INTRODUCTION TO STOCHASTIC PROCESSES

CLASS NOTES FOR MATH/STAT 396

Michael D. Perlman
Department of Statistics
University of Washington
Seattle, Washington 98195
michael@stat.washington.edu

TABLE OF CONTENTS

1. Types and Examples of Stochastic Processes

1.1. Glossary of terms; simple symmetric random walk (SSRW)

A stochastic process is a (finite, countable or uncountable) collection of random variables (rvs) that evolve in time or space according to probabilistic laws. The interesting cases involve *dependent* random variables, and interesting questions often concern events involving the long-term behavior of the process, i.e., events determined by infinitely many of the variables.

Discrete-time process. Here the time index n is discrete, usually $\{0, 1, 2, \ldots\}$, so the stochastic process can be written as $\{Y_n \mid n = 0, 1, 2, \ldots\}$. (Sometimes we may consider $\{Y_n \mid n = 0, \pm 1, \pm 2, \ldots\}$.)

Continuous-time process. Here the time index $t \in [0, \infty)$ and the process is $\{Y_t \mid 0 \le t < \infty\}$. (Sometimes $\{Y_t \mid -\infty < t < \infty\}$.)

Discrete or continuous state space. Here the range of each Y_n or Y_t is discrete (i.e., either finite or countable) or continuous, respectively. (It is assumed that the range of Y_n or Y_t is the *same* for each n or t.

Stationary process. The continuous-time process $\{Y_t\}$ is *stationary* $\equiv$ *time-homogeneous* if for every finite set $t_1 < \cdots < t_k$ the joint distribution of $(Y_{t_1+h}, \ldots, Y_{t_k+h})$ is the same for all h. (Similarly for discrete-time $\{Y_n\}$.)

Stationary/independent increments. The process $\{Y_t\}$ (similarly for $\{Y_n\}$) has *stationary increments* if for every finite sequence

$$s_1 < t_1 \le s_2 < t_2 \le \cdots \le s_k < t_k,$$

the joint distribution of the increments $Y_{t_1+h} - Y_{s_1+h}, \ldots, Y_{t_k+h} - Y_{s_k+h}$ is the same for every h. The process $\{Y_t\}$ has *independent increments* if $Y_{t_1} - Y_{s_1}, \ldots, Y_{t_k} - Y_{s_k}$ are mutually independent.

Markov process. The process $\{Y_t\}$ is a *Markov process* (similarly for $\{Y_n\}$) if for every finite set $r_1 < \cdots < r_k < s < t$,

$$\mathcal{L}(Y_t \mid Y_{r_1}, \ldots, Y_{r_k}, Y_s) = \mathcal{L}(Y_t \mid Y_s), \tag{1.1}$$

that is, if the conditional distribution ($\equiv$ law, abbreviated $\mathcal{L}$) of Y_t given the past and present observations $Y_{r_1}, \ldots, Y_{r_k}, Y_s$ depends only on the present observation Y_s. In fact (see Exercise 1.1), (1.1) is equivalent to the stronger condition that for every $r_1 < \cdots < r_k < s < t_1 < \ldots < t_l$,

$$\mathcal{L}(Y_{t_1}, \ldots, Y_{t_l} \mid Y_{r_1}, \ldots, Y_{r_k}, Y_s) = \mathcal{L}(Y_{t_1}, \ldots, Y_{t_l} \mid Y_s). \tag{1.2}$$

Here (1.2) can be written equivalently in terms of conditional independence:

$$Y_{t_1}, \ldots, Y_{t_l} \perp\!\!\!\perp Y_{r_1}, \ldots, Y_{r_k} \mid Y_s, \tag{1.3}$$

where "$\perp\!\!\!\perp$" denotes (conditional) independence. We can express (1.3) as

$$\text{Future} \perp\!\!\!\perp \text{Past} \mid \text{Present}. \tag{1.4}$$

Note that here the Future and Past appear symmetrically, unlike in (1.2).

Exercise 1.1. Suppose that (Y_1, Y_2, Y_3, Y_4) is a Markov process. Working with conditional probability density functions (pdfs), show that

$$\mathcal{L}(Y_3, Y_4 \mid Y_1, Y_2) = \mathcal{L}(Y_3, Y_4 \mid Y_2). \tag{1.5}$$

Note that this implies that (1.2) follows from (1.1). $\square$

Markov Chain. A discrete-time Markov process with a discrete state space is called a *Markov chain.*

Martingale. The process $\{Y_t\}$ is a *martingale* (similarly for $\{Y_n\}$) if each $\mathrm{E}|Y_t| < \infty$ and for every finite set of time indices $r_1 < \cdots < r_k < s < t$,

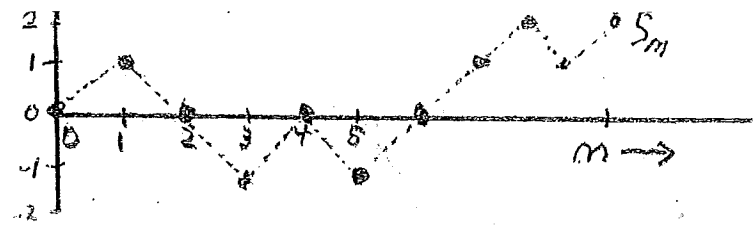
$$\mathrm{E}[Y_t \mid Y_{r_1}, \ldots, Y_{r_k}, Y_s] = Y_s. \tag{1.6}$$

The martingale property is less restrictive than the Markov property in that it only imposes a requirement on the conditional expectation of Y_t, not on its entire conditional distribution, but is more restrictive in that it specifies exactly what this conditional expectation must be.

Example 1.1. *(White noise).* Let $Y_0, Y_1, \ldots$ be independent, identically distributed (iid) random variables. The discrete-time process $\{Y_n \mid n \geq 0\}$ is called a *white noise process.* Trivially it is stationary and Markovian. It is not a martingale and does not have independent increments. It has no interesting limiting properties. □

Basic Example 1.2. *(Simple symmetric random walk (SSRW).*
Let $X_0 \equiv 0$ and $X_i = \pm 1$ with probabilities $(\frac{1}{2}, \frac{1}{2})$, $i = 1, 2, \ldots$. Assume that the X_i's are iid, and let $S_n = X_1 + \cdots + X_n$. Then $\{S_n \mid n = 0, 1, 2, \ldots\}$ is a *random walk.* A typical sample realization looks like

State space: $\{$



Here $\{S_n\}$ is a discrete-time Markov process, a martingale, and has stationary independent increments [verify]. It is not a stationary process. □

Remark 1.1. What is the sample space for the random walk $\{S_n\}$? It is in 1-1 correspondence with the sample space of the sequence of iid ± 1 rvs $X_1, X_2, \ldots$, that is, with the space of all infinite sequences of ± 1's. This space is *not* discrete, rather, it is in 1-1 correspondence with the space of all binary sequences of 0's and 1's, hence in turn is in 1-1 correspondence with the *continuum* $[0, 1]$. [Why?] This shows that the probability of any particular sample realization of the random walk is 0. This can be seen directly by noting that for any sequence $(x_1, x_2, \ldots)$ of ± 1's,

$$\Pr[X_1 = x_1, X_2 = x_2, \ldots] = \lim_{n \to \infty} \Pr[X_1 = x_1, \ldots, X_n = x_n] = \lim_{n \to \infty} \frac{1}{2^n} = 0.$$

Remark 1.2. Since $\mathrm{E}(X_i) = 0$ and $\mathrm{Var}(X_i) = 1$ for $i \geq 1$, $\mathrm{E}(S_n) = 0$ and $\mathrm{Var}(S_n) = n$, so $|S_n| \approx \sqrt{n}$ for each fixed n. However, it is *not* true that for some $c > 0$,

$$\Pr\{|S_n| \leq c\sqrt{n} \ \forall \ n\} = 1.$$

Instead, the *Law of the Iterated Logarithm (LIL)* implies here that

$$\Pr\left\{ \limsup_{n \to \infty} \frac{|S_n|}{\sqrt{2n \log \log n}} = 1 \right\} = 1. \tag{1.7}$$

This is a very deep result which we won't treat here. (Clearly $|S_n| \leq n$ for every n.) The LIL should be compared to another deep result, the *Strong Law of Large Numbers (SLLN)*, which implies here that (since $\mathrm{E}X_i = 0$)

$$\Pr\left\{\lim_{n\to\infty} \bar{X}_n \equiv \lim_{n\to\infty} \frac{S_n}{n} = 0\right\} = 1. \tag{1.8}$$

□

Without proving the LIL or SLLN, we can establish the remarkable fact that *the probabilities in (1.7) and (1.8) must be either 0 or 1.* For this we need to define a tail event. Let $Z_1, Z_2, \ldots$ be any sequence of random variables. An event A defined in terms of these Z_i's is called a *tail event* if its occurrence does not depend on the values of any finite set of the Z_i's. Examples include events such as

$$\left\{\lim_{n\to\infty} Z_n \text{ exists}\right\}, \quad \left\{\lim_{n\to\infty} \bar{Z}_n = c\right\}, \quad \left\{\limsup_{n\to\infty} \frac{S_n}{\sqrt{2n\log\log n}} = c\right\},$$

where $\bar{Z}_n = (Z_1 + \ldots + Z_n)/n$, $S_n = Z_1 + \ldots + Z_n$, and $-\infty \leq c \leq \infty$.

