Stochastic Processes

MATH 447

Nicholas Hayek

Taught by Prof. Louigi Addario-Berry

CONTENTS

We assume working knowledge of probability and no knowledge of measure theory (though a grasp of analysis is essential). See these MATH 356 notes, also taught by Louigi!

I Markov Chains

Conditional expectations will be important in this course. Recall $\mathbb{E}[X|Y = y_0]$, where X, Y are random variables. If Y is continuous, writing $\mathbb{E}[X|Y = y_0] = \frac{\mathbb{P}(X, Y=y_0)}{\mathbb{P}(Y=y_0)}$, will not work. Instead, we consider the slice of the joint density function $f(x, y)$ at $y = y_0$. The result is a one dimensional function $g(x)$ which may not have probability 1. Hence, we divide by $\int g(x)$ to make it into a density function:

$$\mathbb{E}[X|Y = y_0] = \int_{\mathbb{R}} \frac{f(x, y_0)}{\int_{\mathbb{R}} f(x, y_0)dx} x dx$$

DEF 1.1 We frequently write $f_{X|Y}(x) = f(x, y)/\int_{\mathbb{R}} f(x, y)dx$, and call this the *conditional density* of X given Y. For fixed y, then, $\mathbb{E}[X|Y = y] = \mathbb{E}[Z]$, where $Z \sim f_{X|Y}$.

INTRODUCTION

Before providing definitions, we give some examples of stochastic processes:

Eg. 1.1 A simple random walk: $S_{i+1} = S_i + X_i$, where $X_i \sim \text{Ber}(p)$ and $S_0 = 0$. We might ask: does S_i ever return to 0, i.e.

$$\mathbb{P}(\exists i > 0 : S_i = 0)$$

Eg. 1.2 A branching process: as in asexual reproduction, we have an initial node. Each node n has a number of children X_n, where $\frac{X_n}{2} \sim \text{Ber}(p)$. We denote Z_i to be the number of inididuals in the i-th generation. We might ask: does Z_i ever have no children, i.e.

$$\mathbb{P}(\exists i > 0 : Z_i = 0)$$

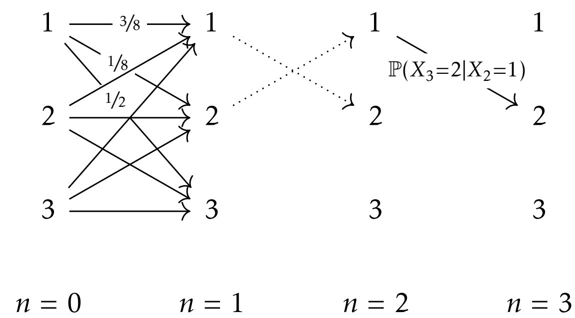
Eg. 1.3 Choose k independent random points in the square $[0, \sqrt{k}]^2$. On average, then, there is 1 point within any unit square $U \subseteq [0, \sqrt{k}]^2$.

DEF 1.2 DEF 1.3 Given a finite or countable set V, a *Markov chain* with *state space* V is a sequence $X_n : n \geq 0$ of random variables, with $X_n \in V$, such that:

$$\mathbb{P}(\underbrace{X_{n+1} = v_{n+1}}_{\text{future}} | \underbrace{X_0 = v_0, ..., X_{n-1} = v_{n-1}}_{\text{past}}, \underbrace{X_n = v_n}_{\text{present}}) = \mathbb{P}(X_{n+1} = v_{n+1}|X_n = v_n)$$

In other words, the future only depends on the past via the present. This is called the
DEF 1.4 *Markov property*.

Sometimes we allow Markov chains to be only finitely large (i.e. $0 \le n \le m$). For instance, we limit ourselves to one weekend of gambling in Las Vegas. A graphical example would look something like:



We can decompose walks on Markov chains by discrete "steps": **PROP 1.1**

$$\mathbb{P}(X_1 = v_1, ..., X_n = v_n | X_0 = v_0) = \prod_{i=1}^{n} \mathbb{P}(X_i = v_i | X_{i-1} = v_{i-1})$$

PROOF.

$$\begin{aligned}
&\mathbb{P}(X_1 = v_1, ..., X_n = v_n | X_0 = v_0) \\
&= \mathbb{P}(X_2 = v_2, ..., X_n = v_n | X_0 = v_0, X_1 = v_1)\mathbb{P}(X_1 = v_1 | X_0 = v_0) \quad \text{by cond. prob.} \\
&\vdots \\
&= \mathbb{P}(X_1 = v_1 | X_0 = v_0) \cdot \mathbb{P}(X_2 = v_2 | X_0 = v_0, X_1 = v_1) \cdots \mathbb{P}(X_n = v_n | X_0 = v_0, ..., X_{n-1} = v_{n-1}) \\
&= \mathbb{P}(X_1 = v_1 | X_0 = v_0) \cdot \mathbb{P}(X_2 = v_2 | X_1 = v_1) \cdots \mathbb{P}(X_n = v_n | X_{n-1} = v_{n-1}) \quad \text{by Markov prop}
\end{aligned}$$

□

Time-Homogeneous Markov Chains

We often write THMC

We say that a Markov chain is *time-homogeneous* if, for all $u, v \in V$ and $n \ge 0$ DEF 1.5

$$\mathbb{P}(X_{n+1} = v | X_n = u) = \mathbb{P}(X_1 = v | X_0 = u)$$

In other words, the chain's behavior is described entirely by $\mathbb{P}(X_1 = v | X_0 = u)$ for each $(v, u) \in V \times V$. In this case, we can describe such probabilities in a *transition matrix* P: DEF 1.6

$$P = \left(p_{u,v}\right)_{(u,v)\in V^2} = \left(\mathbb{P}(X_1 = v | X_0 = u)\right)_{(u,v)\in V^2}$$

Eg. 1.4 Recall the game Snakes and Ladders. A 6×6 grid is indexed $1, ..., 36$. Players start at the 1 cell. The game ends when a player reaches the 36 cell. A die roll dictates how many spots one advances. The are some directed edges between cells (increasing: "ladders", decreasing: "snakes"). One must follow these edges when one lands at its tail. Suppose a ladder exists from 11 to 27. Then

$$\mathbb{P}(X_{11} = 27 | X_{10} = 6, X_9 = 3) = \frac{1}{6} = \mathbb{P}(X_{11} = 27 | X_{10} = 6) = \mathbb{P}(X_2 = 27 | X_1 = 6)$$

We see that Snakes and Ladders is naturally modeled as a time-homogeneous Markov chain.

Eg. 1.5 Sampling without replacement is *not* a Markov chain. If we sample from $|X| = 10$, we have

$$\begin{aligned} \mathbb{P}(X_3 = a \,|X_2 = b) &= 1/9 \\ \mathbb{P}(X_3 = a \,|X_2 = b, X_1 = c) &= 1/8 \\ \mathbb{P}(X_3 = a \,|X_2 = b, X_1 = a) &= 0 \end{aligned}$$

so we do not satisfy the Markov property.

Eg. 1.6 Returning to the Snakes and Ladders example, consider $S \subseteq V$. Let $T_S =$
DEF 1.7 $\inf\{n \geq 0 : X_n \in S\}$, which we call the "*hitting time*" of S. We may ask...

- What is the average number of rounds to finite? We can write this as $\mathbb{E}[T_{\{36\}} | X_0 = 1]$.
- What is the probability of landing on 18 or 19 before the game ends? We can write this as $\mathbb{P}(T_{\{18,19\}} < T_{\{36\}} | X_0 = 1)$.
- What is the average number of visits to 6 before the game ends? We can write this as
$$\mathbb{E}[\#\{n \in [T_{\{36\}}] : X_n = 6\} | X_0 = 1]$$
- What is the expected proportion of time spent on state 5 before the game ends?
- If we allow two players, what is the probability that player 1 wins? Is this still a Markov chain?

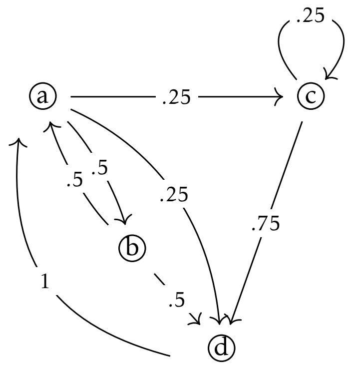
DEF 1.8 A matrix $P = \left(p_{u,v}\right)_{(u,v)\in V^2}$ is called a *stochastic matrix* if every row sums to 1, i.e.

$$\forall\ u \in V, \sum_{v \in V} p_{u,v} = 1$$

Note that any stochastic matrix is the transition matrix of some time-homogeneous Markov chain with state space V and transition probabilities

$$\mathbb{P}(X_{n+1} = v | X_n = u) = \mathbb{P}(X_1 = v | X_0 = u) = p_{u,v}$$

A directed graph, together with its stochastic matrix, can visualize any THMC:



$$\longleftrightarrow \quad \begin{bmatrix} 0 & .5 & .25 & .25 \\ .5 & 0 & 0 & .5 \\ 0 & 0 & .25 & .75 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

Eg. 1.7 Random walks on an undirected weighted graph, where edge weights dictate the proportional probability of transitioning between two states, are a special class of THMCs. In particular, given a graph $G = (V, E)$ with weights $w(e) > 0 : e \in E$, we set

$$p_{u,v} = \frac{w(\{u, v\})}{\sum_{z \in N(u)} w(\{u, z\})}$$

If there are no edges $u \leftrightarrow v$, we write $p_{u,v} = 0$.

Not every THMC can be represented by a random walk on an undirected weighted graph. In particular, see the directed graph listed above, or any transition matrix which is not symmetric.

As a concrete example, we can consider a random walk on the number line $\mathbb{Z}$, where, if $w(k, k+1) = \alpha$, $w(k-1, k) = \frac{\alpha}{2}$.

$$\cdots \overset{\frac{1}{16}}{\text{———}} -3 \overset{\frac{1}{8}}{\text{———}} -2 \overset{\frac{1}{4}}{\text{———}} -1 \overset{\frac{1}{2}}{\text{———}} 0 \overset{1}{\text{———}} 1 \overset{2}{\text{———}} 2 \overset{4}{\text{———}} 3 \overset{8}{\text{———}} \cdots$$

Multi-Step Transition Probabilities

Given a THMC $X = X_n : n \geq 0$ with a transition matrix P, we write

$$\begin{aligned} \mathbb{P}(X_2 = v | X_0 = u) &= \sum_{w \in V} \mathbb{P}(X_2 = v, X_1 = w | X_0 = u) \\ &= \sum_{w \in V} \mathbb{P}(X_1 = w | X_0 = u) \mathbb{P}(X_2 = v | X_1 = w, \cancel{X_0 = u}) \quad \text{by Markov property} \\ &= \sum_{w \in V} p_{u,w} p_{w,v} = (P^2)_{u,v} \quad \text{or write } P^2_{u,v} \end{aligned}$$

Hence, to determine a two-step transition probability, and by extension an n-step transition probability from u to w, we consider $P^n_{u,v}$. **PROP 1.2**

PROOF. See Prop 1.1 to expand probabilities, using Bayes', as needed. We get that

$$\begin{aligned}\mathbb{P}(X_n = v|X_0 = u) &= \sum_{v_1,\dots,v_{n-1}\in V} \mathbb{P}(X_1 = v_1, \dots, X_{n-1} = v_{n-1}, X_n = v|X_0 = u)\\ &= \sum_{v_1,\dots,v_{n-1}\in V} p_{u,v_1} p(v_1, v_2)\cdots p(v_{n-1}, v) = (P^n)_{u,v}\end{aligned}$$ □

PROP 1.3 Thus, if P is a stochastic matrix, then so is P^n.

