$\mathbf{R} = \frac{1}{\lambda}\mathbf{Q} + \mathbf{I} \text{ where } \lambda = \max_i q_i$$

which is a stochastic matrix. The transition function is,

$$\begin{aligned}\boldsymbol{P}(t) = e^{t\boldsymbol{Q}} &= e^{-\lambda t}e^{t\boldsymbol{Q}}e^{\lambda t} = e^{-\lambda t}e^{t(\boldsymbol{Q}+\lambda \boldsymbol{I})}\\ &= e^{-\lambda t}\sum_{k=0}^{\infty}\frac{1}{k!}t^k(\boldsymbol{Q}+\lambda\boldsymbol{I})^k\\ &= \sum_{k=0}^{\infty}\left(\frac{1}{\lambda}\boldsymbol{Q}+\boldsymbol{I}\right)^k\frac{e^{-\lambda t}(\lambda t)^k}{k!}\\ &= \sum_{k=0}^{\infty}\boldsymbol{R}^k\frac{e^{-\lambda t}(\lambda t)^k}{k!}\end{aligned}$$

$\boldsymbol{R}$ is not the matrix of the embedded Markov chain:

$$\tilde{\mathbf{P}}_{ij} = \begin{cases} q_{ij}/q_i, & \text{for } i \neq j \\ 0, & \text{for } i = j \end{cases} \qquad \mathbf{R}_{ij} = \begin{cases} q_{ij}/\lambda, & \text{for } i \neq j \\ 1 - q_i/\lambda, & \text{for } i = j \end{cases}$$

Theorem 156. *For a Markov chain subordinated to a Poisson process, the discrete* $\mathbf{R}$*-chain has the same stationary distribution as the original chain.*

Note on Poisson Subordination:

1. From state i, wait an exponential length of time with rate λ
2. Flip a coin with $P(H) = \frac{q_i}{\lambda}$
3. If heads, transition according to $\mathbf{R}$. Otherwise, stay at i and repeat

---

[35] Transitions for the X_t process occur at the arrival times of the Poisson process. From state i, the process holds an exponentially distributed amount of time with parameter λ, and then transitions to j with probability $\mathbf{R}_{ij}$

Proof. $\vec{\pi}\mathbf{Q} = \vec{\pi}\lambda(\boldsymbol{R} - \boldsymbol{I}) = \lambda\vec{\pi}\boldsymbol{R} - \lambda\vec{\pi}$. ☐

Corollary 157. *The following are equivalent,*

1. $\vec{\pi}\boldsymbol{P}(t) = \vec{\pi}$ *for all* $t \geq 0$
2. $\vec{\pi}\boldsymbol{Q} = \vec{0}$
3. $\vec{\pi}\tilde{\boldsymbol{R}} = \vec{\pi}$ *for some* $\lambda = \max_i q_i$

Example 19: Totally Asymmetric Exclusion Process

The dynamics of the **totally asymmetric simple exclusion process** on $\mathbb{Z}$ are as follows,

1. The alarm clock for a particle rings as a Poisson process of rate 1 at each site independently
2. When the bell at site i rings, if there is a particle at site i and a hole at site $i-1$, they exchange

Equivalently, each particle in the system tries to jump to the left at rate 1, and the jump succeeds whenever the site to the left is unoccupied. Since it is possible for infinitely many particles to move instantaneously, and the minimum of infinitely-many exponential clocks is zero, there is no embedded chain and holding representation for this chain.

Birth-and-Death Chains

Definition 158 (Yule Process)**.** *Let Y_t be the population size at time t. The* ***Yule process*** $(Y_t)_{t\geq 0}$ *is a continuous-time branching process where*[36]*,*

1. $Y_0 = 1$*, and the rate of the process is* $q_{x,x+1} = x \cdot \beta$
2. *Each individual gives birth to an offspring at a constant rate* β
3. *Each individual gives birth independently of other individuals*
4. *The distribution of each division time is Exp*(β)

Note on Yule Process:
The Yule process ia a birth-and-death process in which the birth rate is $x \cdot \lambda$ and the death rate 0.