Theorem 1.1 *(Kolmogorov's 0-1 Law). Let $Z_1, Z_2, \ldots$ be a sequence of independent rvs. Then $\Pr(A) = 0$ or 1 for any tail event A.*

Proof. For each n, A depends only on $Z_{n+1}, Z_{n+2}, \ldots$, hence A is independent of $Z_1, \ldots, Z_n$. Therefore A is independent of the entire sequence $Z_1, Z_2, \ldots$, hence A is independent of itself! Therefore

$$\Pr(A) = \Pr(A \cap A) = \Pr(A)\Pr(A) = (\Pr(A))^2,$$

hence $\Pr(A) = 0$ or 1! □

This explains why, in the context of the SLLN, $\lim \bar{X}_n$ must be a degenerate random variable, i.e., a constant with probability one.

Question 1.1: Consider SSRW. For a fixed integer $m \geq 1$, will the process $\{S_n\}$ ever/always hit the boundary $\{-m, m\}$?

Proposition 1.1: *For SSRW,* $\Pr[\sup_{n\geq 1} |S_n| \geq m] = 1$.

Proof: The process will hit the boundary $\pm m$ if (not only if) $\exists$ a run of $2m$ consecutive $+1$'s. Note that

$$\Pr[X_1 = 1, \ldots, X_{2m} = 1] = \left(\tfrac{1}{2}\right)^{2m} \equiv \epsilon > 0,$$

so as $k \to \infty$,

$$\begin{aligned}\Pr[&(X_1 = 1, \ldots, X_{2m} = 1) \cup \cdots \cup (X_{2(k-1)m+1} = 1, \ldots, X_{2km} = 1)] \\ &= 1 - (1-\epsilon)^k \to 1.\end{aligned}$$

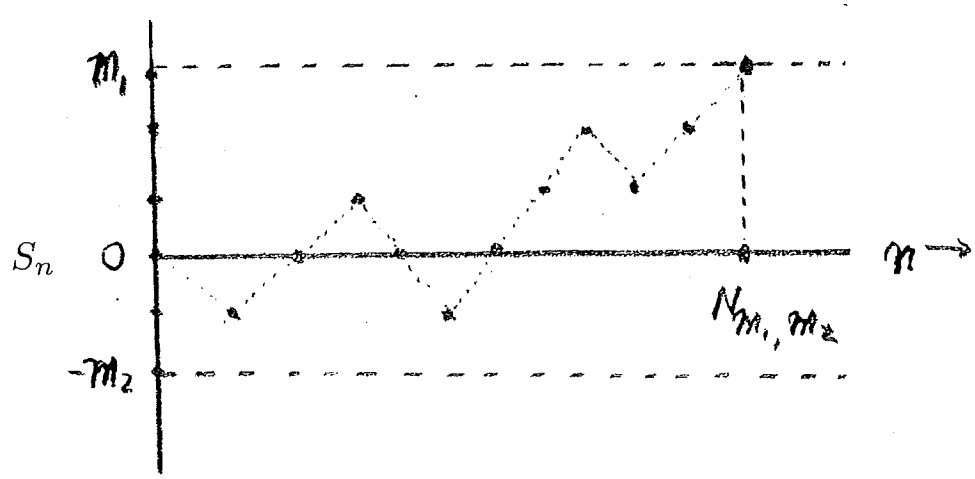
Since $\lim_{k\to\infty} \Pr[\cup_{j=1}^k A_j] = \Pr[\cup_{j=1}^\infty A_j]$, it follows that with probability one, at least one of the events $A_j \equiv \{X_{2(j-1)m+1} = 1, \ldots, X_{2jm} = 1\}$, $j = 1, 2, \ldots$, must occur, hence the claim is verified.

Question 1.2. Is $\Pr[\sup_{n\geq 1} S_n \geq m] = 1$? [see (1.23)]

Question 1.3. Is $\Pr[\, S_n = m$ infinitely often (i.o.)$] = 1$? [Proposition 3.3]

Question 1.4 (Gambler's Ruin). Proposition 1.1 implies that w.pr.1, $\{S_n\}$ hits an asymmetric boundary $\{-m_2, m_1\}$ in finite time. What is

$$p \equiv p_{m_1,m_2} \equiv \Pr[\{S_n\} \text{ hits } m_1 \text{ before } -m_2]? \tag{1.9}$$



Associated with each of these questions is the related question of determining the distribution of the associated (random) *hitting* $\equiv$ *stopping time* N, e.g., the first time that S_n hits the specified boundary. Formally,

$$(1.10)\quad N_m \equiv \min\{n \mid |S_n| = m\}, \quad N_{m_1,m_2} \equiv \min\{n \mid S_n = m_1 \text{ or } -m_2\},$$

etc., are stopping times. These questions will now be addressed.

First consider (1.9). Consider the random variable $S_{N_{m_1,m_2}}$ (well defined since $\Pr[N_{m_1,m_2} < \infty] = 1$), so

$$(1.11)\qquad S_{N_{m_1,m_2}} = \begin{cases} m_1, & \text{with prob. } p; \\ -m_2, & \text{with prob. } 1-p, \end{cases}$$

where $p \equiv p_{m_1,m_2}$ is given by (1.9). Thus

$$(1.12)\qquad \mathrm{E}(S_{N_{m_1,m_2}}) = pm_1 - (1-p)m_2.$$

However we can evaluate $\mathrm{E}(S_{N_{m_1,m_2}})$ in another way, using *Wald's Lemma* for a general random walk (GRW).

Let $Z_1, Z_2, \ldots$ be a sequence of rvs and let N be a random *stopping time* (also called *optional time or Markov time*) based on $\{Z_n \mid n = 1, 2, \ldots\}$. That is, the range of N is $\{1, 2, 3, \ldots\}$ and for each $n = 1, 2, \ldots$, the event $\{N = n\}$ depends only on $Z_1, \ldots, Z_n$. This means that the decision to stop at time n depends only on the values of $Z_1, \ldots, Z_n$, not on any of $Z_{n+1}, \ldots$.

For SSRW, N_{m_1,m_2} is a stopping time for the sequence $\{S_n\}$ [verify]. Note however that $N' \equiv N_{m_1,m_2} - 1$ is *not* a stopping time since the event $\{N' = n\} \equiv \{N_{m_1,m_2} = n+1\}$ depends on the *future* observation S_{n+1}.

Theorem 1.2 *(Wald's Lemma for a randomly stopped sum). Suppose that $Z_1, Z_2, \ldots$ is a sequence of iid rvs with $E|Z_i| < \infty$ and let N be a stopping time for $\{Z_n\}$. If $EN < \infty$, then*

$$(1.13)\qquad \mathrm{E}(S_N) \equiv \mathrm{E}(Z_1 + \ldots + Z_N) = (\mathrm{E}N)(\mathrm{E}Z_1).$$

Proof? $\mathrm{E}(S_N) = \mathrm{E}[\mathrm{E}(Z_1 + \cdots + Z_N | N)] \stackrel{?}{=} \mathrm{E}[N\mathrm{E}(Z_1)] = \mathrm{E}(N)\mathrm{E}(Z_1)$. [No]

Proof. Define the indicator function $1_n \equiv 1_{\{N \geq n\}} = \left\{ \begin{array}{ll} 1, & N \geq n, \\ 0, & N \leq n-1 \end{array} \right\}$. Since N is a stopping time, 1_n depends on $Z_1, \ldots, Z_{n-1}$, so $1_n \perp\!\!\!\perp Z_n$. Thus

$$\begin{aligned} \mathrm{E}(S_N) &= \mathrm{E}\left(\sum_{n=1}^{N} Z_n\right) = \mathrm{E}\left(\sum_{n=1}^{\infty} 1_n Z_n\right) \stackrel{*}{=} \sum_{n=1}^{\infty} \mathrm{E}\left(1_n Z_n\right) \\ &= \sum_{n=1}^{\infty} \mathrm{E}(1_n)\mathrm{E}(Z_n) = \left(\sum_{n=1}^{\infty} \Pr[N \geq n]\right) \mathrm{E}(Z_1) \\ &= \mathrm{E}(N)\mathrm{E}(Z_1) \quad [\text{why?}]. \end{aligned}$$