PROOF. $\sum_{v\in V} P^n_{u,v} = \sum_{v\in V} \mathbb{P}(X_n = v|X_0 = u) = 1.$ □

Theorem 1.1 Markov Property

If $X_n : n \geq 0$ is a THMC with state space V, then for all $u_0, \dots, u_{n-1}, u, v \in V$,

$$\mathbb{P}(X_{n+m} = v|X_0 = u_0, \dots, X_{n-1} = u_{n-1}, X_n = u) = \mathbb{P}(X_{n+m} = v|X_n = u) = P^m_{u,v}$$

PROOF. One shows this by combining the Markov property with Prop 1.2. □

We *also* call this the Markov property.

In particular, Thm 1.1 is the Markov property combined with a claim about $P^m_{u,v}$

DEF 1.9 We say that a Markov chain has an *initial distribution* $\alpha = (\alpha_v : v \in V)$ if $\mathbb{P}(X_0 = v) = \alpha_v$ for each $v \in V$. If this is the case, we often write α as a subscript of our state probabilities. For instance,

$$\mathbb{P}_\alpha(X_n = v) = \sum_{u\in V} \mathbb{P}_\alpha(X_0 = u, X_n = v) = \sum_{u\in V} \mathbb{P}_\alpha(X_0 = u)\mathbb{P}_\alpha(X_n = v|X_0 = u) = \sum_{u\in V} \alpha_u P^n_{u,v}$$

PROP 1.4 For an event E, depending only on $X_0, \dots, X_n$, with $\mathbb{P}(X_n = u, E) > 0$, we have

$$\mathbb{P}(X_{n+m} = v|X_n = u, E) = P^m_{u,v}$$

PROOF. For any such event E, we can determine whether E occurs exactly when we know the realized values u_i of X_i for $i = 1, \dots n-1$. Hence, we may write $\mathcal{S}$ to be the set of tuples $(u_0, \dots, u_{n-1})$ that guarantee E. It follows that

$$\mathbb{P}(X_n = u, E) = \sum_{\mathbf{s}\in\mathcal{S}} \mathbb{P}(\mathbf{X} = \mathbf{s}, X_n = u)$$

Similarly, we have

$$\begin{aligned}\mathbb{P}(X_{n+m} = v, X_n = u, E) &= \sum_{\mathbf{s}\in\mathcal{S}} \mathbb{P}(X_{n+m} = v, X_n = u, \mathbf{X} = \mathbf{s}) \\ &= \sum_{\mathbf{s}\in\mathcal{S}} \mathbb{P}(X_{n+m} = v | X_n = u, \mathbf{X} = \mathbf{s})\mathbb{P}(X_n = u, \mathbf{X} = \mathbf{s}) \\ &= P^m_{u,v} \sum_{\mathbf{s}\in\mathcal{S}} \mathbb{P}(X_n = u, \mathbf{X} = \mathbf{s}) = P^m_{u,v}\mathbb{P}(X_n = u, E)\end{aligned}$$

Divide and use Bayes, and the result follows. □

If X is a THMC with transition matrix P, then, for all $k \geq 1$, $X_{kn} : n \geq 0$ is a THMC with transition matrix P^k. **PROP 1.5**

PROOF.

For any $n \neq 0$, any sequence $u_0, ..., u_{n+1} \in V$ satisfies

$$\mathbb{P}(X_{(n+1)k} = u_{n+1} | X_0 = u_0, X_k = u_1, ..., X_{nk} = u_n) = P^k_{u_n,u_{n+1}} \qquad \square$$

Theorem 1.2 Chapman-Kolmogorov

For any Markov chain X with state space V, any $m, n \geq 0$, and $u, v \in V$,

$$\mathbb{P}(X_{m+n} = v | X_0 = u) = \sum_{w\in V} \mathbb{P}(X_n = w | X_0 = u)\mathbb{P}(X_{m+n} = v | X_n = w)$$

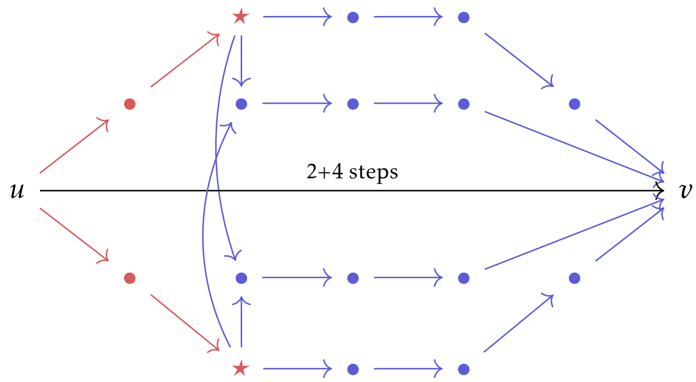
If the X is time homogeneous, then this is $P^{n+m}_{u,v}$, which agrees with Prop 1.2.

PROOF.

$$\begin{aligned}\mathbb{P}(X_{m+n} = v | X_0 = u) &= \sum_{w\in V} \mathbb{P}(X_{m+n} = v, X_n = w | X_0 = u) \\ &= \sum_{w\in V} \mathbb{P}(X_{m+n} = v | X_n = w, X_0 = u)\mathbb{P}(X_n = w | X_0 = u) \\ &= \sum_{w\in V} \mathbb{P}(X_{m+n} = v | X_n = w)\mathbb{P}(X_n = w | X_0 = u)\end{aligned}$$

Below is a visualization of the Chapman-Kolmogorov equation: to get from u to v over 6 steps, we pick some partition (e.g. $2 + 4 = 6$), and sum the probabilities of reaching v via length 2 (red) and length 4 (blue) paths.

You should count 6 such combinations in the sample below.



STATIONARY AND LIMITING DISTRIBUTIONS

DEF 1.10 Recall from probability the *law of large numbers*: if $Y_n : n \geq 1$ are IID with common mean μ, then $\frac{S_n}{n} \to \mu$ in probability, where $S_n = \sum_{i=1}^n Y_i$, i.e. $\forall \varepsilon > 0$,

$$\limsup_{n\to\infty} \mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) = 0$$

If $Y_i \in \mathbb{Z}$ is a THMC, for $k, \ell \in \mathbb{Z}$ and $i \in [n-1]$,

$$\begin{aligned} P_{k,\ell} &:= \mathbb{P}(S_{n+1} = \ell | S_n = k, S_i = u_i \ \forall i) = \mathbb{P}(Y_{n+1} = \ell - k | S_n = k, S_i = u_i \ \forall i) \\ &= \mathbb{P}(Y_{n+1} = \ell - k | Y_1 = u_1 - u_0, Y_2 = u_2 - u_1, ..., Y_n = k - u_{n-1}) \\ &= \mathbb{P}(Y_{n+1} = \ell - k) = \mathbb{P}(Y_1 = \ell - k) \end{aligned}$$

We conclude that $S_n : n \geq 0$ may be viewed as a THMC, with transition matrix P.

From now on, we denote by $\mathbb{P}_v(E)$ the probability $\mathbb{P}(E|v)$.

Eg. 1.8 A general two-state chain, with states A and B, can be described by

$$P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$$

Let $q_n = \mathbb{P}_A(X_n = A) = \mathbb{P}(X_n = A | X_0 = A)$. Then

$$\begin{aligned} q_{n+1} &= \mathbb{P}_A(X_{n+1} = A, X_n = A) + \mathbb{P}_A(X_{n+1} = A, X_n = B) \\ &= \mathbb{P}_A(X_{n+1} = A | X_n = A)\mathbb{P}_A(X_n = A) + \mathbb{P}_A(X_{n+1} = A | X_n = B)\mathbb{P}_A(X_n = B) \\ &= (1-\alpha)q_n + \beta(1-q_n) = \beta + (1-\alpha-\beta)q_n \end{aligned}$$

This recurrence has a unique solution. In particular, one can find

$$q_n = \frac{\beta}{\alpha+\beta} + (1-\alpha-\beta)^n \frac{\alpha}{\alpha+\beta}$$

It follows that $q_n \to \frac{\beta}{\alpha+\beta}$, and hence $\mathbb{P}_A(X_n = B) = 1 - q_n \to \frac{\alpha}{\alpha+\beta}$. Likewise:

$$\mathbb{P}_B(X_n = B) = \frac{\alpha}{\alpha+\beta} + (1_\alpha - \beta)^n \frac{\beta}{\alpha+\beta}$$

So $\mathbb{P}_B(X_n = B) \to \frac{\alpha}{\alpha+\beta}$.

Let $\pi := (\pi_A, \pi_B) := \left(\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta}\right)$ be the distribution of our initial state X_0, Then

$$\mathbb{P}_\pi(X_1 = A) = \pi_A \mathbb{P}_A(X_1 = A) + \pi_B \mathbb{P}_B(X_1 = A) = \pi_A$$

and, similarly, $\mathbb{P}_\pi(X_1 = B) = \pi_B$. Hence, if X_0 has initial distribution π, then X_1 also has distribution π. By induction, X_n has distribution π $\forall n \geq 0$.

When we say $X = \text{Markov}(P)$, we mean that X is a THMC with transition matrix P.

A probability distribution π is called *stationary* if $\pi P = \pi$. Similarly, a probability distribution λ is called a *limiting distribution* if, for each $u, v \in V$ DEF 1.11 DEF 1.12

$$(P^n)_{u,v} \to \lambda_v \text{ as } n \to \infty$$

In other words, $\mathbb{P}_u(X_n = v) \to \lambda_v$. Note that, for any initial distribution α, we have $\alpha P^n \to \lambda$, i.e. $(\alpha P^n)_v \to \lambda_v$, where λ is limiting. As proof:

PROOF.

$$(\alpha P^n)_v = \sum_{u \in V} \alpha_u P^n_{u,v} \to \sum_{u \in V} \alpha_u \lambda_v = \lambda_v$$

□

If λ is a limiting distribution for P, then λ is stationary for P. **PROP 1.6**

PROOF. Fix any initial distribution α, we have

$$\lambda = \lim_{n\to\infty}(\alpha P^n) = \lim_{n\to\infty}(\alpha P^{n-1}P) = (\lim_{n\to\infty}\alpha P^{n-1})P = \lambda P \qquad \square$$

Stationary distributions need not be unique, but limiting distributions are (as the limit $\lim_{n\to\infty} \alpha P^n$ is well-defined). In general, then, stationary distributions need not be limiting distributions.

When $n = 0$, $P^n = I$, which encapsulates the idea that, at timestep 0, we will be at our initial positions.