Remark 159. *A Yule process* $(Y_t)_{t\geq 0}$ *satisfies the following identity,*

$$Y_{t+\tau} \sim Y_t^{(1)} + Y_t^{(2)} \quad \text{and} \quad \tau \sim Exp(\beta)$$

where τ *is the time of the first division, and* $Y_t^{(1)}, Y_t^{(2)}$ *are Yule processes.*

(Y_t) cannot be written as a Poisson subordinated discrete time walk. However, it exists for all time without exploding. Conversely, if $q_{x,x+1} = \beta \cdot x^2$, then the process explodes in finite time.

[36] The rate represents the expected time until the next division. Thus, $P(\min\{C_1, \cdots, C_x\} > t) = e^{-x\beta}$.

Remark 160. *Define the following,*

$$\begin{aligned}
&\textit{Arrival } i \textit{ is the birth of member } i+1 \textit{ of the population}\\
&\hat{C}_i \textit{ is the minimum time for one } C_i \textit{ to finish ringing}\\
&Z_i \textit{ is the time of arrival } i\\
&T_n := \min\{t \mid Y_t = n\}
\end{aligned}$$

We need to compute $P(T_n \leq t)$*,*

$$\begin{aligned}
P(T_n \leq t) &= P\left(\sum_{i=1}^{n-1} \hat{C}_i \leq t\right) \textit{ where } \hat{C}_i \sim Exp(i \cdot \beta)\\
&= P\left(\bigcap_{i=1}^{n-1} \{Z_i \leq t\}\right) \textit{ where } Z_i \sim Exp(\beta)\\
&= \prod_{i=1}^{n-1} P(Z_i \leq t) \textit{ since } (Z_i)_{i \in [n-1]} \textit{ i.i.d}\\
&= (1-e^{-t\beta})^{n-1} \textit{ by induction}
\end{aligned}$$

Remark 161. *We can write* $Y_t = \sum_{i=1}^{\infty} \mathbb{1}_{\{T_j \leq t\}}$*. Then,*

$$\mathbb{E}[Y_t] = \mathbb{E}\left[\sum_{i=1}^{\infty} \mathbb{1}_{\{T_j \leq t\}}\right] = \sum_{i=1}^{\infty} P(T_j \leq t) = \sum_{i=1}^{\infty} \left(1 - e^{-t\cdot\beta}\right)^{i-1} = e^{t\cdot\beta}$$

using the sum of a geometric series.

Definition 162 (Birth-and-Death)**.** ***Birth-and-death processes*** *are a class of time-reversible, continuous-time Markov chains. They satisfy,*

1. *Births occur from* i *to* $i+1$ *with rate* β
2. *Deaths occur from* i *to* $i-1$ *with rate* 1

making $q_{x,x+1} = \beta \cdot x$ *and* $q_{x,x-1} = x$[37]*.*

Theorem 163. *If* $\beta > 1$*, then the probability of extinction is less than 1.*

Proof. The proof is by a Poisson Process argument, where,

$$\left(\frac{\beta}{1+\beta}\right) \quad \text{and} \quad \left(\frac{1}{1+\beta}\right)$$

are the parameters of each thinned process. Computing the number of divisions before death gives that $\mathbb{E}[Y] = \beta$. □

The following questions are equivalent,

1. How many times is the Exp(β) birth clock C of an individual the minimum before the Exp(1) death clock rings?
2. How many offspring does the individual produce?

Corollary 164. *Let* (Y_t) *be a birth-and-death process with birth rate* β *and death rate* 1*. If* $\beta \leq 1$*, the underlying branching process goes extinct,*

$$Y_t \xrightarrow[t\to\infty]{a.s} 0 \quad \textit{where } 0 \textit{ is an absorbing state}$$

---

[37] This represents an exponential 1 clock and a population of size x.