This completes the proof, except for justification of the interchange of E and $\sum$ at (*). This relies on the *Dominated Convergence Theorem*, which implies that $\mathrm{E}\left(\sum_{n=1}^{\infty} Y_n\right) = \sum_{n=1}^{\infty} \mathrm{E}(Y_n)$ provided that either all $Y_n \geq 0$ or $\mathrm{E}(\sum_{n=1}^{\infty} |Y_n|) < \infty$. Set $Y_n = 1_n Z_n$. Since $|1_n Z_n| \geq 0$ we have

$$\begin{aligned} \mathrm{E}\left(\sum_{n=1}^{\infty} |1_n Z_n|\right) &= \sum_{n=1}^{\infty} \mathrm{E}|1_n Z_n| = \sum_{n=1}^{\infty} \mathrm{E}(1_n)\mathrm{E}|Z_n| \\ &= \left(\sum_{n=1}^{\infty} \Pr[N \geq n]\right) \mathrm{E}|Z_1| \\ &= \mathrm{E}(N)\mathrm{E}|Z_1| < \infty. \end{aligned}$$

Thus the interchange of E and $\sum$ at (*) is justified. □

Now return to (1.12). Recall that $X_1 = \pm 1$ with probs. $(1/2, 1/2)$, so $\mathrm{E}(X_1) = 0$. If we can show $\mathrm{E}(N_{m_1,m_2}) < \infty$ (Proposition 1.2 below), then we can use Wald's Lemma to see that $\mathrm{E}(S_{N_{m_1,m_2}}) = (\mathrm{E}N_{m_1,m_2})(\mathrm{E}X_1) = 0$. Thus from (1.12), $pm_1 - (1-p)m_2 = 0$, so

$$\begin{aligned} p \equiv p_{m_1,m_2} &= \frac{m_2}{m_1 + m_2} \equiv \Pr[\{S_n\} \text{ hits } m_1 \text{ before } -m_2], \\ 1 - p \equiv p_{m_2,m_1} &= \frac{m_1}{m_1 + m_2} \equiv \Pr[\{S_n\} \text{ hits } -m_2 \text{ before } m_1]. \end{aligned} \tag{1.14}$$

Note that this agrees with our intuition that the larger m_i is, the less likely the RW will hit that boundary first.

Gambling interpretation: Suppose that two gamblers (G1 and G2) compete by tossing a fair coin repeatedly. If Heads occurs, then G1 pays G2 \$1, while if Tails occurs then G2 pays G1 \$1. Thus S_n is the amount that G1 has *lost* after n tosses. Then if G1 and G2 initially begin with $\$m_1$ and $\$m_2$ respectively, p_{m_1,m_2} is the probability that G1 is *ruined* (loses all his money) first.

Proposition 1.2. *For SSRW,* $\mathrm{E}(N_{m_1,m_2}) < \infty$.

Proof. $\mathrm{E}(N_{m_1,m_2}) = \sum_{n=1}^{\infty} \Pr[N_{m_1,m_2} \geq n]$. For each n choose k_n so that

$$k_n(m_1+m_2) < n \leq (k_n+1)(m_1+m_2),$$

so $k_n \geq \frac{n}{m_1+m_2} - 1$. Thus

$$\begin{aligned}
&\Pr[N_{m_1,m_2} \geq n+1] \\
&\leq \Pr[\text{no consec. runs of } m_1+m_2 \text{ 1's in } X_1,\ldots,X_n] \\
&\leq \Pr[\text{no consec. runs of } m_1+m_2 \text{ 1's in } X_1,\ldots,X_{m_1+m_2}, \\
&\qquad \ldots, \text{no consec. runs} \ldots \text{in } X_{(k_n-1)(m_1+m_2)+1},\ldots,X_{k_n(m_1+m_2)}] \\
&= \Pr[\cdots]\cdots\Pr[\cdots] \qquad (k_n \text{ terms}) \\
&= (1-\epsilon)^{k_n} \qquad [\epsilon = 2^{-(m_1+m_2)}] \\
&\leq (1-\epsilon)^{\frac{n}{m_1+m_2}-1} \\
&= \tfrac{1}{1-\epsilon}\cdot\delta^n, \qquad (1.15)
\end{aligned}$$

where $\delta = (1-\epsilon)^{(1/(m_1+m_2))} < 1$. Thus

$$\Pr[N_{m_1,m_2} \geq n] \leq \frac{1}{\delta(1-\epsilon)}\delta^n, \quad \text{so} \tag{1.16}$$

$$\mathrm{E}(N_{m_1,m_2}) \leq \frac{1}{\delta(1-\epsilon)} \cdot \sum_{n=1}^{\infty} \delta^n < \infty. \tag{1.17}$$

Exercise 1.2. Use (1.16) to show that $\mathrm{E}(N_{m_1,m_2}^r) < \infty$ for all $r = 1,2,3,\ldots$.
Hint: Show first that for any random variable N with range $\{0,1,\ldots\}$,

$$\mathrm{E}(N^r) \leq r\sum_{n=1}^{\infty} n^{r-1}\Pr[N \geq n]. \tag{1.18}$$

Exercise 1.3. Show that for any random variable N with range $\{0, 1, \ldots\}$,

$$2\sum_{n=2}^{\infty}(n-1)\Pr[N \geq n] \leq \mathrm{E}(N^2) \leq 2\sum_{n=1}^{\infty} n\Pr[N \geq n]. \qquad \square \tag{1.19}$$

For Proposition 1.3 below we need to extend Wald's Lemma to the second moment of a randomly stopped sum. Since the X_i's that comprise a simple RW are bounded ($|X_i| = 1$), the following result suffices:

Exercise 1.4*. Let $Z_1, Z_2, \ldots$ be iid *bounded* rvs, that is, $|Z_i| \leq B < \infty$. Set $\mu = \mathrm{E}(Z_1)$, $\sigma^2 = \mathrm{Var}(Z_1)$. If N is a stopping time for $\{Z_n\}$ with $\mathrm{E}(N^2) < \infty$, then

$$\mathrm{E}(S_N - N\mu)^2 = (\mathrm{E}N)\sigma^2. \tag{1.20}$$

Hint: First show that (1.20) will follow if it can be established for the special case $\mu = 0$. For this case, modify the proof of Wald's Lemma. $\square$

Remark 1.3. Using martingale theory, this result can be established under the much weaker conditions that $\mathrm{E}(Z_i^2) < \infty$ and $\mathrm{E}(N) < \infty$. $\square$

Proposition 1.3. *For SSRW,*

$$\mathrm{E}(N_{m_1,m_2}) = m_1 m_2. \tag{1.21}$$

Proof. $\mathrm{E}(X_1) = 0$, $\mathrm{Var}(X_1) = 1$, and by Exercise 1.2, $\mathrm{E}(N^2_{m_1,m_2}) < \infty$, so by (1.20) with $\mu = 0$ and $\sigma^2 = 1$, $\mathrm{E}(S^2_{N_{m_1,m_2}}) = \mathrm{E}(N_{m_1,m_2})$. But

$$S^2_{N_{m_1,m_2}} = \begin{cases} m_1^2, & \text{with prob. } p = \frac{m_2}{m_1+m_2}, \\ m_2^2, & \text{with prob. } 1-p = \frac{m_1}{m_1+m_2}, \end{cases} \tag{1.22}$$

so (1.21) follows since

$$\mathrm{E}(S^2_{N_{m_1,m_2}}) = \frac{m_1^2 m_2}{m_1+m_2} + \frac{m_2^2 m_1}{m_1+m_2} = m_1 m_2. \qquad \square$$

For the gambling interpretation of SSRW, it follows from (1.21) that if the total $m_1 + m_2$ is fixed, the expected duration of the game is greatest when $m_1 = m_2$, i.e., when the two players have equal initial capitals.