A stochastic matrix P is called *regular* if $\exists n \geq 1$ such that $P^n > 0$ on all entries. DEF 1.13

Theorem 1.3 Fundamental Theorem of Markov Chains

Every finite, regular stochastic matrix P has a limiting distribution π.

Incorporating some of the formulations above, this is equivalent to saying: *For a regular stochastic matrix, there exists a unique distribution $\pi = (\pi_v : v \in V)$ such that $\pi P = \pi$ and $\mathbb{P}_u(X_n = v) \to \pi_v$ $\forall u, v \in V$.*

In this case, there is a unique stationary distribution, and it is the unique limiting distribution.

A stationary distribution always exists!

Let $\rho = \langle 1, \ldots, 1 \rangle$. Then note that $P\rho = \rho$, since the sum of any row in P must be 1. Hence, P has eigenvalue 1. It follows that it has a left eigenvector, i.e. $\pi : \pi P = \pi$. This is exactly a stationary distribution, as long as we scale suitably such that π is a distribution.

However, then process of scaling into a distribution is non-trivial. Since π may have negative coordinates, and hence $\sum \pi_i = 0$, we must consider instead $|\pi|$, i.e. prove it is also an eigenvalue.

This is true, but requires the fact that P is stochastic

Periodicity of States

DEF 1.14 For $u, v \in V$, we say that v is *accessible* from u if $\exists n \geq 0$ such that $(P^n)_{u,v} > 0$. Equivalently, in the directed graph generated by P, there is a directed path from u to v. When v is accessible from u, we write $u \to v$.

DEF 1.15 States u and v *communicate* if $u \to v$ and $v \to u$. When u and v communicate, we write $u \leftrightarrow v$. Observe that communication is a equivalence relation. Hence, the state space V can be written as a disjoint union of mutually-communicating states, called a

DEF 1.16 *communication class*. Note that, in the directed graph generated by P, these correspond to the strongly connected components.

DEF 1.17 We say that P is *irreducible* if there is only one communication class.

Clearly, if P is regular, then it is irreducible

PROP 1.7 $u \to v \iff \mathbb{P}_u(T_{\{v\}} < \infty) > 0$.

DEF 1.18 The *period* of a state $u \in V$ is

$$d(u) := \gcd(n > 0 : P^n_{u,u} > 0)$$

DEF 1.19 If $d(u) = 1$, we call u *aperiodic*. By extension, P is aperiodic if $d(u) = 1\ \forall u \in V$, and X is aperiodic if $X = \text{Markov}(P)$ for P aperiodic.

PROP 1.8 If $u \leftrightarrow v$, then $d(u) = d(v)$.

PROOF. Let $I = \{n > 0 : P^n_{u,u} > 0\}$, and similarly J for v. Hence, $d(u) = \gcd(I)$ and $d(v) = \gcd(J)$. Let $a, b > 0$ such that $P^a_{u,v} > 0$ and $P^b_{v,u} > 0$. Then

$$P^{a+b}_{u,u} \geq P^a_{u,v} P^b_{v,u} > 0$$

$\implies a + b \in I$, so $d(u) | a + b$. Now, if $n \in J$, then

$$P^{a+b+n}_{u,u} \geq P^a_{u,v} P^n_{v,v} P^b_{v,u} > 0$$

$\implies a + b + n \in I$, so $d(u) | n + a + b$. But, by the previous line, $d(u) | n$. Since $n \in J$ is arbitrary, we can write $d(u) | \gcd(J) = d(v)$.

Symmetrically, we could conclude that $d(v) | d(u)$, so indeed $d(v) = d(u)$. □

PROP 1.9 Let $I = \{n > 0 : P^n_{u,u} > 0\}$. If $\gcd(I) = 1$, then $\exists a, b \in I$ such that $\gcd(a, b) = 1$.

PROOF.

This is not true for any I (and thus relies not only on number theory). Let $\ell, m \in I$, with $\ell < m$. Let $k = m - \ell$. If $k = 1$, then $\gcd(\ell, m) = 1$. Otherwise, since $\gcd(I) = 1$, there is an $n \in I$ with $k \nmid n$. We then write $n = qk + r$, with $r \in [k-1]$. Then $m' \in (q+1)m \in I$, since $P_{u,u}^{(q+1)m} \geq (P_{u,u}^m)^{q+1}$. Symmetrically, we can argue $\ell' = (q+1)\ell \in I$.

Similarly, $\ell^* := \ell' + n \in I$, since $P_{u,u}^{\ell'+n} \geq P_{u,u}^{\ell'} P_{u,u}^n$. We have

$$\begin{aligned} m' - \ell^* &= (q+1)m - (q+1)\ell - n = (q+1)(m-\ell) - n \\ &= (q+1)k - n = k - r \in [k-1] \end{aligned}$$

TODO... □

Theorem 1.4 Postage Stamp Lemma

If P is irreducible and aperiodic, then $\forall u, v \in V, \exists N$ such that $P_{u,v}^n > 0 \; \forall n \geq N$.

Before proving this, we note that, for $a, b \geq 1$ with $\gcd(a, b) = 1$, then for any $q \geq ab$, we can write $q = ja + kb$ for integers $j, k \geq 0$.

PROOF.

Fix $u, v \in V$. Since P is aperiodic, there are $a, b \geq 1$ with $P_{u,u}^a, P_{u,u}^b > 0$ and $\gcd(a, b) = 1$, by Prop 1.9. Since P is irreducible, there is some $m > 0$ with $P_{u,v}^m > 0$. Thus, let $N = m + ab$. For any $n \geq N$, let $q = n - m$. We have that $q \geq ab$, so we can find $j, k \geq 0$ with $q = ja + kb$. Then

$$P_{u,v}^n = P_{u,v}^{q+m} = P_{u,v}^{ja+kb+m} \geq P_{u,u}^{ja} P_{u,u}^{kb} P_{u,v}^m \geq (P_{u,u}^a)^j (P_{u,u}^b)^k P_{u,v}^m$$

All are positive, so $P_{u,v}^n > 0$, as desired. □

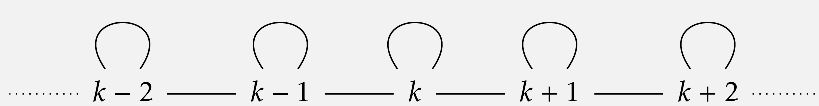
Theorem 1.5 Characterization of Regular Markov Chains

Let $P = (p_{u,v})_{u,v \in V}$ be a stochastic matrix, where $|V| < \infty$. Then

$$P \text{ is regular} \iff P \text{ is irreducible and aperiodic}$$

PROOF.

We first note why finiteness is necessary. Consider:



with all edges having weight 1. This graph is clearly aperiodic and irreducible, but not regular.

($\implies$) If P is regular, then $\exists n > 0$ s.t. $P_{u,v}^n > 0$ for all $u, v \in V$. Then, for all $u, v \in V$, we have $u \to v$ and $v \to u$. Hence, P is irreducible. Now, if P is irreducible, then for all $u \in V$, there is some $v \in V$ such that $P_{v,u} > 0$ (think about this in graph theoretic terms). Then, let $n > 0$ be such that $P^n_{u,u}$ is positive. We have

$$P^{n+1}_{u,u} \geq P^n_{u,v} P_{u,v} > 0$$

So, with $I = \{m > 0 : P^m_{u,u} > 0\}$, $d(u) = \gcd(I) \leq \gcd(n, n+1) = 1$. It follows that $d(u) = 1$, so u is aperiodic (and hence P is aperiodic).

($\Longleftarrow$) By Thm 1.4, for each $u, v \in V$, there exists $N : P^n_{u,v} > 0\ \forall n \geq N$. Let N^* be the maximum value of N determined over all pairs $(u, v) \in V^2$. Then, for $n \geq N^*$ and all $u, v \in V$, $P^n_{u,v} > 0$. It follows that all entries of P^n are positive, and we are done. □

Finding Stationary Distributions

Recall that $x = (x_v : v \in V)$ is a stationary distribution if $xP = x$. Let V be finite. Then, for a stationary distribution x, we have

$$\begin{aligned} x_1 p_{1,1} + \cdots + x_n p_{n,1} &= x_1 \\ x_1 p_{1,2} + \cdots + x_n p_{n,2} &= x_2 \\ &\vdots \\ x_1 p_{1,n} + \cdots + x_n p_{n,n} &= x_n \end{aligned}$$

We have n equations, n unknowns, and a homogeneous system, so there is not a unique solution. If

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0.5 & 0.5 \\ 0.5 & 0 & 0.5 \end{pmatrix}$$

We can compute $x = \langle t, 2t, 2t \rangle$. But, noting that x is a probability distribution, and hence $5t = 1$, this yields $x = \langle 1/5, 2/5, 2/5 \rangle$. We'll consider some special cases.

UNDIRECTED GRAPHS

Let $G = (V, E)$ be undirected. Then we define a THMC by

This is distinct from Example 1.7

$$P_{u,v} = \begin{cases} \frac{1}{\deg(u)} & \{u, v\} \in E \\ 0 & \text{o.w.} \end{cases}$$

Let $x = (\deg(v) : v \in V)$. We have

$$\begin{aligned} (xP)_v &= \sum_{u \in V} \deg(u) P_{u,v} = \sum_{u \in N(v)} \deg(u) \cdot \frac{1}{\deg(u)} \\ &= \deg(v) \end{aligned}$$

Hence, $xP = x$. Recalling that $\sum_{v \in V} \deg(v) = 2|E|$, we conclude that

$$\left(\frac{\deg(v)}{2|E|} : v \in V \right)$$

is a stationary distribution.

UNDIRECTED WEIGHTED GRAPHS

This is not distinct from Example 1.7

Let $G = (V, E)$ be undirected. Then, we define a THMC by

$$P_{u,v} = \begin{cases} \frac{w(\{u,v\})}{\sum_{z \in N(u)} w(\{u,z\})} & v \in N(u) \\ 0 & \text{o.w.} \end{cases}$$

Let $x = (\sum_{e:e\ni v} w(e) : v \in V)$. Then we can compute $xP = x$, and similar to above,

$$x = \left(\frac{\sum_{e:e\ni v} w(e)}{2\sum_{e\in E} w(e)} : v \in V \right)$$

is a stationary distribution.

Transience and Recurrence

Recall $T_S = \inf\{n \geq 0 : X_n \in S\}$, the "hitting time" of S. We let $R_S = \inf\{n > 0 : X_n \in S\}$. Note that if $T_S > 0$, $T_S = R_S$. Otherwise, R_S gives the first "*return time*" to the set S. DEF 1.20

A state $v \in V$ is called *recurrent* if $\mathbb{P}_v(R_{\{v\}} < \infty) = 1$. If all states of v are recurrent, we may P and $X = \text{Markov}(P)$ recurrent. Otherwise, we call v *transient*, and similarly extend the notion to the transition matrix and chain when all state are transient. DEF 1.21 DEF 1.22

For a given state $v \in V$, we call $L_v = |\{n \geq 0 : X_n = v\}|$ the *local time* of v. Note that, if $v = X_j$ and v is recurrent, then $L_v = \infty$. DEF 1.23

Intuitively, if $L_v > k$ when $X_0 = v$, then $L_v > k-1$ when $X_{i_1} = v$, where i_1 is the first time we return to v.