Conversely, when $\beta > 1$*, there exists a random variable* Y_∞*,*

$$Y_t \underset{t\to\infty}{\overset{a.s}{\to}} Y_\infty$$

and $P(Y_\infty = 0)$ *is the extinction probability of the branching process,*

$$P(Y_\infty = \infty) = 1 - P(Y_\infty = 0)$$

Definition 165. *Let* (Y_t) *be an irreducible continuous-time Markov chain.* (Y_t) *is null-recurrent if and only if* $\mathbb{E}[\tau_0^+] = \infty$[38].

Corollary 166. *The underlying discrete-time behavior does not determine positive or null recurrence for the chain. The holding time parameters can be made sufficiently large to see this.*

Theorem 167. *An irreducible, continuous-time Markov chain is positive recurrent if and only if there exists a stationary probability vector* $\vec{\pi}$*. That is,* $\sum_{i\in S}\pi_i = 1$ *and* $\vec{\pi}\mathbf{Q} = 0$*, where* $\mathbf{Q}$ *is the generator of the chain.*

Proof. The proof uses Martingales, and it was not seen in class. ☐

Example 20: Reflected Birth-and-Death Chain

We can modify the birth-and-death chain to make the boundary condition at zero reflect. Equivantly, we define,

$$\mathbf{Q}'_{x,y} = \begin{cases} q_{x,y} & x \neq 0 \\ 1 & x = 0 \text{ and } y \neq 0 \\ 0 & x = 0 \text{ and } y = 0 \end{cases}$$

Q. Is the reflected birth-and-death chain recurrent?
The modified chain is recurrent if and only if $\beta \leq 1$ since the process goes extinct, i.e., reaches zero, with probability 1.

Q. Is the reflected birth-and-death chain positive-recurrent?
The modified chain is positive recurrent if and only if $\beta < 1$.

Martingales

Examples of Martingales

Definition 168 (Martingale). *A* ***martingale*** $(M_t)_{t\geq 0}$ *is a stochastic process that satisfies, for all* $t \geq 0$*,*

1. $\mathbb{E}[M_t \mid (M_r)_{r\in[0,s]}] = M_s$ *for all* $0 \leq s \leq t$
2. $\mathbb{E}[|M_t|] < \infty$

where $\mathcal{F}_k = \sigma(M_0, \cdots, M_s) = (M_r)_{r\in[0,s]}$ *is the history of the chain*[39].

---

[38] Otherwise, (Y_t) is positive-recurrent. This definition is the same as the discrete case.

[39] The condition that $\mathbb{E}[|M_t|] < \infty$ is omitted for the examples in this section.

Corollary 169. *A discrete-time martingale (M_t) satisfies,*

1. $\mathbb{E}[M_t \mid Y_0, \cdots, M_{t-1}] = M_s$ *for all* $t \geq 0$
2. $\mathbb{E}[|M_t|] < \infty$

Remark 170. *If* $(M_n)_{n\geq 0}$ *is a martingale, then* $\mathbb{E}[M_n]$ *is constant.*

Proof. By the Law of Total Expectation, for all $0 \leq s \leq t$,

$$\mathbb{E}[M_t] = \mathbb{E}[\mathbb{E}[M_0, \cdots, M_s]] = \mathbb{E}[M_s]$$

That is, $\mathbb{E}[M_t] = \mathbb{E}[M_0]$ for all t. □

Example 21: Simple Random Walk

The simple symmetric random walk $S_n := \sum_{i=1}^n X_i$ has,

$$X_i = \begin{cases} +1 & \text{with probability } 1/2 \\ -1 & \text{with probability } 1/2 \end{cases}$$

for $n \geq 1$ with $S_0 = 0$. Then,

$$\begin{aligned} \mathbb{E}[S_{n+1} \mid S_0, \cdots, S_n] &= \mathbb{E}[X_{n+1} + S_n \mid S_0, \cdots, S_n] \\ &= \mathbb{E}[X_{n+1} \mid S_0, \cdots, S_n] + \underbrace{\mathbb{E}[S_n \mid S_0, \cdots, S_n]}_{\mathbb{E}[g(X)|X]=g(X)} \\ &= \underbrace{\mathbb{E}[X_{n+1}]}_{=0} + S_n \\ &= S_n \end{aligned}$$