1.2. Gambler's ruin vs. an infinitely rich opponent

For a SSRW $\{S_n\}$, $\Pr[N_{m_1,m_2} < \infty] = 1$ and $\mathrm{E}(N_{m_1,m_2}) < \infty$; that is, with probability one the game will end, and the expected waiting time is finite. Suppose now that $m_2 \to \infty$, that is, your opponent has infinite capital. From (1.14), $p_{m_1,m_2} = \frac{m_2}{m_1+m_2} \to 1$, so

$$p_{m_1,\infty} \equiv \lim_{m_2\to\infty} p_{m_1,m_2} \equiv \Pr[\,\{S_n\} \textit{ eventually } \text{hits } m_1] = 1. \tag{1.23}$$

Thus with probability one, you will lose if you play against an opponent with infinite capital. Equivalently,

$$\Pr[N_{m_1,\infty} < \infty] = 1, \qquad \Pr[N_{\infty,m_2} < \infty] = 1. \tag{1.24}$$

where $N_{m_1,\infty} = \min\{n \mid S_n = m_1\}$, $N_{\infty,m_2} = \min\{n \mid S_n = -m_2\}$.

This has an interesting consequence. Clearly

$$S_{N_{m_1,\infty}} = m_1, \qquad S_{N_{\infty,m_2}} = -m_2, \tag{1.25}$$

so $\mathrm{E}(S_{N_{m_1,\infty}}) = m_1 \neq 0$ and $\mathrm{E}(S_{N_{\infty,m_2}}) = -m_2 \neq 0$. Therefore

$$\mathrm{E}(N_{m_1,\infty}) = \infty, \qquad \mathrm{E}(N_{\infty,m_2}) = \infty, \tag{1.26}$$

for if these were finite then Wald's Lemma yields the contradictions that

$$\begin{aligned} m_1 =& \mathrm{E}(S_{N_{m_1,\infty}}) = (\mathrm{E}N_{m_1,\infty}) \cdot 0 = 0, \\ m_2 =& \mathrm{E}(S_{N_{\infty,m_2}}) = (\mathrm{E}N_{\infty,m_2}) \cdot 0 = 0. \end{aligned}$$

Short "Proof" of (1.26):

$$\mathrm{E}(N_{m_1,\infty}) = \mathrm{E}(\lim_{m_2\to\infty} N_{m_1,m_2}) \stackrel{?}{=} \lim_{m_2\to\infty} \mathrm{E}(N_{m_1,m_2}) = \lim_{m_2\to\infty} m_1 m_2 = \infty.$$

(Wrong: "?" is not necessarily valid.)

Thus we have the almost-paradoxical situation that the expected waiting times to hit the lower and upper boundaries are both infinite, while the expected waiting time to hit *some* boundary is finite! That is,

$$\mathrm{E}(N_{m_1,m_2}) \equiv \mathrm{E}\left(\min(N_{m_1,\infty},\, N_{\infty,m_2})\right) < \infty. \tag{1.27}$$

An explanation is that $N_{m_1,\infty}$ and N_{∞,m_2} are *negatively correlated.* If one of them, say $N_{m_1,\infty}$, is large, then the other, N_{∞,m_2}, will tend to be small.

1.3. Recurrence of states in a SSRW

How often must the SSRW hit a specified state $k \in \{0, \pm 1, \ldots\}$? The surprising answer is that, even though the expected waiting time to hit state k is infinite by (1.26), *state k will be hit infinitely often (i.o.) with probability one!* (Note that $\{S_n = k \text{ i.o.}\}$ is *not* a tail event w.r.t. $\{X_n\}$.)

Definition 1.1. State k is *recurrent (transient)* if

$$\Pr[S_n = k \text{ for some } n \geq 1 \mid S_0 = k] = 1 \ (< 1). \tag{1.28}$$ □

In fact, if state k is recurrent (transient) then $\Pr[S_n = k \text{ i.o.} \mid S_0 = k] = 1 \ (= 0)$ (also see Propositions 3.3, 3.4):

Proposition 1.4. *(i) For SSRW, each state $k = 0, \pm 1, \pm 2, \ldots$ is recurrent.*

(ii) For any states k and l,

$$\Pr[S_n = l \text{ infinitely often} \mid S_0 = k] = 1. \tag{1.29}$$

Proof. (i) It suffices to show this for $k = 0$. Then

$$\begin{aligned}
&\Pr[\{S_n\} \text{ returns to } 0 \mid S_0 = 0] \\
&= \tfrac{1}{2} \Pr[\{S_n\} \text{ returns to } 0 \mid S_1 = 1] + \tfrac{1}{2} \Pr[\{S_n\} \text{ returns to } 0 \mid S_1 = -1] \\
&= \tfrac{1}{2} \Pr[N_{\infty,1} < \infty] + \tfrac{1}{2} \Pr[N_{1,\infty} < \infty] \\
&= 1
\end{aligned}$$

by (1.24) with $m_1 = m_2 = 1$.

(ii) (sketch) By (1.24),

$$\Pr[\{S_n\} \text{ hits } l \mid S_0 = k] = 1. \tag{1.30}$$

Since the first hitting time for state l is a stopping time, the process "renews itself" at this (random) time (this requires the Strong Markov Property), hence by (i), the process now returns to state l infinitely often. □

Remark 1.4. However, the expected waiting time T for a return to any state k is *infinite.* For example, if $k = 0$ then by (1.26),

$$\begin{aligned}
\mathrm{E}(T) &= \tfrac{1}{2}\mathrm{E}[T \mid S_1 = 1] + \tfrac{1}{2}\mathrm{E}[T \mid S_1 = -1] \\
&= \tfrac{1}{2}\mathrm{E}[1 + N_{\infty,1}] + \tfrac{1}{2}\mathrm{E}[1 + N_{1,\infty}] = \infty.
\end{aligned}$$

From Proposition 1.4 *the SSRW in one dimension is recurrent.* We will see that this remains true in two dimensions, *but not in three or more dimensions.* In these higher dimensions SSRW is *transient*, as will be shown shortly. First, however, for 1-dimensional SSRW we study the hitting probabilities for boundaries other than constant boundaries, for example linear boundaries $\{|S_n| = cn\}$. For this we need the following well-known result.

Lemma 1.1 *(The Borel-Cantelli Lemma). Let $\{A_n\}$ be a sequence of events (not necessarily independent). If $\sum \Pr(A_n) < \infty$ then*

$$\Pr[\textit{infinitely many } A_n \textit{ occur}] = 0. \tag{1.31}$$

Proof. Since $\{B_n\} \equiv \{\cup_{k\geq n} A_k\}$ is a decreasing sequence of events and $\sum \Pr(A_n)$ is a convergent infinite series,

$$\begin{aligned}\Pr[\text{inf. many } A_n \text{ occur}] &= \Pr[\cap_{n=1}^{\infty}(\cup_{k\geq n} A_k)] \\ &= \lim_{n\to\infty} \Pr[\cup_{k\geq n} A_k] \\ &\leq \lim_{n\to\infty} \sum\nolimits_{k\geq n} \Pr(A_k) = 0. \qquad \square\end{aligned}$$

Proposition 1.5. *(i) linear boundaries:* $\Pr[|S_n| \geq cn \ i.o.] = 0 \ \ \forall c > 0$.

(ii) square-root-nlog n boundaries: $\Pr[|S_n| \geq \sqrt{cn \log n} \ i.o.] = 0 \ \forall c > 2$.

Proof. (i) By symmetry and Markov's inequality, for any $\lambda > 0$,

$$\begin{aligned}\Pr[|S_n| \geq cn] &= 2\Pr[S_n \geq cn] \\ = 2\Pr[e^{\lambda S_n} \geq e^{\lambda cn}] &\leq 2\,(\mathrm{E}e^{\lambda S_n})/e^{\lambda cn} \\ = 2\,(\mathrm{E}e^{\lambda X_1})^n/e^{\lambda cn} &= 2\,[(e^{\lambda} + e^{-\lambda})/2]^n/e^{\lambda cn} \\ = 2\,[(e^{\lambda(1-c)} + e^{-\lambda(1+c)}/2]^n &\equiv 2\,[g(\lambda)]^n.\end{aligned}$$

But $g(0) = 1$ and $g'(0) = -c < 0$ [verify], so there exists $\tilde{\lambda} > 0$ such that $0 < g(\tilde{\lambda}) < 1$, hence

$$\Pr[|S_n| \geq cn] \leq 2[g(\tilde{\lambda})]^n. \tag{1.32}$$

Thus $\sum \Pr[|S_n| \geq cn] < \infty$ so the result follows from the B-C Lemma.