PROP 1.10 Let $X = \text{Markov}(P)$. For any state $v \in V$ and $k > 1$,

$$\mathbb{P}_v(L_v > k) = \mathbb{P}_v(L_v > 1)^k$$

PROOF. Using the law of total probability:

As $R_v = t \iff R_v = t \wedge X_t = v$

$$\begin{aligned} \mathbb{P}_v(L_v > k) = \mathbb{E}\left[\mathbb{P}_v(L_v > k | R_v)\right] &= \sum_{t=1}^{\infty} \mathbb{P}_v(R_v = t)\mathbb{P}_v(L_v > k | R_v = t) \\ &= \sum_{t=1}^{\infty} \mathbb{P}_v(R_v = t)\mathbb{P}_v(L_v > k | R_v = t, X_t = v) \\ &= \sum_{t=1}^{\infty} \mathbb{P}_v(R_v = t)\mathbb{P}_v(L_v > k-1) \\ &= \mathbb{P}_v(L_v > k-1)\sum_{t=1}^{\infty} \mathbb{P}_v(R_v = t) \\ &= \mathbb{P}_v(L_v > k-1)\mathbb{P}_v(R_v < \infty) = \mathbb{P}_v(L_v > k-1)\mathbb{P}_v(L_v > 1) \end{aligned}$$

The result follows by induction. □

PROP 1.11

$$\mathbb{P}_v(L_v = \infty) = \begin{cases} 1 & v \text{ recurrent} \\ 0 & v \text{ transient} \end{cases}$$

PROOF. This follows directly from Prop 1.10 + monotonicity of probability. $\square$

PROP 1.12

$$\sum_{n=0}^{\infty} P_{v,v}^n = \begin{cases} \infty & v \text{ recurrent} \\ \frac{1}{1-\mathbb{P}_v(R_{\{v\}}<\infty)} & v \text{ transient} \end{cases}$$

PROOF. This follows from linearity of expectation, and the fact that, for a non-negative integer variable X, $\mathbb{E}[X] = \sum_{i=0}^{\infty} \mathbb{P}(X > i)$ (We use, and prove, this fact later).

$$\begin{aligned} \sum_{n=0}^{\infty} P_{v,v}^n &= \sum_{n=0}^{\infty} \mathbb{P}_v(X_n = v) = \sum_{i=0}^{\infty} \mathbb{E}[\mathbb{1}_{X_i=v}] \\ &= \mathbb{E}\left[\sum_{i=0}^{\infty} \mathbb{1}_{X_i=v}\right] = \mathbb{E}_v[L_v] = \sum_{i=0}^{\infty} \mathbb{P}(L_v > i) \\ &= \sum_{i=0}^{\infty} \mathbb{P}_v(L_v > 1)^i = \begin{cases} \infty & X \text{ recurrent} \\ \frac{1}{1-\mathbb{P}_v(L_v>1)} & X \text{ transient} \end{cases} \end{aligned}$$

Note that $\mathbb{P}_v(L_v > 1) \equiv \mathbb{P}_v(R_v < \infty)$ $\square$

PROP 1.13 If $u \leftrightarrow v$, then u is transient $\iff$ v is transient.

PROOF. Fix $a, b \geq 0$ with $P_{u,v}^a, P_{v,u}^b > 0$. Then

$$\begin{aligned} \sum_{n=0}^{\infty} P_{v,v}^n \geq \sum_{n=0}^{\infty} P_{v,v}^{a+b+n} &= \sum_{n=0}^{\infty} P_{v,u}^b P_{u,u}^n P_{u,v}^a \\ &= P_{v,u}^b P_{u,v}^a \sum_{n=0}^{\infty} P_{u,u}^n \end{aligned}$$

Thus, if v is transient, then $\sum_{n=0}^{\infty} P_{v,v}^n < \infty$, so it must be that $\sum_{n=0}^{\infty} P_{u,u}^n < \infty$, i.e. u is transient. The argument is identical in reverse. $\square$

Eg. 1.9 If $u \leftrightarrow v$ and u is recurrent, then $\mathbb{P}_u(T_{\{v\}} < \infty) = 1$.

Eg. 1.10 The following chain is completely transient:

$$\begin{tikzcd} \cdots \arrow[r, bend left] & -2 \arrow[l, bend left] \arrow[r, bend left, "2/3"] & -1 \arrow[l, bend left, "1/3"] \arrow[r, bend left, "2/3"] & 0 \arrow[l, bend left, "1/3"] \arrow[r, bend left, "2/3"] & 1 \arrow[l, bend left, "1/3"] \arrow[r, bend left, "2/3"] & 2 \arrow[l, bend left, "1/3"] \arrow[r, bend left] & \cdots \arrow[l, bend left] \end{tikzcd}$$

In fact, we could replace $2/3$ by p, and $1/3$ by $1-p$. In this case, the chain is irreducible. To see that it is transient, we have

$$P_{0,0}^{2n} = \binom{2n}{n} p^n (1-p)^n$$

Then

$$\sum_{n=0}^{\infty} P_{0,0}^{2n} < \sum_{n=0}^{\infty} 2^{2n} p^n (1-p)^n = \sum_{n=0}^{\infty} (4p(1-p))^n < \infty \quad \text{if } p \neq \frac{1}{2}$$

By Prop 1.12. Notice that $\sum_{n=0}^{\infty} P_{0,0}^{n} = \sum_{n=0}^{\infty} P_{0,0}^{2n}$, since it is only possible to return on even-length cycles.

We conclude that the chain is transient when $p \neq \frac{1}{2}$.

FACT

Stirling's Formula provides

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

in that

$$\lim_{m\to\infty} \frac{m!}{\left(\frac{m}{e}\right)^m \sqrt{2\pi m}} = 1$$

These facts, though out of the scope of this course, can be derived from a careful treatment of Reimann sums

This fact implies

$$e\left(\frac{n}{e}\right)^n \leq n! \leq \frac{e(n+1)}{4}\left(\frac{n+1}{e}\right)^n$$

Eg. 1.11 We return to the previous example, letting $p = \frac{1}{2}$. Then

$$\sum_{n=0}^{\infty} P_{0,0}^{2n} = \sum_{n=0}^{\infty} \frac{(2n)!}{(n!)^2}[p(1-p)]^n \sim \sum_{n=0}^{\infty} \frac{\left(\frac{2n}{e}\right)^{2n} \sqrt{2\pi(2n)}}{\left[\left(\frac{n}{e}\right)^n \sqrt{2\pi n}\right]^2}[p(1-p)]^n$$

$$= \sum_{n=0}^{\infty} \frac{[4p(1-p)]^n}{\sqrt{\pi n}} = \sum_{n=0}^{\infty} \frac{1}{\sqrt{\pi n}} = \infty$$

We conclude that the chain is recurrent when $p = \frac{1}{2}$.

PROP 1.14

If V is finite, then there is at least one recurrent state.

PROOF.

Fix an initial distribution $\alpha = (\alpha_v : v \in V)$. Then $\mathbb{P}_\alpha\left(\sum_{v\in V} L_v = \infty\right) = 1$. We conclude that there is at least one state $v \in V$ with $\mathbb{P}_\alpha(L_v = \infty) > 0$. But also:

$$\begin{aligned}
\mathbb{P}_\alpha(L_v = \infty) = \sum_{n=0}^{\infty} \mathbb{P}_\alpha(L_v = \infty, T_v = n) &= \sum_{n=0}^{\infty} \mathbb{P}_\alpha(L_v = \infty | T_v = n)\mathbb{P}_\alpha(T_v = n) \\
&= \sum_{n=0}^{\infty} \mathbb{P}_\alpha(L_v = \infty | T_v = n, X_n = v)\mathbb{P}_\alpha(T_v = n) \\
&= \sum_{n=0}^{\infty} \mathbb{P}_\alpha(L_v = \infty | X_n = v)\mathbb{P}_\alpha(T_v = n) \\
&= \sum_{n=0}^{\infty} \mathbb{P}_v(L_v = \infty)\mathbb{P}_\alpha(T_v = n)
\end{aligned}$$

So $\mathbb{P}_v(L_v = \infty) > 0 \implies \mathbb{P}_v(L_v = \infty) = 1$, by Prop 1.11. $\square$

PROP 1.15 Finite, irreducible chains are recurrent.

PROOF. Since the chain is finite it has at least one recurrent state, by Prop 1.14. Then all states must be recurrent, since the chain is irreducible, by Prop 1.13. $\square$

Canonical Decompositions

Fix a transition matrix P and list the communication classes of V as

$$D_1, D_2, \ldots \quad \text{(transient)} \qquad C_1, C_2, \ldots \quad \text{(recurrent)}$$

Note that we can split the chain up in this way by Prop 1.13. Set $D = \cup_{i \geq 0} D_i$. Then the

DEF 1.24 *canonical decomposition* of the chain is defined to be

$$D \sqcup C_1 \sqcup C_2 \sqcup \cdots$$

DEF 1.25 We say that a communication class C is *closed* if, for any $u \in C$, $v \notin C$, $p_{u,v} = 0$. Intuitively, if $X_0 \in C$, or we enter C at some later time, we will never leave C.

PROP 1.16 If C is a recurrent communication class, then C is closed.

PROOF. Fix $u \in C, v \notin C$. Suppose $v \mapsto u$. If $p_{u,v} > 0$, then $u \mapsto v$, so $v \in C$ ↯. Suppose $v \not\mapsto u$. Then $\mathbb{P}_u(R_u = \infty) \geq \mathbb{P}_u(X_1 = v) = p_{u,v}$. But $\mathbb{P}_u(R_u = \infty) = 0$, since u is recurrent. It follows that $p_{u,v} = 0$. $\square$

The converse of Prop 1.16 is not true in generality, but it is in the finite case:

PROP 1.17 Finite, closed communication classes are recurrent.

PROOF. From any starting state in C, we must visit some state $u \in C$ infinitely often, as $|C| < \infty$ and $X_t \in C \ \forall t$. But recurrence is a class property by Prop 1.13. Hence, all of C is recurrent. $\square$

When our communication classes are closed, we have

$$\begin{tikzcd}
 & & D \arrow[dll, Rightarrow] \arrow[d, Rightarrow] \arrow[drr, Rightarrow] & & & \\
C_1 & & C_2 & & C_3 & \cdots
\end{tikzcd}$$

Proof of Fundamental Theorem of Markov Chains

Recall Thm 1.3:

Every finite, regular stochastic matrix P has a limiting distribution π.

We will prove this in two steps. First, we will find some stationary distribution. Then, we will prove that this is a limiting distribution.

Theorem 1.6 Existence Theorem

Let P be irreducible and recurrent. Let $(X_n : n \geq 0) = \text{Markov}(P)$. Fix $u \in V$, which we call a reference vertex, and, for any $v \in V$, define

$$\gamma_v = \mathbb{E}_u[|\{0 \leq n < R_u : X_n = v\}|]$$

Let $\gamma = (\gamma_v : v \in V)$. Then $\gamma P = \gamma$, and $0 < \gamma_v < \infty \ \forall v \in V$.