since X_{n+1} is independent of $X_1, \cdots, X_n$ and consequently X_{n+1} is independent of $S_0, \cdots, S_n$. Next, we prove that,

$$\mathbb{E}[|S_n|] = \mathbb{E}\left[\left|\sum_{i=1}^n X_i\right|\right] \leq \mathbb{E}\left[\sum_{i=1}^n |X_i|\right] = \sum_{i=1}^n \mathbb{E}[|X_i|] = n < \infty$$

Example 22: Biased Random Walk

Let $p + q = 1$. The biased random walk $S_n := \sum_{i=1}^n X_i$ has,

$$X_i = \begin{cases} +1 & \text{with probability } p \\ -1 & \text{with probability } q \end{cases}$$

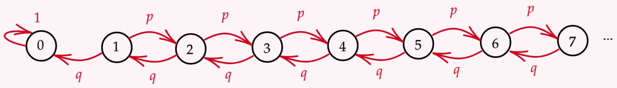
for $n \geq 1$ with $S_0 = 0$. While it is not a martingale,

$$\begin{aligned} \mathbb{E}[S_{n+1} \mid S_0, \cdots, S_n] &= \underbrace{\mathbb{E}[X_{n+1}]}_{p-q} + S_n \\ &= (p - q) + S_n \end{aligned}$$

the modified process $(\hat{S}_n)_{n\geq 1}$ is a martingale,

$$\hat{S}_n = S_n - (p-q)\cdot n$$

Example 23: Biased Random Walk on $\mathbb{N}_0$ with Absorption



Let $(S_n)_{n\geq 0}$ be the biased random walk on $\mathbb{N}_0$ with an absorbing boundary at zero. $(\hat{S}_n)$ is not a martingale,

$$\begin{aligned}\mathbb{E}[\hat{S}_{n+1} \mid \hat{S}_0, \cdots, \hat{S}_n] &= \mathbb{E}[S_{n+1} - (p-q)(n+1) \mid \hat{S}_0, \cdots, \hat{S}_n] \\ &= \mathbb{E}[X_{n+1}] + S_n - (p-q)(n+1) \\ &= S_n - (p-q)(n+1) + (p-q)\mathbb{1}_{\{Y_n \neq 0\}}\end{aligned}$$

Define $(S'_n)_{n\geq 0}$ by $S'_n := S_n - (p-q)\cdot \min\{n, T\}$ for $T := \inf\{n \mid S_n = 0\}$. Then (S'_n) is a martingale.

Example 24: Absorption Time for Biased Random Walk

If $p < q$, then (S_n) is absorbed with probability 1,

$$Y_k \underset{k\to\infty}{\rightarrow} 0$$

Therefore,

$$\begin{aligned}\mathbb{E}[S'_0] &= \mathbb{E}[S'_k] \\ &= \mathbb{E}[S_n] - \mathbb{E}[(p-q)\cdot \min\{n, T\}]\end{aligned}$$

Taking $k \to \infty$ gives the absorption time $\mathbb{E}[T] = \frac{\mathbb{E}[X_0]}{(q-p)}$.

Example 25: Branching Process

Let $(Z_k)_{k\geq 0}$ be a branching process with offspring distribution $X = (x_1, \cdots,)$. Put $\mathbb{E}[X] = \mu$. Then,