(ii) (approximate argument:) By the Central Limit Theorem, $S_n/n^{1/2} \approx N(0,1)$, so with Φ and ϕ the $N(0,1)$ cdf and pdf, respectively,

$$\tfrac{1}{2}\Pr[|S_n| \geq \sqrt{cn\log n}] \approx 1 - \Phi(\sqrt{c\log n}) < \frac{\phi(\sqrt{c\log n})}{\sqrt{c\log n}} = \frac{c'}{n^{c/2}\sqrt{\log n}}.$$

Thus the series $\sum \Pr[|S_n| \geq \sqrt{cn\log n}]$ is convergent for $c > 2$, so the result follows from the Borel-Cantelli Lemma. □

Remark 1.5. The inequality (1.32) is an example of a *Chernoff bound* – this appears in Ch. 8 of Ross (the 394/5 textbook).

Remark 1.6. Compare Proposition 1.5(ii) to the LIL for $\{S_n\}$.

1.4. SSRW is transient in dimension $d \geq 3$

There are several ways to extend the definition of SSRW to higher dimensions – we'll consider two ways. Let $\mathbf{Z}^d$ denote the integer lattice in $\mathbf{R}^d$.

1. Combining independent 1-dimensional SSRW's: Set $\mathbf{S}_0 = \mathbf{0} \equiv (0,\ldots,0)$ and for $n \geq 1$,

$$\mathbf{S}_n = \begin{pmatrix} S_n^1 \\ \vdots \\ S_n^d \end{pmatrix} \equiv \begin{pmatrix} X_1^1 + \cdots + X_n^1, \\ \vdots \\ X_1^d + \cdots + X_n^d \end{pmatrix} \in \mathbf{Z}^d, \tag{1.33}$$

where $\{X_i^j\}$ are iid ± 1 rvs. That is, this SSRW comprises d independent 1-dimensional SSRW's, one for each coordinate. This is not a "pure" RW, since its range excludes some of the lattice points in $\mathbf{Z}^d$. [Figure]

As when $d = 1$, with probability one this SSRW must eventually exit any finite neighborhood of the origin. This follows from Proposition 1.1:

$$\Pr\left[\sup_{n\geq 1}\left(\max_{j=1,\ldots,d}|S_n^j|\right) \geq m\right] \geq \Pr\left[\sup_{n\geq 1}|S_n^1| \geq m\right] = 1. \tag{1.34}$$

Furthermore, if we let $N_m^j = \min\{n \mid |S_n^j| = m\}$ for $j = 1,\ldots,d$, then by Proposition 1.2,

$$\mathrm{E}\Big(\min\{n \mid \max_{j=1,\ldots,d}|S_n^j| = m\}\Big) = \mathrm{E}\Big(\min_{j=1,\ldots,d} N_m^j\Big) \leq \mathrm{E}\Big(N_m^1\Big) < \infty. \tag{1.35}$$

For $d \geq 3$, unlike the cases $d = 1$ (or 2), once this SSRW exits a finite neighborhood of the origin, there is a positive probability that it will never return to this neighborhood. This is now demonstrated.

Proposition 1.6. *Suppose that $d \geq 3$.*

(a) Any state $\mathbf{k}$ is transient for $\{\mathbf{S}_n\}$, i.e., $\Pr[\mathbf{S}_n = \mathbf{k}$ i.o. $\mid \mathbf{S}_0 = \mathbf{k}] = 0$.

(b) For any finite set $A \subseteq \mathbf{Z}^d$, $\Pr[\{\mathbf{S}_n\} \in A$ i.o.$] = 0$.

Proof. (a) It suffices to consider $\mathbf{k} = \mathbf{0}$. But $\Pr[\mathbf{S}_n = \mathbf{0}] = 0$ for n odd and

$$\begin{aligned}
\Pr[\mathbf{S}_{2n} = \mathbf{0}] &= \Pr[S_{2n}^1 = 0, \ldots, S_{2n}^d = 0] \\
&= \Pr[S_{2n}^1 = 0] \cdots \Pr[S_{2n}^d = 0] \\
&= \left[\tbinom{2n}{n} \tfrac{1}{2^{2n}}\right]^d \\
&\sim \frac{1}{n^{d/2}} \qquad\qquad \text{as } n \to \infty
\end{aligned}$$

by Stirling's formula ($n! \sim n^n e^{-n}\sqrt{2\pi n}$). Thus $\sum \Pr[\mathbf{S}_n = \mathbf{0}] < \infty$ when $d \geq 3$, so the result follows from the Borel-Cantelli Lemma.

(b) Since A is finite, A recurs infinitely often iff at least one of its states recurs infinitely often, but this has probability 0. □

For dimension $d = 2$, the RW $\{\mathbf{S}_n\}$ remains recurrent. This will follow from the general theory of Markov chains in §3. Thus the critical changeover from recurrence to transience occurs between dimensions 2 and 3 [Polya]. This is a general property of random walks, even in continuous time, e.g., for Brownian motion.

2. A "pure" d-dimensional SSRW: Set $\tilde{\mathbf{S}}_0 = \mathbf{0} \equiv (0, \ldots, 0)$ and for $n \geq 1$,

$$\tilde{\mathbf{S}}_n = \mathbf{U}_1 + \cdots + \mathbf{U}_n, \tag{1.36}$$

where $\mathbf{U}_1, \ldots, \mathbf{U}_n$ are iid rvs with $\Pr[\mathbf{U}_i = \pm\mathbf{e}_j] = \frac{1}{2d}$ for $j = 1, \ldots, d$ and $\mathbf{e}_j$ is the jth coordinate vector $(0, \ldots, 0, 1, 0, \ldots, 0)'$ in $\mathbf{Z}^d$. Here the 1-dimensional component RWs are *not* independent [verify]. But again this RW is recurrent for dimension $d = 2$ but transient for $d \geq 3$.

(Note that when $d = 2$, the process $\{\tilde{\mathbf{S}}_n\}$ can be transformed to the process $\{\mathbf{S}_n\}$ simply by rotating $\mathbf{Z}^2$ by 45°.)

First we observe that, like $\{\mathbf{S}_n\}$, $\{\tilde{\mathbf{S}}_n\}$ must eventually exit any finite neighborhood of the origin.

Exercise 1.5. (i) For any integer $m \geq 1$, show that

$$\Pr\left[\sup_{n\geq 1}\left(\max_{j=1,\dots,d}|\tilde{S}_n^j|\right) \geq m\right] = 1 \tag{1.37}$$

(ii)* Show that (1.35) holds with S_n^j replaced by $\tilde{S}_n^j$.
Hint: Use arguments similar to Propositions 1.1 and 1.2 with $\epsilon = (\frac{1}{2d})^{2m}$.

Again, when $d \geq 3$ there is a positive probability that like $\{\mathbf{S}_n\}$, $\{\tilde{\mathbf{S}}_n\}$ will never return to any finite neighborhood of the origin in $\mathbf{Z}^d$:

Proposition 1.7 *(Polya's Theorem). Suppose that $d \geq 3$.*

(i) Any state $\mathbf{k}$ is transient for $\{\tilde{\mathbf{S}}_n\}$, i.e., $\Pr[\tilde{\mathbf{S}}_n = \mathbf{k}\ i.o. \mid \mathbf{S}_0 = \mathbf{k}] = 0$.

(ii) For any finite set $A \subseteq \mathbf{Z}^d$, $\Pr[\{\tilde{\mathbf{S}}_n\} \in A\ i.o.] = 0$.