By $\mathbb{E}_u$, we mean the expectation, under the assumption that $X_0 = u$

PROOF.

Observe that $\gamma_u = 1$. Write

$$\gamma_v = \mathbb{E}_u\left[\sum_{n=0}^{R_u-1} \mathbb{1}_{X_n=v}\right] = \mathbb{E}_u\left[\sum_{n=1}^{R_u} \mathbb{1}_{X_n=v}\right] = \mathbb{E}_u\left[\sum_{n=1}^{\infty} \mathbb{1}_{X_n=v}\mathbb{1}_{n\leq R_u}\right]$$

We utilize a second indicator variable in order to use linearity of expectation (otherwise, our sum would index over a random variable, which is not valid). Then

$$\gamma_v = \sum_{n=1}^{\infty} \mathbb{E}_u[\mathbb{1}_{X_n=v}\mathbb{1}_{n\leq R_u}] = \sum_{n=1}^{\infty} \mathbb{P}_u(X_n = v, n \leq R_u)$$

Now, using the law of total probability,

$$\begin{aligned}
\mathbb{P}_u(X_n = v, n \leq R_u) &= \sum_{w\in V} \mathbb{P}_u(X_{n-1} = w, X_n = v, n \leq R_u) \\
&= \sum_{w\in W} \mathbb{P}_u(X_n = v | X_{n-1} = w, n \leq R_u)\mathbb{P}_u(X_{n-1} = w, n \leq R_u) \\
&= \sum_{w\in V} \mathbb{P}_u(X_{n-1} = w, n \leq R_u)P_{w,v}
\end{aligned}$$

So

$$\begin{aligned}\gamma_v &= \sum_{n=1}^{\infty}\sum_{w\in W} P_{w,v}\mathbb{P}_u(X_{n-1}=w, n\le R_u) = \sum_{w\in V} P_{w,v}\left(\sum_{n=1}^{\infty}\mathbb{P}_u(X_{n-1}=w, n-1<R_u)\right)\\ &= \sum_{w\in V} P_{w,v}\sum_{n=0}^{\infty}\mathbb{P}_u(X_n=w, n<R_u) = \sum_{w\in V} P_{w,v}\sum_{n=0}^{\infty}\mathbb{E}_u[\mathbb{1}_{X_n=w}\mathbb{1}_{n<R_u}]\\ &= \sum_{w\in V} P_{w,v}\mathbb{E}_u\left[\sum_{n=0}^{R_u-1}\mathbb{1}_{X_n=w}\right] = \sum_{w\in V} P_{w,v}\gamma_w = (\gamma P)_v\end{aligned}$$

$\implies \gamma P = \gamma$. Furthermore, $\gamma_v > 0$, since $u \mapsto v$. Letting $n : P^n_{v,u} > 0$, we have $\gamma_u = (\gamma P^n)_u \ge \gamma_v P^n_{v,u}$. Noting that $\gamma_u = 1$, this shows $\gamma_v < \infty$. □

PROP 1.18 As a corollary, under the same conditions, if $\mathbb{E}_u[R_u]$ is finite, $\pi = (\pi_v : v \in V)$, where

$$\pi_v = \frac{\gamma_v}{\sum_{w\in V}\gamma_w}$$

is a stationary distribution.

PROOF. Observe

$$\sum_{w\in V}\gamma_W = \sum_{w\in V}\mathbb{E}_i\left[\sum_{n=0}^{R_n-1}\mathbb{1}_{X_n=w}\right] = \mathbb{E}_u\left[\sum_{n=0}^{R_n^{-1}}\sum_{w\in V}\mathbb{1}_{X_n=w}\right] = \mathbb{E}_u[R_u]$$

So we require that $\mathbb{E}_u[R_u] < \infty$, then a stationary distribution exists. □

An an exercise, if u is recurrent and $u \mapsto v$, then $\mathbb{P}_u(T_v < \infty) = 1\ \forall v \in V$.

PROP 1.19 π, as defined above, is a limiting distribution.

PROOF. Consider two independent copies $X_n, Y_n = \text{Markov}(P)$. Then (X_n, Y_n) is a Markov chain with transition matrix $Q = (q_{(v,w),(x,y)})_{(v,w),(x,y)\in V\times V}$. In particular,

$$q_{(v,w),(x,y)} = p_{v,x}p_{w,y}$$

Fix some state $u \in V$. Let α be the initial distribution that has $X_0 = u$ and $Y_0 \sim \pi$.

$$\alpha_{(x,y)} = \begin{cases}\pi(y) & x = u\\ 0 & \text{o.w.}\end{cases}$$

Remark that, as π is stationary, $Y_n \sim \pi\ \forall n \ge 0$. For any $u \in V$, we'd like that $P^n_{u,v} \to \pi(v)$.

Let $M = \inf\{n \ge 0 : X_n = Y_n\}$ be the first meeting time of the X_n and Y_n chains. If P is finite and regular, then Q is finite and regular. Then

$$\mathbb{P}_\alpha(M < \infty) \geq \mathbb{P}_\alpha(T_{(v,v)} < \infty) = \sum_{(x,y)\in V\times V} \alpha(x,y)\mathbb{P}_{(x,y)}(T_{v,v} < \infty) = 1$$

It follows that $\lim_{n\to\infty} \mathbb{P}_\alpha(M > n) = 0$. We claim that $\mathbb{P}_\alpha(X_n = v, M \leq n) = \mathbb{P}_\alpha(Y_n = v, M \leq n)\ \forall n \geq 0$. Assuming this, then

$$\begin{aligned}
\mathbb{P}_u(X_n = v) = \mathbb{P}_\alpha(X_n = v) &= \mathbb{P}_\alpha(X_n = v, M \leq n) + \mathbb{P}_\alpha(X_n = v, M > n)\\
\mathbb{P}_\pi(Y_n = v) = \mathbb{P}_\alpha(Y_n = v) &= \mathbb{P}_\alpha(Y_n = v, M \leq n) + \mathbb{P}_\alpha(Y_n = v, M > n)\\
\implies \mathbb{P}_u(X_n = v) - \pi(v) &= \mathbb{P}_\alpha(X_n = v, M > n) - \mathbb{P}_\alpha(Y_n = v, M > n)\\
\implies |\mathbb{P}_u(X_n = v) - \pi(v)| &= |\mathbb{P}_\alpha(X_n = v, M > n) - \mathbb{P}_\alpha(Y_n = v, M > n)| \leq \mathbb{P}_\alpha(M > n)
\end{aligned}$$

But $P_\alpha(M > n) \to 0$, so it must be that $\mathbb{P}_u(X_n = v) \to \pi(v)$.

We still must show the claim, however.

$$\begin{aligned}
\mathbb{P}_\alpha(X_n = v, M \leq n) &= \sum_{k=0}^{n}\sum_{w\in V} \mathbb{P}_\alpha(X_n = v, M = k, X_k = w)\\
&= \sum_{k=0}^{n}\sum_{w\in V} \mathbb{P}_\alpha(M = k, X_k = w)\mathbb{P}_\alpha(X_n = v | X_k = w, M = k)\\
&\overset{\text{MP}}{=} \sum_{k=0}^{n}\sum_{w\in V} \mathbb{P}_\alpha(M = k, X_k = w) P_{w,v}^{n-k}\\
&\textit{and now we reverse!}\\
&= \sum_{k=0}^{n}\sum_{w\in V} \mathbb{P}_\alpha(M = k, Y_k = w) P_{w,v}^{n-k}\\
&\overset{\text{MP}}{=} \sum_{k=0}^{n}\sum_{w\in V} \mathbb{P}_\alpha(M = k, Y_k = w)\mathbb{P}_\alpha(X_n = v | Y_k = w, M = k)\\
&= \mathbb{P}_\alpha(Y_n = v, M \leq n) \qquad \square
\end{aligned}$$

HITTING TIMES AND ABSORBING STATES

We call a state $v \in V$ *absorbing* if $\mathbb{P}_v(X_1 = v) = 1$. DEF 1.26

Eg. 1.12 Consider a game where we continually bet \$1 with winning probability p, until we earn \$$k$ or run out of money. Modeling the state space as our balance, we have absorbing states at \$0 and \$$k$.

Observe that each absorbing state forms a recurrent communication class of size 1. As a consequence, we have a variant of the canonical form, where $A \subseteq V$ are absorbing states, and we write $V = A \sqcup (V \setminus A)$. For a stochastic matrix P and block matrices Q, R, we have:

$$P = \begin{array}{c} \begin{array}{cc} V\setminus A & A \end{array} \\ \left[\begin{array}{c|c} Q & R \\ \hline 0 & I_{|A|} \end{array}\right] \end{array} \implies P^2 = \begin{array}{c} \begin{array}{cc} V\setminus A & A \end{array} \\ \left[\begin{array}{c|c} Q^2 & QR + RI \\ \hline 0 & I_{|A|} \end{array}\right] \end{array}$$

$$\implies P^n = \begin{array}{c} \begin{array}{cc} V\setminus A & A \end{array} \\ \left[\begin{array}{c|c} Q^n & (I_{|V\setminus A|} + Q + \cdots + Q^{n-1})R \\ \hline 0 & I_{|A|} \end{array}\right] \end{array}$$

Schematically, then

$$\begin{tikzcd} V \setminus A \arrow[loop above, "Q"] \arrow[r, "R"] & A \arrow[loop above, "I_{|A|}"] \end{tikzcd}$$

PROP 1.20 Let V be finite. Then all states are either absorbing or transient.

PROOF. Fix an initial distribution $\alpha = (\alpha_v : v \in V)$ and an absorbing state $c \in V$. Then

$$\mathbb{E}_\alpha(T_c) = \sum_{v\in V} \mathbb{P}_\alpha(X_0 = v)\mathbb{E}_\alpha(T_c|X_0 = v) = \sum_{v\in V} \alpha_v \mathbb{E}_v(T_c) =$$

Recall from probability that $\mathbb{E}X = \sum_{n\geq 1} \mathbb{P}(X \geq n)$ whenever X is supported on the non-negative integers. A pictorial proof is as follows

$$\begin{aligned} \mathbb{E}[X] = \mathbb{P}(X = 1)& \\ &+ \mathbb{P}(X = 2) + \mathbb{P}(X = 2) \\ &+ \mathbb{P}(X = 3) + \mathbb{P}(X = 3) + \mathbb{P}(X = 3) \\ &\vdots \\ &+ \mathbb{P}(X = n) + \mathbb{P}(X = n) + \mathbb{P}(X = n) + \cdots + \mathbb{P}(X = n) \\ &= \mathbb{P}(X \geq 1) + \mathbb{P}(X \geq 2) + \cdots + \mathbb{P}(X \geq n) \end{aligned}$$

Assume v is not absorbing. Then

$$\begin{aligned} \mathbb{E}_v(T_c) = \sum_{n\geq 1} \mathbb{P}_v(T_c \geq n) &= \sum_{n\geq 1} \sum_{w\neq c\in V} \mathbb{P}_v(X_{n-1} = w) = \sum_{n\geq 1} \sum_{w\neq c\in V} P^{n-1}_{v,w} \\ &= \sum_{n\geq 1} \sum_{w\in V\setminus A} Q^{n-1}_{w,v} + \sum_{n\geq 1} \sum_{w\neq c\in A} P^{n-1}_{v,w} \end{aligned}$$

where Q is as in the absorbing canonical form above. If there is an absorbing state $b \neq c$ such that $v \mapsto b$, then $\mathbb{E}_v(T_c) = \infty$. Otherwise, the second sum above is 0, and we get

$$\mathbb{E}_v(T_c) = \sum_{w\in V\setminus A} \sum_{n\geq 1} Q^{n-1}_{v,w} = \sum_{w\in V\setminus A} \sum_{n\geq 0} Q^{n}_{v,w}$$

Here, $I + Q + Q^2 + \cdots = (I - Q)^{-1}$. □

Why is this invertible? Genearlly, if $A^n \to 0$ as $n \to 0$, then $I - A$ is invertible.