$$M_k := \frac{Z_k}{\mu^k}$$

is a martingale. Since $Z_{k+1} \sim \sum_{j=1}^{Z_k} X_j$,

$$\begin{aligned}
\mathbb{E}[M_{k+1} \mid M_1, \cdots, M_k] &= \mathbb{E}\left[\frac{Z_{k+1}}{\mu^{k+1}} \,\middle|\, M_1, \cdots, M_k \right] \\
&= \mathbb{E}\left[\frac{\sum_{j=1}^{Z_k} X_j}{\mu^k \cdot \mu} \,\middle|\, Z_k \right] \\
&= \frac{1}{\mu^k \cdot \mu} \sum_{j=1}^{Z_k} \mathbb{E}[X_j] \text{ since } \mu, k \text{ fixed} \\
&= \frac{1}{\mu^k \cdot \mu} \sum_{j=1}^{Z_k} \mu \\
&= \frac{Z_k}{\mu^k} \\
&= M_k
\end{aligned}$$

Example 26: Pólya's Urn

If (R_k, B_k) are the number of red and blue balls at time k.

$$M_k := \frac{R_k}{R_k + B_k}$$

is a martingale. Recall that,

$$\begin{aligned}
&(R_k, B_k) \to (R_k, B_{k+1}) \text{ with probability } \frac{B_k}{R_0 + B_0 + k} \\
&(R_k, B_k) \to (R_{k+1}, B_k) \text{ with probability } \frac{R_k}{R_0 + B_0 + k} \\
&(B_k, k) \to (B_k, k+1) \text{ with probability } \frac{R_0 + B_0 + k - B_n}{R_0 + B_0 + k} \\
&(B_k, k) \to (B_{k+1}, k+1) \text{ with probability } \frac{B_k}{R_0 + B_0 + k}
\end{aligned}$$

Since (M_k) is a Markov Chain,

$$\begin{aligned}
\mathbb{E}[M_{k+1} \mid M_1, \cdots, M_k] &= \mathbb{E}\left[\frac{R_{k+1}}{R_{k+1} + B_{k+1}} \,\middle|\, (R_k, B_k) \right] \\
&= \mathbb{E}\left[\frac{R_{k+1}}{B_0 + R_0 + k + 1} \,\middle|\, (R_k, B_k) \right] \\
&= \frac{1}{B_0 + R_0 + k + 1} \cdot \mathbb{E}\left[R_{k+1} \mid (R_k, B_k) \right]
\end{aligned}$$

where A_{k+1} be the event that the $(k+1)$th ball is red, given that the number of red and blue balls at time k are R_k and B_k.

$$\underbrace{P\left(R_{k+1} = R_k + 1\right)}_{P(A_{k+1} | (R_k, B_k)) = \mathbb{E}\left[\mathbb{1}_{A_{k+1}}\right]} \cdot \frac{R_k + 1}{R_k + B_k + 1}$$

$$+ \quad \underbrace{P\left(R_{k+1} = R_k\right)}_{P(A_{k+1}^c \mid (R_k, B_k)) = \mathbb{E}\left[\mathbb{1}_{A_{k+1}^c}\right]} \quad \cdot \frac{R_k}{R_k + B_k + 1}$$

is the desired expectation. Simplifying,

$$\begin{aligned}
&= \frac{R_k^2 + R_k + B_k R_k}{(R_0 + B_0 + k) \cdot (R_0 + B_0 + k + 1)} \\
&= \frac{R_k \cdot (R_k + B_k + 1)}{(R_k + B_k) \cdot (R_k + B_k + 1)} \\
&= \frac{R_k}{R_k + B_k} \\
&= M_k
\end{aligned}$$

Example 27: Absorption Probabilities

Let (Y_k) be a Markov chain with an absorbing state i. Then,

$$M_k := P(A_i \mid Y_k)$$

is a martingale, where A_i is the event that the chain is absorbed at i. By the Total Conditional Law of Expectation,

$$\begin{aligned}
\mathbb{E}[M_{k+1} \mid M_0, \cdots, M_k] &= \mathbb{E}[P(A_i \mid Y_{k+1}) \mid M_0, \cdots, M_k] \\
&= \mathbb{E}[P(A_i \mid Y_{k+1}) \mid Y_k] \\
&= \mathbb{E}[P(A_i \mid Y_{k+1}, Y_k) \mid Y_k] \\
&= P(A_i \mid Y_k) \\
&= M_k
\end{aligned}$$

where the third inequality follows by the Markov Property.