Proof. (i) It suffices to consider $\mathbf{k} = \mathbf{0}$. But $\Pr[\tilde{\mathbf{S}}_n = \mathbf{0}] = 0$ for n odd, and $\tilde{\mathbf{S}}_{2n} = 0$ iff the sequence $\mathbf{U}_1, \dots, \mathbf{U}_{2n}$ consists of k_1 $\mathbf{e}_1$'s, k_1 $-\mathbf{e}_1$'s, ..., k_d $\mathbf{e}_d$'s, and k_d $-\mathbf{e}_d$'s, where $k_1, \dots, k_d$ are nonnegative integers such that $k_1 + \cdots + k_d = n$. Therefore by the formula for the cell probabilities of the distribution $\text{Multinomial}_{2d}(2n; \frac{1}{2d}, \dots, \frac{1}{2d})$,

$$\begin{aligned}
\Pr[\tilde{\mathbf{S}}_{2n} = \mathbf{0}] &= \sum_{k_1+\cdots+k_d=n} \frac{(2n)!}{(k_1!)^2\cdots(k_d!)^2}\frac{1}{(2d)^{2n}} \\
&= \frac{(2n)!}{2^{2n}(n!)^2}\sum_{k_1+\cdots+k_d=n}\left(\frac{n!}{k_1!\cdots k_d! d^n}\right)^2 \\
&\equiv \frac{(2n)!}{2^{2n}(n!)^2}\sum_{k_1+\cdots+k_d=n} p_{k_1,\dots,k_d}^2 \\
&\leq \frac{(2n)!}{2^{2n}(n!)^2}\max_{k_1+\cdots+k_d=n} p_{k_1,\dots,k_d}.
\end{aligned} \tag{1.38}$$

where we have used the fact that $\{p_{k_1,\ldots,k_d}\}$ is just the set of all cell probabilities for the distribution $\text{Multinomial}_d(n; \frac{1}{d}, \ldots, \frac{1}{d})$ By Stirling's formula,

$$\frac{(2n)!}{2^{2n}(n!)^2} \sim \frac{1}{(\pi n)^{1/2}} \qquad \text{[verify]} \tag{1.39}$$

for large n, while by Lemma 1.2 below (set $\alpha_j = k_j + 1$ and $\mu = n + d$) and Stirling's formula,

$$p_{k_1,\ldots,k_d} \equiv \frac{n!}{k_1! \cdots k_d! d^n} \leq \frac{n!}{[\Gamma(\frac{n}{d}+1)]^d d^n} \sim \frac{d^{d/2}}{(2\pi n)^{(d-1)/2}}. \tag{1.40}$$

Therefore $\Pr[\tilde{\mathbf{S}}_{2n} = \mathbf{0}] = O(n^{-d/2})$, so $\sum_n \Pr[\tilde{\mathbf{S}}_n = \mathbf{0}] < \infty$ when $d \geq 3$.

(ii) Since A is finite, A recurs infinitely often iff at least one of its states recurs infinitely often, but this has probability 0.

Lemma 1.2. *Let Γ denote the gamma function, $\alpha = (\alpha_1, \ldots, \alpha_d)$, and*

$$\Omega(\mu) = \{\alpha \mid \alpha_1 \geq 1, \ldots, \alpha_d \geq 1, \sum\nolimits_{j=1}^d \alpha_j = \mu\}, \tag{1.41}$$

a simplex in $\mathbf{R}^d$. Then for any $\mu > 0$,

$$\min_{\Omega(\mu)} \prod\nolimits_{j=1}^d \Gamma(\alpha_j) = \left[\Gamma\left(\frac{\mu}{d}\right)\right]^d. \tag{1.42}$$

Proof. First we show that $\gamma(x) \equiv \log \Gamma(x)$ is convex on $(0, \infty)$:

$$\gamma(x) \equiv \log \int_0^\infty t^{x-1} e^{-t} dt, \tag{1.43}$$

$$\gamma'(x) = \frac{\int_0^\infty (\log t) t^{x-1} e^{-t} dt}{\int_0^\infty t^{x-1} e^{-t} dt}, \tag{1.44}$$

$$\gamma''(x) = \frac{\int_0^\infty (\log t)^2 t^{x-1} e^{-t} dt}{\int_0^\infty t^{x-1} e^{-t} dt} - \left[\frac{\int_0^\infty (\log t) t^{x-1} e^{-t} dt}{\int_0^\infty t^{x-1} e^{-t} dt}\right]^2 \overset{\text{[why?]}}{>} 0. \tag{1.45}$$

Therefore

$$\delta(\alpha) \equiv \sum\nolimits_{j=1}^d \gamma(\alpha_j) = \log \prod\nolimits_{j=1}^d \Gamma(\alpha_j)$$

is convex and symmetric ($\equiv$ permutation-invariant) on $\Omega(\mu)$. Together, these properties imply that the minimum of $\delta(\alpha)$ on $\Omega(\mu)$ is attained when $\alpha = (\frac{\mu}{d}, \ldots, \frac{\mu}{d})$ [verify!], which yields (1.42).

1.5. Some exact probability calculations for 1-dimensional SSRW

For any integer $m \geq 1$, now set $N_m = \min\{n | S_n = m\}$, the first time that $S_n = m$. Note that the range of the random stopping time N_m is $\{m, m+2, m+4, \ldots\}$ [why?] Then [verify]:

$$\Pr[S_n = m] = \binom{n}{\frac{n+m}{2}} \frac{1}{2^n} \qquad \text{[binomial distn.]}. \tag{1.46}$$

$$\Pr[N_m = n] = \frac{m}{n} \binom{n}{\frac{n+m}{2}} \frac{1}{2^n} \qquad \text{[Recursion; reflection]}. \tag{1.47}$$

$$\Pr[S_1 > 0, \ldots, S_{n-1} > 0 \mid S_n = m] = \frac{m}{n} \qquad \text{[Ballot Thm.]}. \tag{1.48}$$

Also, let $N_0 = \inf\{n \geq 1 \mid S_n = 0\}$, i.e., N_0 is the first time that the RW *returns* to 0. Note that the range of N_0 is $\{2, 4, 6, \ldots\}$. Then:

$$N_0 \sim 1 + N_1 \qquad \text{[Condition on first step]}. \tag{1.49}$$

$$\begin{aligned} \Pr[N_0 = 2k] &= \frac{1}{2k-1} \binom{2k}{k} \frac{1}{2^{2k}} && \text{[by (1.47)]} \\ &= \frac{1}{2k-1} \Pr[S_{2k} = 0]. && \text{[by (1.46)]} \end{aligned} \tag{1.50}$$

This shows that

$$\Pr[\text{1st return to 0 occurs at time } 2k \mid S_{2k} = 0] = \frac{1}{2k-1}. \tag{1.51}$$

Exercise 1.6*. For any integer $n = 1, 2, \ldots$, establish the possibly surprising result that

$$\Pr[S_{2n} = 0] = \Pr[S_1 \neq 0,\ S_2 \neq 0, \ldots, S_{2n} \neq 0] \tag{1.52}$$

$$= 2\Pr[S_1 > 0, S_2 > 0, \ldots, S_{2n} > 0] \tag{1.53}$$

$$= \Pr[S_1 \geq 0, S_2 \geq 0, \ldots, S_{2n} \geq 0]. \tag{1.54}$$

1.6. The Poisson process.

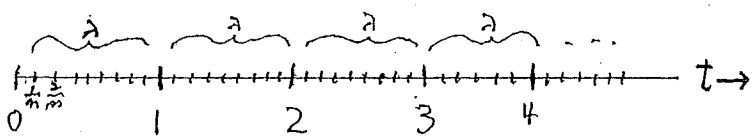
We shall first construct the *Poisson process* (PP) in one dimension. Recall the Poisson approximation to the binomial distribution:

Proposition 1.8. *Let* $N^{(n)} \sim \text{Binomial}(n, p_n)$, *where* $n \to \infty$ *and* $p_n \to 0$ *s.t.* $\mathrm{E}(N^{(n)}) \equiv np_n = \lambda > 0$. *Then* $N^{(n)} \xrightarrow{d} \text{Poisson}(\lambda)$ *as* $n \to \infty$.

[See PK p.23. Note that the range of $N^{(n)}$ is $\{0, 1, \ldots, n\}$, which converges to the Poisson range $\{0, 1, \ldots\}$ as $n \to \infty$.] □

This result says that if a very large number n of elves toss identical coins independently, each with a very small success probability p_n, so that the expected number of successes $np_n = \lambda$, then the total number of successes approximately follows a Poisson distribution. Suppose now that these n elves are spread uniformly over the unit interval $(0, 1]$, and that n more elves with identical coins are spread uniformly over the interval $(1, 2]$, and n more spread over $(2, 3]$, and so on:

Expected:



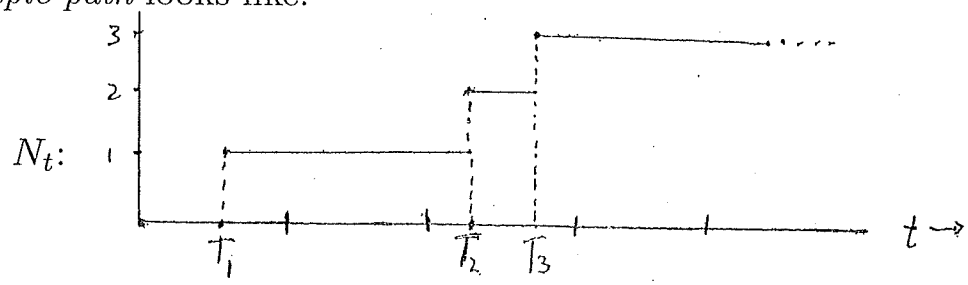
For $0 < t < \infty$, let $N_t^{(n)}$ denote the total number of successes occurring in the interval $(0, t]$ (set $N_0^{(n)} = 0$.) Then $\mathrm{E}(N_t^{(n)}) = \lambda t$, and as $n \to \infty$,

$$N_t^{(n)} \xrightarrow{d} N_t \sim \text{Poisson}(\lambda t),$$
$$\mathrm{E}(N_t) = \lambda t.$$