Theorem 1.7 Hitting State Properties

Let P be a finite transition matrix, such that all states are either transient or absorbing. Let A be the absorbing states, with $|A| = m$. Then

$$\lim_{n\to\infty} P^n = \begin{bmatrix} 0 & (I-Q)^{-1}R \\ 0 & I_m \end{bmatrix}$$

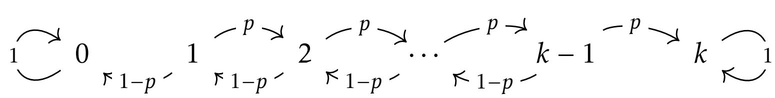
Furthermore, for any transient states u, v,

$$\mathbb{E}_u(L_v) = (I-Q)^{-1}_{u,v} \quad \text{and} \quad \mathbb{E}_u[T_A] = \sum_{v\in V}(I-Q)^{-1}_{u,v}$$

Finally, for any transient state u and absorbing state c,

$$\mathbb{P}_u(X_{T_A} = c) = ((I-Q)^{-1}R)_{u,c}$$

Eg. 1.13 Consider Gambler's Ruin, with reward limit k:



Then $A = \{0, k\}$. Let $p_i = \mathbb{P}_i(X_{T_A} = k)$. Computing manually, this is $p_i = pp_{i+1} + (1-p)p_{i-1}$, with boundary conditions $p_0 = 0$ and $p_k = 1$. We conclude that

$$p_i = \begin{cases} \dfrac{1-\left(\frac{1-p}{p}\right)^i}{1-\left(\frac{1-p}{p}\right)^k} & p \neq \frac{1}{2} \\ \frac{i}{k} & p = \frac{1}{2} \end{cases}$$

Similarly, we manually compute $\mathbb{E}_i(T_A) = 1 + p\mathbb{E}_{i+1}(T_A) + (1-p)\mathbb{E}_{i-1}(T_A)$, with boundary conditions $\mathbb{E}_0(T_A) = \mathbb{E}_k(T_A) = 0$. This gives

$$\mathbb{E}_i(T_A) = \begin{cases} \frac{k(1-r^i)-i(1-r^k)}{p(1-r)(1-r^k)} & p \neq \frac{1}{2} \\ i(k-i) & p = \frac{1}{2} \end{cases} \quad \text{where } r = \frac{1-p}{p}$$

Reversibility

Some motivation: on an atomic scale, we cannot pin-point an "arrow of time," meaning, it is impossible to distinguish between a video on the atomic scale and that video in reverse. Entropy is hard to gauge! For instance, consider the following chains:

By symmetry, it must be that $(\pi_1, \pi_2, \pi_3) = (1/3, 1/3, 1/3)$ is the stationary and limiting distribution for each chain. Can we tell if a "video" of the left chain is played in reverse? By symmetry, no. What about for the right video? We should observe "net counter-clockwise" behavior, as the clockwise transition from a given state only has probability $1/3$. Thus, we could distinguish between a video and its reverse.

Sometimes, it may only be possible to observe the arrow of time in a chain while close to initial conditions. The Ehrenfest Urn is a box with a partition down the middle, with particles initially on the left side of the box.

At each time step, we choose a random point in the box, and move it to its opposing side. When the urn reaches equilibrium, it will be impossible to distinguish an arrow of time. However, at the start, we expect left particles to move to the right.

These notions motivate the following:

DEF 1.27 Let P be a stochastic matrix with stationary distribution π. Then the *time reversal* of P is the matrix $\hat{P}$ with

$$\pi_u \hat{P}_{u,v} = \pi_v P_{v,u}$$

In other words, after one time step, the probability of reaching state v from u in the time reversal is equal to the probability of reaching state u from v in the original chain.

Theorem 1.8 Norris' Theorem

Let P be irreducible with stationary distribution π. Let $X_0 \sim \pi$ and consider the chain $X_n : n \in [N] = \text{Markov}(P)$. Then $(Y_0, ..., Y_N) := (X_N, ..., X_0)$ is $\text{Markov}(\hat{P})$.

PROOF. Not provided in class. See textbook. □

Eg. 1.14 Consider the Ehrenfest Urn example. Let X_n denote the number of particles on the left container at time n. Suppose that there are k total particles, such that $V = [k]$. Then

$$\mathbb{P}(X_{n+1} = i + 1 | X_n = i) = \frac{k-i}{k} \qquad \mathbb{P}(X_{n+1} = i - 1 | X_n = i) = \frac{i-1}{k}$$

What might a stationary distribution be? We could guess $\text{Bin}(k, \frac{1}{2})$, so that $\pi(i) = \frac{1}{2^k}\binom{k}{i}$ (why does this make sense?). Checking,

$$\begin{aligned}
\mathbb{P}_\pi(X_1 = i) &= \pi(i-1)\mathbb{P}_{i-1}(X_1 = i) + \pi(i+1)\mathbb{P}_{i+1}\mathbb{P}_{i+1}(X_1 = i) \\
&= \frac{1}{2^k}\left[\binom{k}{i-1}\frac{k-(i-1)}{k} + \binom{k}{i+1}\frac{i+1}{k}\right] \\
&= \frac{1}{2^k}\left[\frac{(k-1)!}{i(k-i)!}\cdot i + \frac{(k-1)!}{i!(k-1)!}\cdot(k-1)\right] = \frac{1}{2^k}\binom{k}{i} = \pi_i
\end{aligned}$$

We conclude that $X_0 \sim \pi \implies X_1 \sim \pi$, so $\pi P = \pi$.

We can now compute the time reversal matrix, based on Def 1.27:

$$\hat{P}_{i,i+1} = \frac{\pi_{i+1}}{\pi_i} P_{i+1,i} = \frac{\binom{k}{i+1}}{\binom{k}{i}} \frac{i+1}{k} = \frac{k-i}{k} = P_{i,i+1}$$

One shows a similar result for $P_{i,i-1}$. $P = \hat{P}$. In conclusion, $P = \hat{P}$.

A chain $X = \text{Markov}(P)$ and its stochastic matrix P are called *reversible* if $\hat{P} = P$. DEF 1.28

Theorem 1.9 Reversible Chains are Random Weighted Walks

Let P be a stochastic matrix, with a stationary distribution π. Then P is reversible if and only if P can be represented as a random walk on an undirected weighted graph, i.e. there exists a graph $G = (V, E)$ with weights $w : E \to [0, \infty)$ such that

$$P_{u,v} = \frac{w(\{u, v\})}{\sum_{w \in N(u)} w(\{u, w\})} \ \forall u, v \in V$$

PROOF.

($\implies$) Let P be reversible with stationary distribution π. Let $G = (V, E)$ have edge weights $w(\{x, y\}) = \pi_x P_{x,y} = \pi_y P_{y,x}$, by reversibility. Then, for $u, v \in V$,

$$\frac{w(\{u, v\})}{\sum_{w \in N(u)} w(\{u, w\})} = \frac{\cancel{\pi_u} P_{u,v}}{\sum_{w \in N(u)} \cancel{\pi_u} P_{u,w}} = P_{u,v}$$

($\impliedby$) Fix $G = (V, E)$ with $w : E \to [0, \infty)$. Suppose P is given by a random walk on G:

$$P_{u,v} = \frac{w(\{u, v\})}{\sum_{w \in N(u)} w(\{u, w\})}$$

Then, for $x \in V$, a stationary distribution is provided by $\pi(x) = \frac{\sum_{e \in \delta(x)} w(e)}{2 \sum_{e \in E} w(e)}$. Notate $D := 2 \sum_{e \in E} w(e)$. For each $u, v \in V$,

$$\begin{aligned} \pi_u P_{u,v} &= \frac{\sum_{e \in \delta(u)} w(e)}{D} \cdot \frac{w(\{u, v\})}{\sum_{w \in N(u)} w(\{u, w\})} = \frac{w(\{u, v\})}{D} \\ &= \frac{\sum_{e \in \delta(v)} w(e)}{D} \cdot \frac{w(\{v, u\})}{\sum_{e \in \delta(v)} w(e)} = \pi_v P_{v,u} \end{aligned}$$

Hence, P is reversible. $\square$

If π is satisfies $\pi_u P_{u,v} = \pi_v P_{v,u} \ \forall u, v \in V$, then π is stationary for P. PROP 1.21

PROOF.

Check that

$$\mathbb{P}_\pi(X_1 = v) = \sum_{u \in V} \pi_u P_{u,v} = \sum_{u \in V} \pi_v P_{v,u} = \pi_v \implies \pi P = \pi \qquad \square$$

Eg. 1.15 This example is unrelated to our recent work. Consider a cipher of the alphabet, i.e. a bijection $f : C \to A$, where $A = \{a, b, ..., z\}$, and C are a set of cipher symbols. From a collection of ciphered text, how can we determine f?

One naive solution is to model the frequency of symbols in C, and "align" it with the frequency of the alphabet in English text. However, the mapping from letters to their frequency is not injective, so we cannot identify letters uniquely from their frequency.

Alternatively, we can let $M : A \times A \to [0, 1]$ describe the frequency of transitions, i.e. let

$$M(x, y) = \text{ proportion of "}x\text{" that are followed by a "}y\text{" in English text}$$

For instance, $M(s, t)$ may be large, but $M(x, j) = 0$. Then for a given text $s = (s_1, ..., s_L)$ of cipher characters, and a hypothesized mapping f,

$$\text{Plausibility}(f) = \prod_{i=1}^{L-1} M(f(s_i), f(s_{i+1}))$$

This is exactly the probability that a Markov chain with transition matrix M produces the sequence $(f(s_1), ..., f(s_L))$. We can maximize the plausibility score algorithmically:

```
function MONTE-CARLO
    f ← any guess f : C → A
    while not satisfied do
        compute Pl(f)
        generate f_* from f by a random transposition
        if Pl(f_*) > Pl(f) then
            f ← f_*
        else
            toss coin
            if Heads then
                Do nothing
            else
                f ← f_*
```

Metropolis-Hastings

Let V be a finite state space. Let $\pi = (\pi_v)_{v \in V}$ be a probability distribution on states. Let T be a transition matrix such that $T_{x,y} > 0 \iff T_{y,x} > 0$. We wish to use T to generate a sample from π.