Limit Theorems

Definition 171 (Bracket). *The **bracket** of a martingale $(M_k)_{k \geq 0}$ is*[40]*,*

$$\langle M_k \rangle := \sum_{n=0}^{k} \mathbb{E}\left((M_{n+1} - M_n)^2 \mid \mathcal{F}_n \right)$$

Example 28: Simple Random Walk

We saw that the simple random walk on $\mathbb{Z}_0$ is a martingale.

$$(M_{n+1} - M_n)^2 = 1 \quad \text{for all } n \in \mathbb{N}_0$$

so $\langle M_k \rangle = k$ for all $k \in \mathbb{N}$.

[40] This is effectively the accumulated variance of the process (Y_k).

Remark 172. *Let* $(M_k)_{k\geq 0}$ *be a martingale. Then,*

$$\mathbb{E}\left\langle M_k \right\rangle = \mathbb{E}[(M_{n+1} - M_0)^2]$$

Proof. By definition,

$$\begin{aligned}\mathbb{E}\left[(M_{n+1} - M_n)^2 \mid \mathcal{F}_n\right] &= \mathbb{E}\left[M_{n+1}^2 + M_n^2 - 2 \cdot M_{n+1}M_n \mid \mathcal{F}_n\right] \\ &= \mathbb{E}\left[M_{n+1}^2 \mid \mathcal{F}_n\right] + M_n^2 + 2 \cdot M_n^2\end{aligned}$$

Therefore,

$$\begin{aligned}\mathbb{E}\left[\langle M_k \rangle\right] &= \mathbb{E}\left[\sum_{n=0}^{n} \mathbb{E}\left[(M_{n+1} - M_n)^2 \mid \mathcal{F}_n\right]\right] \\ &= \sum_{n=0}^{n} \mathbb{E}\left[\mathbb{E}\left[(M_{n+1} - M_n)^2 \mid \mathcal{F}_n\right]\right] \\ &= \sum_{n=0}^{n} \mathbb{E}\left[\mathbb{E}\left[M_{n+1}^2 \mid \mathcal{F}_n\right] + M_n^2 - 2 \cdot M_n^2\right] \\ &= \sum_{n=0}^{n} \mathbb{E}[M_{n+1}^2] + \mathbb{E}[M_n^2] - 2 \cdot \mathbb{E}[M_n^2] \\ &= \sum_{n=0}^{n} \mathbb{E}[M_{n+1}^2] - \mathbb{E}[M_n^2] \\ &= \mathbb{E}[M_{n+1}^2] - \mathbb{E}[M_0^2] \\ &= \mathbb{E}[(M_{n+1} - M_0)^2]\end{aligned}$$

☐

Example 29: Pólya's Urn (Bracket Process)

Since $\frac{R_n}{n+n_0} \geq \frac{R_n}{n+n_0+1}$, we have that,

$$\mathbb{E}\left[\left(\frac{R_{n+1}}{n+1+n_0} - \frac{R_n}{n+n_0}\right)^2 \middle| \mathcal{F}_n\right] \leq \frac{1}{(n+1+n_0)^2}$$

Implying that, for all $n \in \mathbb{N}$,

$$\langle M_n \rangle \leq \sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6} < \infty$$

Theorem 173. *Let* $(M_k)_{k\geq 0}$ *be a martingale. Suppose that,*

$$P\left(\sup_n \langle M_n \rangle < \infty\right) = 1$$

Then, $\lim_{n\to\infty} M_n$ *exists. Denote it by* M_∞*. Then,*

1. $P\left(\lim_{n\to\infty} M_n = M_\infty\right) = 1$

2. $\sup \mathbb{E}[\langle M_n\rangle] < \infty$
3. $M_n = \mathbb{E}[M_\infty \mid \mathcal{F}_n]$