Because the elves are probabilistically independent and independence is preserved under convergence in distribution, for any sequence of fixed points $0 \equiv t_0 < t_1 < t_2 \cdots$, the increments $N_{t_1} - N_{t_0}$, $N_{t_2} - N_{t_1}$, ..., are mutually independent. Thus the PP has *independent increments* with

$$N_{t_i} - N_{t_{i-1}} \sim \text{Poisson}(\lambda(t_i - t_{i-1})), \tag{1.55}$$
$$\mathrm{E}[N_{t_i} - N_{t_{i-1}}] = \lambda(t_i - t_{i-1}). \tag{1.56}$$

Because of (1.55)-(1.56), the process $\{N_t \mid 0 \le t < \infty\}$ is called a *homogeneous Poisson process* (PP) with *intensity* λ. Its sample paths are nondecreasing step functions with jump size 1. A typical realization $\equiv$ *sample path* looks like:



Here the jump points $0 < T_1 < T_2 < \cdots$ are *random variables.* Such a process is called a *point process* because it is completely determined by the locations of the jump points $T_1, T_2, \ldots$. Considered as a function of t, $\{N_t \mid 0 \le t < \infty\}$ is a *stochastic process* $\equiv$ *random function.*

A *non-homogeneous* PP also can be defined, with λ replaced by an *intensity function* $\lambda(t) \ge 0$. A non-homogeneous PP retains all the above properties of a homogeneous PP except that in (1.55) and (1.56), $(t_i - t_{i-1})\lambda$ is replaced by $\int_{t_{i-1}}^{t_i} \lambda(t)dt$. A non-homogeneous PP can be thought of as the limit of non-homogeneous elf-coin-tossing processes, where the elves are distributed non-uniformly along the line.

Proposition 1.9. *T_1, $T_2 - T_1$, $T_3 - T_2, \ldots$ are i.i.d. Exponential (λ) rvs. In particular, $\mathrm{E}(T_i - T_{i-1}) = \frac{1}{\lambda}$, which reflects the intuitive fact that the expected waiting time to the next success is inversely proportional to the intensity rate λ.*

Partial Proof. $P[T_1 > t] = P[\text{no successes occur in } (0,t]\,] = P[N_t = 0] = e^{-\lambda t}$, since $N_t \sim \text{Poisson}(\lambda t)$. Thus T_1 has pdf $\lambda e^{-\lambda t}$ on $(0, \infty)$. □

Exercise 1.7 *(Completion of the proof of Proposition 1.9).*

(i) Show that T_1 and $T_2 - T_1$ are independent exponential(λ) rvs.

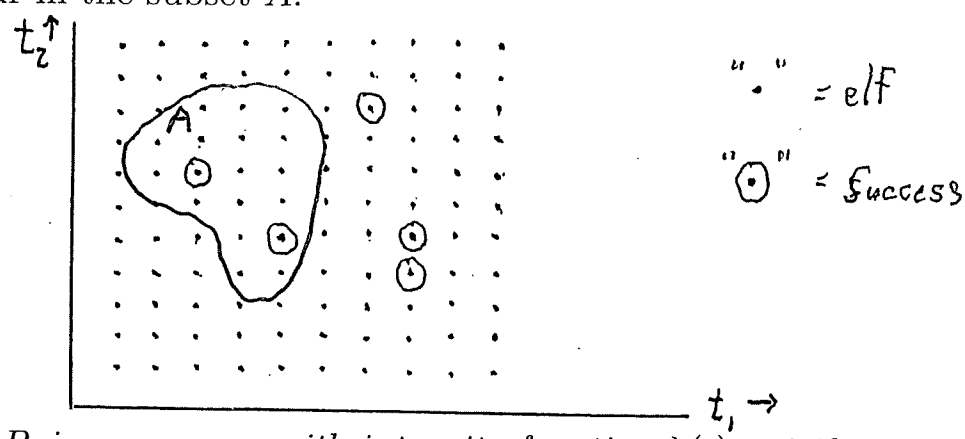
(ii) Show that T_1, $T_2 - T_1$, ..., $T_n - T_{n-1}$ are iid exponential (λ) rvs. □

Remark 1.7. Because the sum of iid exponential rvs has a gamma distribution, $T_k \sim \text{Gamma}(k, \lambda)$. Since $\{T_k > t\} = \{X(t) \le k - 1\}$, this implies a relation between the cdfs of the gamma and Poisson distributions: see PK Theorem 5.4, p.242. □

In view of Proposition 1.9 and the fact that a PP is completely determined by the location of the jumps, there is a duality between a homogeneous PP $\{N_t \mid 0 \le t < \infty\}$ and sums of i.i.d. exponential random variables. Given a sequence of i.i.d. Exponential(λ) rvs $V_1, V_2, \ldots$, define $T_1 = V_1$, $T_2 = V_1 + V_2$, $T_3 = V_1 + V_2 + V_3$, etc. Then $T_1, T_2, T_3, \ldots$ determine the jump points of the PP, from which the entire sample path can be constructed.[4]

PPs arise in many applications, for example as a model for radioactive decay over time. Here, an "elf" is an individual atom – each atom has a tiny probability p of decaying (a "success") in unit time, but there are a large number n of atoms. PPs also serve as models for the number of traffic accidents over time (or location) on a busy freeway.

PPs can be extended in several ways: from homogeneous to non-homogeneous as mentioned above, and/or to point processes in more than one dimension. In general, a point process on an open region $R \subseteq \mathbf{R}^n$ is a random set function $\{N(A) \mid A \subseteq R\}$, where $N(A)$ is the (random) number of points that occur in the subset A.



This constitutes a *Poisson process with intensity function* $\lambda(t) \ge 0$ if

$$N(A) \sim \text{Poisson}\Big(\int_A \lambda(t)dt\Big), \tag{1.57}$$

$$N(A_1) \perp\!\!\!\perp \cdots \perp\!\!\!\perp N(A_k) \quad \text{if } A_1, \ldots, A_k \text{ are disjoint.} \tag{1.58}$$

[4] There must be infinitely many jumps in $(0, \infty)$. This follows from the Borel-Cantelli Theorem, which says that if $\{A_n\}$ is a sequence of *independent* events, then P[infinitely many A_n occur] is 0 or 1 according to whether $\sum P(A_n)$ is $< \infty$ or $= \infty$. Now let A_n be the event that at least one success occurs in the interval $(n-1, n]$.

This can be thought of as a limit of elf-coin-tossing processes where many elves are distributed in R according to density function $\lambda(t)$. The PP is homogeneous if $\lambda(t) \equiv \lambda > 0$ (a constant), otherwise it is non-homogeneous.

Examples of random processes that may be PPs include the spatial distribution of weeds in a field, of ore deposits in a region, of erroneous pixels in a picture transmitted from a Mars orbiter, or of galaxies in the cosmos. These are called "spatial" processes because the random points occur in at random locations in a region. These "may be" PPs because the independence property may not hold if spatial correlation is present.

The waiting-time paradox for a homogeneous Poisson Process.

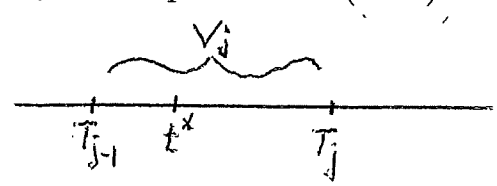
Does it seem that your waiting time for a bus is usually longer than you had expected? This can be explained by the memory-free property of the exponential distribution of the waiting times.