We can build a new transition matrix P with $\pi P = \pi$, in which case it suffices to show that $\pi_u P_{u,v} = \pi_v P_{v,u}$, by Prop 1.21.

For $x, y \in V$, let

$$A(x, y) := \frac{\pi_y T_{y,x}}{\pi_x T_{x,y}}$$

Intuitively, this asks "how much more likely is it that I encounter $y \mapsto x$ versus $x \mapsto y$ on my first transition" Then, we let

$$P_{x,y} := \begin{cases} T_{x,y} & x \neq y, A(x, y) \geq 1 \\ T_{x,y} A(X, y) & x \neq y, A(x, y) < 1 \\ T_{x,y} + \sum_{z:A(x,z)<1} T_{x,z}(1 - A(x, z)) & x = y \end{cases}$$

To verify that this is indeed a stochastic matrix, we have

$$\begin{aligned} \sum_{y \in V} P_{x,y} &= \left[\sum_{y \neq x: A(x,y) \geq 1} T_{x,y} \right] + \left[\sum_{y \neq x: A(x,y) < 1} T_{x,y} A(x, y) + T_{x,x} + \sum_{z: A(x,z) < 1} T_{x,z}(1 - A(x, z)) \right] \\ &= \sum_{y \in V} T_{x,y} = 1 \end{aligned}$$

Algorithmically, P behaves the following way:

Given $X_n = x$, to generate X_{n+1}...

- Choose "proposal state" Y with $\mathbb{P}(Y = y) = T_{x,y}$.
- Generate $U \in \text{Unif}[0, 1]$.
- If $U \leq A(x, Y)$, set $X_{n+1} = Y$
- Else, set $X_{n+1} = x$

We claim that P is reversible with stationary distribution π.

PROOF.

Fix $x, y \in V$. We show that $\pi_x P_{x,y} = \pi_y P_{y,x}$. If $x = y$, this is clear. Otherwise, assume $A(x, y) \geq 1$. Then

$$\pi_x P_{x,y} = \pi_x T_{x,y} = \frac{\pi_y T_{y,x}}{A(x, y)} = \pi_y T_{y,x} A(y, x) = \pi_y P_{y,x}$$

□

The algorithm, or, equivalently, the definition for $P_{x,y}$, is called the *Metropolis-Hastings algorithm*, or "rejection sampling." DEF 1.29

A similar subject to look into: "mixing times"

II Branching Processes

This topic covers a special case of Markov chains, called branching processes.

Eg. 2.1 Let $X \sim \mathrm{Ber}(\frac{1}{2})$. Consider a root note. If $X = H$, we give that node 2 children. Else, we give it no children. We repeat this process for each child.

This is a model for asexual reproduction (or, alternatively, one's surname passing down patrilineally).

We can generalize this slightly by having the number of children of parent i in generation n be distributed according to $B_{n,i} \sim B$, where $B \in \mathbb{Z}_{\geq 0}$ are IID. We let Z_n denote the number of total children at generation n.

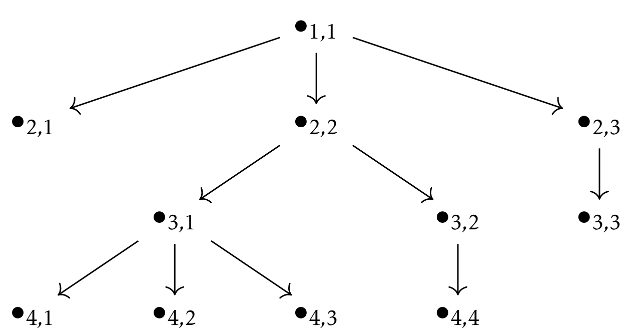
We can represent a realized sampling from B as a matrix of non-negative integers, where $B_{n,i} = (B)_{n,i}$. If generation $n-1$ has Z_{n-1} individuals, then the next generation has

$$B_{n,1} + B_{n,2} + \cdots + B_{n,Z_{n-1}}$$

individuals. For instance, consider

$$B = \begin{bmatrix} 3 & 1 & 0 & 0 & \cdots \\ 0 & 2 & 1 & 4 & \cdots \\ 3 & 1 & 0 & 1 & \cdots \\ 0 & 0 & 0 & 0 & \cdots \\ 2 & 1 & 2 & 3 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \end{bmatrix}$$

Then, if we consider only the first initial child,



Let $B_{n,i} \sim B$ be IID, non-negative, integer random variables. Let Z_0 be some fixed integer.
DEF 2.1 Inductively define $Z_n = \sum_{i=1}^{Z_{n-1}} B_{n,i}$. Then Z_n is called a *branching process*, and we write $Z_n \sim \text{Bienaymé}(B)$.

PROP 2.1 Branching processes are time-homogeneous Markov chain, with transition matrix

$$P_{k,\ell} = \mathbb{P}(B_{1,1} + \ldots + B_{1,k} = \ell) = \mathbb{P}(kB = \ell)$$

PROOF.

Our state space consists of non-negative integers. We verify the Markov property:

$$\begin{aligned}\mathbb{P}(Z_{n-1} = \ell | Z_0 = s_0, ..., Z_{n-1} = s_{n-1}, Z_n = k) &= \mathbb{P}(B_{n+1,1} + ... + B_{n+1,k} = \ell | \cdots) \\ &= \mathbb{P}(B_{n+1,1} + ... + B_{n+1,k} = \ell) \\ &= \mathbb{P}(B_{1,1} + ... + B_{1,l} = \ell) =: \clubsuit\end{aligned}$$

Then, we have

$$\begin{aligned}\mathbb{P}(Z_{n+1} = \ell | Z_n = k) &= \sum_{\mathbf{s}\geq 0} \mathbb{P}(Z_0 = s_0, .., Z_{n-1} = s_{n-1}, Z_{n+1} = \ell | Z_n = k) \\ &= \sum_{\mathbf{s}\geq 0} \mathbb{P}(Z_0 = s_0, ..., Z_{n-1} = s_{n-1} | Z_k) \cdot \\ &\quad \mathbb{P}(Z_{n+1} = \ell | Z_n = k, Z_0 = s_0, ..., Z_{n-1} = s_{n-1}) \\ &= \sum_{\mathbf{s}\geq 0} \mathbb{P}(Z_0 = s_0, ..., Z_{n-1} = s_{n-1} | Z_n = k) \clubsuit = \clubsuit\end{aligned}$$

So, indeed, Z_n is time-homogeneous. By $\clubsuit$, we have the transition matrix

$$P_{k,\ell} = \mathbb{P}(B_{1,1} + ... + B_{1,k} = \ell) \qquad \square$$

We are curious as to the survival or extinction of Z_n. We say that Z_n *survives* if $Z_n > 0 \ \forall n \geq 0$, or in probability, $\mathbb{P}(Z_n > 0 \ \forall n \geq 0) = 1$. We call it *extinct* otherwise. DEF 2.2 DEF 2.3

PROP 2.2 If $Z_n \sim$ Bienaymé(B), then

$$\mathbb{E}[Z_n | Z_0 = k] = k(\mathbb{E}[B])^n$$

PROOF.

We proceed by induction on n. When $n = 0$, then $\mathbb{E}[Z_0 | Z_0 = k] = k = k(\mathbb{E}B)^0$.

Now, we have

$$\begin{aligned}\mathbb{E}[Z_{n+1} | Z_n = k] &= \mathbb{E}\left[B_{n+1,1} + ... + B_{n+1,k} | Z_n = k\right] \\ &= \sum_{i=1}^{k} \mathbb{E}\left[B_{n+1,i} | Z_n = k\right] = k\mathbb{E}[B]\end{aligned}$$

Thus, $\mathbb{E}[Z_{n+1} | Z_n] = Z_n \mathbb{E}[B]$. Then, using the tower law of expectation,

$$\begin{aligned}\mathbb{E}[Z_{n+1}] &= \mathbb{E}_{Z_n}[\mathbb{E}[Z_{n+1} | Z_n]] = \mathbb{E}_{Z_n}[Z_n \mathbb{E}[B]] \\ &= \mathbb{E}[Z_n]\mathbb{E}[B] = k(\mathbb{E}[B])^n \mathbb{E}[B] = k(\mathbb{E}[B])^{n+1}\end{aligned}$$

using induction. $\square$

PROP 2.3 If $\mathbb{E}[B] < 1$, then $\mathbb{P}($ survival $| Z_0 = k) = 0 \ \forall k \geq 0$.

PROOF. Fix $\varepsilon > 0$. Let $N \geq 0$ be large enough such that $k(\mathbb{E}[B])^n < \varepsilon$. Then

$$\begin{aligned}\mathbb{P}(\text{ survival } |Z_0 = k) &\leq \mathbb{P}(Z_N \geq 1|Z_0 = k) \leq \mathbb{E}[Z_N|Z_0 = k] \\ &\leq \sum_{\ell \geq 1} \mathbb{P}(Z_n \geq \ell|Z_0 = k) = \mathbb{E}[Z_N|Z_0 = k] = k(\mathbb{E}[B])^N < \varepsilon \qquad \square\end{aligned}$$

Generating Functions

DEF 2.4 Let X be a random variable. Recall from probability the *moment generating function* G_X, where

$$G_X(s) = \mathbb{E}[s^X]$$

In some textbooks, this is $\mathbb{E}[e^{sX}]$, but these notions are equivelant up to a change of variables

For an integer-valued random variable,

$$G_X(s) = \sum_{k \in \mathbb{Z}} s^k \mathbb{P}(X = k)$$

and, similarly, for a continuous random variable,

$$G_X(s) = \int_{\mathbb{R}} s^x f_X(x) dx$$

We will focus on integer-valued variables for branching processes.

Eg. 2.2 When X is integer-valued,

$$G_X(1) = \mathbb{E}[1^X] = \sum_{k \in \mathbb{Z}} 1^k \mathbb{P}(X = k) = 1$$

Similarly,

$$|G_X(s)| = |\mathbb{E}[s^X]| = \left| \sum_{k \in \mathbb{Z}} s^k \mathbb{P}(X = k) \right| \leq \sum_{j \in \mathbb{Z}} |s|^k \mathbb{P}(X = k) \leq G_X(1)$$

We conclude that $G_X(s)$ is defined for all $|s| \leq 1$. (In fact, it is infinitely differentiable on this interval).