Proof. The proof was not given in class. □

Theorem 174. *Let $(M_k)_{k\geq 0}$ be a martingale. If $M_n \geq 0$ for all $n \in \mathbb{N}$, then there exists a random variable $M_\infty < \infty$ such that,*

$$P\left(\lim_{n\to\infty} M_n = M_\infty\right) = 1$$

Proof. The proof was not given in class. □

Example 30: Branching Processes (Bracket Process)

Let $(M_k)_{k\geq 0}$ be the martingale for a branching process $(Z_k)_{k\geq 0}$, as defined above. Then,

$$\begin{aligned}
\langle M_k\rangle &= \sum_{k=1}^{n} \mathbb{E}\left[(M_{k+1}-M_k)^2 \,\middle|\, \mathcal{F}_k\right] \\
&= \sum_{k=1}^{n} \mathbb{E}\left[\left(\frac{Z_{k+1}}{\mu^{k+1}} - \frac{Z^k}{\mu^k}\right)^2 \,\middle|\, \mathcal{F}_k\right] \\
&= \sum_{k=1}^{n} \mathbb{E}\left[\left(\frac{1}{\mu^{k+1}} \cdot \sum_{i=1}^{Z^k} X_i - \frac{Z^k}{\mu^k}\right)^2 \,\middle|\, \mathcal{F}_k\right] \\
&= \sum_{k=1}^{n} \mathbb{E}\left[\frac{1}{\mu^{2k}} \cdot \left(\sum_{i=1}^{Z^k}\left(\frac{X_i}{\mu} - 1\right)\right)^2 \,\middle|\, \mathcal{F}_k\right] \\
&= \sum_{k=1}^{n} \mathbb{E}\left[\frac{1}{\mu^{2k}} \cdot \left(\sum_{i=1}^{Z^k}\left(\frac{X_i}{\mu} - \frac{\mu}{\mu}\right)\right)^2 \,\middle|\, \mathcal{F}_k\right] \\
&= \sum_{k=1}^{n} \frac{1}{\mu^{2k}} \cdot \mathbb{E}\left[\left(\sum_{i=1}^{Z^k}\left(\frac{X_i}{\mu} - \frac{\mu}{\mu}\right)\right)^2 \,\middle|\, \mathcal{F}_k\right] \\
&= \frac{Z_n}{\mu^{2n}} \cdot \frac{\mathrm{Var}\,(X_1)}{\mu^2}
\end{aligned}$$

using the definition of the population variance. Now, we can bound $\sup \langle M_n\rangle$ since $\langle M_n\rangle$ is an increasing sequence,

$$\sup \langle M_n\rangle \leq \sum_{k=0}^{\infty} \frac{Z_k}{\mu^{2k}} \frac{\mathrm{Var}\,(X_1)}{\mu^2}$$

and use the mean exponential growth rate to conclude that,

$$\mathbb{E}\left[\sup \langle M_n\rangle\right] \leq \sum_{k=0}^{\infty} \mathbb{E}\left[\frac{Z_k}{\mu^{2k}}\right] \frac{\mathrm{Var}\,(X_1)}{\mu^2} = \sum_{k=0}^{\infty} \frac{1}{\mu^k} \frac{\mathrm{Var}\,(X_1)}{\mu^2}$$

with $\mu > 1$ and $\mathrm{Var}\,(X_1) < \infty$, the ratio M_k converges to a non-degenerate limit random variable.

10 Runs of the Pólya's Urn Process:

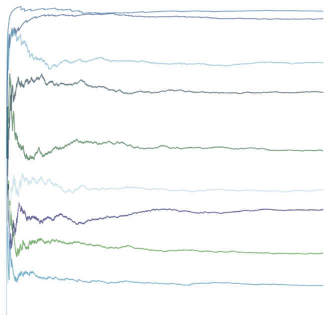


Figure 7: It was proven in class that,

$$M_n := \frac{R_n}{n} \to M_\infty$$

where $M_\infty \sim \mathrm{Unif}([0,1])$.