Eg. 2.3 Let $X = \text{Ber}(p)$. Then $\mathbb{E}[s^X] = (1 - p) \cdot 1 + ps = 1 - p + ps$. For kX, $\mathbb{E}[s^{kX}]$ becomes $1 - p + ps^k$

Eg. 2.4 Let $X = \text{Bin}(n, p) = B_1 + \ldots + B_n : B_i \sim \text{Ber}(p)$. Then

$$\mathbb{E}[s^X] = \mathbb{E}[s^{B_1 + \ldots + B_n}] = \prod_{i=1}^{n} \mathbb{E}[s^{B_i}] = [1 - p + ps]^n$$

Considering the special case $\text{Bin}(n, \frac{\lambda}{n})$, where $\lambda > 0$, and we think of n as large,

$$\mathbb{E}[s^X] = \left(1 - \frac{\lambda}{n} + \frac{\lambda s}{n}\right)^n = \left(1 + \frac{\lambda(s-1)}{n}\right)^n \approx e^{\lambda(s-1)}$$

Eg. 2.5 When $X = \text{Poisson}(\lambda)$, where $\mathbb{P}(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$,

$$\mathbb{E}[s^X] = \sum_{k\geq 0} \frac{s^k e^{-\lambda}\lambda^k}{k!} = e^{-\lambda} \sum_{k\geq 0} \frac{(\lambda s)^k}{k!} = e^{-\lambda} e^{\lambda s} = e^{\lambda(s-1)}$$

We conclude that the moments of a Poisson variable can be estimated by a suitable binomial variable.

Eg. 2.6 Suppose $X_n = A_{n,1} + \ldots + A_{n,n}$, where $A_{n,i}$ are independent Bernoulli (not necessarily identically distributed), such that $\mathbb{E}[X_n] = \lambda$. We can ask: under what conditions on the expectation of $\mathbb{E}[A_{n,i}]$ does $G_{X_n}(s) \to e^{\lambda(s-1)}$?

We now connect these ideas to branching processes. Recall that an offspring distribution B is a non-negative integer-valued random variable.

Eg. 2.7 Let $X \sim 2\text{Ber}(p)$. When $p = 1/3$, we have that $G_X(s) = \frac{2}{3} + \frac{1}{3}s^2$. Check these graphs for $p = 1/2$ and $p = 1/3$.

We want to get an idea of how G_B looks on $s \in [0, 1]$, where B is an offspring distribution:

$$\begin{aligned}
G_B(0) &= 0^0\mathbb{P}(B = 0) + \sum_{k\geq 1} 0^k \mathbb{P}(B = k) && = \mathbb{P}(B = 0) \\
G_B(1) & && = 1 \\
G_B(s) &= \sum_{k\geq 0} s^k \mathbb{P}(B = k) && \geq 0 \text{ for } s \geq 0 \\
G_B'(s) &= \sum_{k\geq 0} ks^{k-1}\mathbb{P}(B = k) = \sum_{k\geq 1} ks^{k-1}\mathbb{P}(B = k) && \geq 0 \text{ for } s \geq 0 \\
G_B'(0) &= 1 \cdot 0^0 \mathbb{P}(B = 1) + \sum_{k\geq 2} k0^{k-1}\mathbb{P}(B = k) && = \mathbb{P}(B = 1) \\
G_B'(1) &= \sum_{k\geq 1} k \cdot 1^{k-1}\mathbb{P}(B = k) = \sum_{k\geq 0} k\mathbb{P}(B = k) && = \mathbb{E}[B] \\
G_B''(s) &= \sum_{k\geq 0} k(k-1)s^{k-2}\mathbb{P}(B = k) = \sum_{k\geq 2} k(k-1)s^{k-2}\mathbb{P}(B = k) && \geq 0, \text{ and} > 0 \text{ when } \mathbb{P}(B \geq 2) > 0
\end{aligned}$$

Alternatively, $G_B''(s) > 0$ when $\mathbb{P}(B \leq 1) < 1$.

Thus, we can characterize $G_B(s)$ as a positive, non-decreasing, convex function, starting at $\mathbb{P}(B = 0)$ and ending at 1 for $s \in [0, 1]$. In particular, there exists some $0 \leq s < 1$ with $G_B(s) = s$ if and only if $\mathbb{E}[B] > 1$ or $\mathbb{P}(B = 1) = 1$

If $Z_n : n \geq 0$ is Bienaymé(B) with $Z_0 = 1$, then, for $n \geq 1$, we have **PROP 2.4**

$$G_{Z_n}(s) = \underbrace{G_B(G_B(\cdots(G_B(s))))}_{n \text{ times}}$$

PROOF. We'll show this by induction. When $n = 1$, $G_{Z_1} = G_{B_{1,1}}(s) = G_B(s)$.

Now, for $n \geq 1$, we have

$$G_{Z_{n+1}} = \mathbb{E}[s^{Z_{n+1}}] = \sum_{k\geq 0} \mathbb{P}(Z_n = k)\mathbb{E}[s^{Z_{n+1}}|Z_n = k] \text{ by tower property}$$

Then, we have

$$\begin{aligned}\mathbb{E}[s^{Z_{n+1}}|Z_n = k] &= \mathbb{E}\left[s^{B_{n+1,1}+\ldots+B_{n+1,k}}|Z_n = k\right] = \mathbb{E}\left[s^{B_{n+1,1}+\ldots+B_{n+1,k}}\right] \\ &= \prod_{i=1}^{k} \mathbb{E}[s^{B_{n+1,i}}] \text{ by independence} \\ &= G_B(s)^k\end{aligned}$$

Then, continuing from above,

$$\sum_{k\geq 0} \mathbb{P}(Z_n = k)\mathbb{E}[s^{Z_{n+1}}|Z_n = k] = \sum_{k\geq 0} \mathbb{P}(Z_n = k)G_B(s)^k = \mathbb{E}\left[G_B(s)^{Z_n}\right] = G_{Z_n}(G_B(s))$$

and, applying induction, we're done. □

PROP 2.5 As a corollary,

$$\mathbb{P}(Z_n = 0|Z_0 = 1) = G_B(G_B(\cdots(G_B(0))))$$

PROOF. We know $\mathbb{P}_1(Z_n = 0) = G_{Z_n}(0)$. □

PROP 2.6 Write $\tau = \inf(n \geq 0 : Z_n = 0)$, i.e. the extinction time of Z_n. Note that $\tau = \infty$ if and only if $Z_n > 0\ \forall n \geq 0$. Then $\mathbb{P}_1(\tau < \infty) = \lim_{n\to\infty} \mathbb{P}_1(Z_n = 0)$.

PROOF. This is an application of the monotone limit theorem. First we take a detour in probability. For events $(E_n : n \geq 1)$ such that $E_1 \subseteq E_2 \subseteq \cdots$, we have

$$\begin{aligned}\mathbb{P}\left(\cup_{n\geq 1} E_n\right) &= \mathbb{P}\left(E_1 \cup (E_2 \setminus E_1) \cup \cdots \cup (E_n \setminus E_{n-1})\right) \\ &= \mathbb{P}(E_1) + \sum_{n\geq 2} \mathbb{P}(E_n \setminus E_{n-1}) = \mathbb{P}(E_1) + \lim_{N\to\infty} \sum_{n=2}^{N} \mathbb{P}(E_n \setminus E_{n-1}) \\ &= \lim_{N\to\infty} \mathbb{P}(E_N) \qquad \text{by additivity}\end{aligned}$$

Setting $E_n = \{Z_n = 0\}$ for $n \geq 1$ (note that these events are nested), we conclude

$$\mathbb{P}(\cup_{n\geq 1}\{Z_n = 0\}) = \lim_{n\to\infty} \mathbb{P}(Z_n = 0)$$

But $\cup_{n\geq 1}\{Z_n = 0\}$ is the event $\tau < \infty$, and we are done. □

Theorem 2.1 Fundamental Theorem of Branching Processes

Let $(Z_n : n \geq 0)$ be Bienaymé(B), with $Z_0 = 1$. Then

$$\mathbb{P}_1(\text{extinction}) = \mathbb{P}_1(\tau < \infty)$$

is the smallest non-negative solution s of $G_B(s) = s$. In particular, $\mathbb{P}_1(\tau < \infty) < 1$ if and only if either $\mathbb{E}[B] > 1$ or $\mathbb{P}(B = 1) = 1$.

PROOF.

Let $s := \mathbb{P}_1(\tau < \infty)$. We've shown that $G_{Z_n}(0) = \mathbb{P}_1(Z_n = 0) \to s$ as $n \to \infty$, by Prop 2.6. But, since G_B is continuous on $[0, 1]$, we also have that $G_B(G_{Z_n}(0)) \to G(s)$ as $n \to \infty$. But $G_B(G_{Z_n}(0)) = \mathbb{P}_1(Z_{n+1} = 0)$ by Prop 2.5, so it must also converge to s. Thus, $G_B(s) = s$.

It remains to show that s is the smallest solution to $G(s) = s$. In particular, suppose $t \geq 0$ satisfies $G_B(t) = t$. Then $t \geq 0$, so $G_B' \geq 0$, and thus, by monotonicity, $G_B(t) \geq G_B(0)$. But then $t \geq G_B(0)$. Again by monotonicity, $G_B(t) \geq G_B(G_B(0))$, i.e. $t \geq G_B(G_B(0))$. By induction, and Prop 2.5, $t \geq \lim_{n\to\infty} \mathbb{P}_1(Z_n = 0) = \mathbb{P}_1(\tau < \infty) = s$. □

POISSON PROCESSES

Examples and Intuition

Recall Poisson variables: they are useful for modeling outcomes of a large number of independent trials, where each trail has success probability on the order of $O(\frac{1}{N})$, frequently $\frac{c}{N}$. For instance, $\text{Bin}(N, \frac{c}{N}) \approx \text{Poi}(c)$. Poisson processes operate similarly, with some added spacial or temporal structure.

Eg. 2.8 We consider 1 time dimension. Then, Poisson processes might be well-suited to model arrivals at a Metro self-checkout line; or radioactive decay events; or childbirths. An example which is not well suited might be train arrival times (in a given timestep, the probability of a train arriving is dictated, or influenced, by a schedule; the probability of a train arriving in timestep $n + 1$ is reduced by the event of a train arriving at timestep n).

We consider 2 spacial dimensions. The rainfall on a square meter of pavement is well-modeled (with spacial units one square mm, for instance). We could also consider the locations of shooting stars in a meteor shower.

Some key properties of Poisson processes, which we should keep in mind when modeling:

1. An event at one time or location does not affect the probability of another event at another time or location.
2. Events cannot occur simultaneously or at the same location.
3. Events occur at a rate, i.e. for an event E, $\mathbb{P}[E \in \mathcal{S}] \approx c|\mathcal{S}|$, where $\mathcal{S}$ is some measurable time-space region with "volume" $\mathcal{S}$.

Suppose we have a two-dim region R, which we separate into area ε-sized regions. Suppose the rate of an event (which we will call a "point") in an ε-region is c. Take some subregion $R_1 \subseteq R$ with area A_1. Then, the total number of ε-regions which contain an event is approximately

$$\approx \mathrm{Bin}\left(\frac{A_1}{\varepsilon}, c\varepsilon\right) \approx \mathrm{Poi}(A_1 \cdot c)$$

If we let $R_2 \subseteq R$ be disjoint from R_1, with area A_2, then we similarly model the number of points in A_2 by $\mathrm{Poi}(cA_2)$, which, as a variable, is independent from $\mathrm{Poi}(cA_1)$ above.

INDEX OF DEFINITIONS