We will model the bus arrival times as the jump times $T_1 < T_2 < \cdots$ of a homogeneous PP $\{N_t\}$ on $[0, \infty)$ with intensity λ. Thus the interarrival times $V_i \equiv T_i - T_{i-1}$ $(i \geq 1,\ T_0 \equiv 0)$ are i.i.d Exponential (λ) rvs and

$$\mathrm{E}(V_i) = \frac{1}{\lambda}, \qquad i \geq 1. \tag{1.59}$$

Now suppose that you arrive at the bus stop at a fixed time $t^* > 0$. Let the index $j \geq 1$ be such that $T_{j-1} < t^* < T_j$ $(j \geq 1)$, so V_j is the length of the interval that contains your arrival time. We expect from (1.59) that

$$\mathrm{E}(V_j) = \frac{1}{\lambda}. \tag{1.60}$$



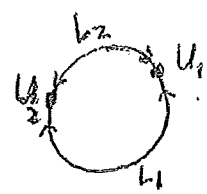
Paradoxically, however,

$$\mathrm{E}(V_j) = \mathrm{E}(T_j - t^*) + \mathrm{E}(t^* - T_{j-1}) \ > \ \mathrm{E}(T_j - t^*) \ = \ \frac{1}{\lambda}, \tag{1.61}$$

since $T_j - t^* \sim \mathrm{Expo}\,(\lambda)$ by the memory-free property of the exponential distribution: if the next bus has not arrived by time t^* then the additional waiting time to the next bus still has the $\mathrm{Expo}\,(\lambda)$ distribution. Thus *you appear always to be unlucky to arrive at the bus stop during a longer-than-average interarrival time!*

This paradox is resolved by noting that the index j is random, not fixed: it is the random index such that V_j includes t^*. The fact that this interval includes a prespecified point t^* tends to make V_j larger than average: a larger interval is more likely to include t^* than a shorter one. Thus it is not so suprising that $\mathrm{E}(V_j) > \frac{1}{\lambda}$.

Exercise 1.8. A simpler example of this phenomenon can be seen as follows. Let U_1 and U_2 be two random points chosen independently and uniformly on the (circumference of the) unit circle and let L_1 and L_2 be the lengths of the two arcs thus obtained:

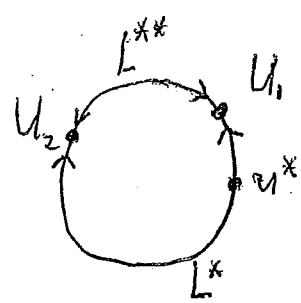


Thus, $L_1 + L_2 = 2\pi$ and $L_1 \stackrel{\mathrm{d}}{=} L_2$ by symmetry, so

$$\mathrm{E}(L_1) = \mathrm{E}(L_2) = \pi. \tag{1.62}$$

(i) Find the distributions of L_1 and of L_2.

(ii) Let L^* denote the length of the arc that contains the point $u^* \equiv (1, 0)$ and let L^{**} be the length of the other arc.



Find the distributions of L^* and L^{**}. Find $\mathrm{E}(L^*)$ and $\mathrm{E}(L^{**})$ and show that $\mathrm{E}(L^*) > \mathrm{E}(L^{**})$.

Hint: There is a simplifying geometric trick. □

Remark 1.8. In (1.61), it is tempting to apply the memory-free property in reverse to assert that also $t^* - T_{j-1} \sim \mathrm{Expo}\,(\lambda)$. This is actually true whenever $j \geq 2$, but not when $j = 1$: $t^* - T_0 \equiv t^* \not\sim \mathrm{Expo}\,(\lambda)$.

This may be achieved, however, by assuming that the bus arrival times $\ldots, T_{-2}, T_{-1}, T_0, T_1, T_2, \ldots$ follow a "doubly-infinite" homogeneous PP on the entire real line $(-\infty, \infty)$. Just as the PP on $(0, \infty)$ can be thought of in terms of many coin-tossing elves spread homogeneously over $(0, \infty)$, this PP can be thought of in terms of many coin-tossing elves spread homogeneously over $(-\infty, \infty)$. The PP properties remain the same, in particular, the interarrival times $T_i - T_{i-1}$ are i.i.d. Exponential (λ) rvs. In this case it *is* true that $t^* - T_{j-1} \sim \text{Expo}\,(\lambda)$, hence we have the *exact* result that

$$\mathrm{E}(V_j) = \frac{2}{\lambda}. \tag{1.63}$$

(In fact, $V_j \sim \text{Expo}\,(\lambda) + \text{Expo}\,(\lambda) \stackrel{\text{d}}{=} \text{Gamma}(2, \lambda)$.) $\square$

2. Markov Chains With Stationary Transition Probabilities

Let $\{X_n \mid n = 0, 1, 2, \ldots\}$ be a Markov chain with state space $\mathcal{S}$, either finite or countable. We assume that the one-step transition probabilities

$$p_{ij} \equiv \Pr[X_{n+1} = j \mid X_n = i], \quad (i, j) \in \mathcal{S}^2,$$

are *stationary*, that is, do not depend on the time n. Then *the joint distribution of the process $\{X_n\}$ is completely determined by the initial distribution $\mathbf{p}_0 \equiv (p_i \mid i \in \mathcal{S})$ of X_0 and the one-step transition probability matrix (tpm)*

$$P \equiv (p_{ij} \mid (i, j) \in \mathcal{S}^2). \tag{2.1}$$

For, by the Markov property, for any n and any $(i_0, i_1, i_2, \ldots, i_n) \in \mathcal{S}^{n+1}$,

$$\begin{aligned}
&\Pr[X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n] \\
&= \Pr[X_0 = i_0] \Pr[X_1 = i_1 | X_0 = i_0] \cdots \Pr[X_n = i_n | X_{n-1} = i_{n-1}] \\
&= p_{i_0} p_{i_0, i_1} \cdots p_{i_{n-1}, i_n}, \qquad (2.2)
\end{aligned}$$

which depends only on $\mathbf{p}_0$ and P. In particular $\Pr[X_1 = i_1] = \sum_{i_0} p_{i_0} p_{i_0, i_1}$, that is, X_1 has distribution given by $\mathbf{p}_0 P$.

What is the *two-step* tpm $P^{(2)} \equiv (p_{ij}^{(2)})$? It is simply given by P^2, since

$$\begin{aligned}
p_{ij}^{(2)} &= \Pr[X_{n+2} = j \mid X_n = i] \quad \text{(use total probability via } X_{n+1}) \\
&= \sum_{k \in \mathcal{S}} \Pr[X_{n+1} = k \mid X_n = i] \Pr[X_{n+2} = j \mid X_{n+1} = k, X_n = i] \\
&= \sum_{k \in \mathcal{S}} \Pr[X_{n+1} = k \mid X_n = i] \Pr[X_{n+2} = j \mid X_{n+1} = k] \\
&= \sum_{k \in \mathcal{S}} p_{ik} p_{kj} \\
&= \text{the } ij\text{th entry of } P^2. \qquad (2.3)
\end{aligned}$$

Similarly, the *n-step* tpm $P^{(n)} = P^n$, the nth power of P. Thus, if X_0 has initial distribution $\mathbf{p}_0$, then X_n has the distribution $\mathbf{p}_0 P^n$, so *the long-term behavior of the Markov chain is determined by that of P^n.*

Example 2.1a *(Simplest nontrivial example: the two-state Markov chain).* Suppose that the state space $\mathcal{S} = \{1, 2\}$ and the tpm is

$$P = \begin{array}{c c} & \begin{array}{cc} 1 & 2 \end{array} \\ \begin{array}{c} 1 \\ 2 \end{array} & \begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix} \end{array}, \qquad 0 \le a,\, b \le 1. \tag{2.4}$$

Note: This process can be realized by tossing two coins, labeled 1 and 2, with $\mathrm{Pr}_1[\text{Heads}] = a$ and $\mathrm{Pr}_2[\text{Heads}] = b$; if Heads occurs we switch coins for the next toss, if Tails occurs the same coin is tossed again. Let X_n denote the coin used for the $(n+1)$st toss.

P^n is easily determined as follows: write

$$P = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} -a & a \\ b & -b \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} -a \\ b \end{pmatrix} (1,\, -1) \equiv I + B. \tag{2.5}$$

Note that $B^2 = -(a+b)B$, so

$$PB = B + B^2 = [1 - (a+b)]B \equiv \delta B,$$

where $\delta = 1 - (a+b)$ satisfies $-1 \le \delta \le 1$. Therefore

$$\begin{aligned} P^2 &= P(I+B) = P + \delta B = I + B + \delta B, \\ P^3 &= P(I + B + \delta B) = \cdots = I + B + \delta B + \delta^2 B, \\ P^n &= I + B + \delta B + \cdots + \delta^{n-1} B \\ &= I + (1 + \delta + \cdots + \delta^{n-1})B \\ &= I + \left(\tfrac{1-\delta^n}{1-\delta}\right) B \quad \text{if } \delta < 1. \end{aligned} \tag{2.6}$$

Case 1: $\delta = 1$. Here $a = b = 0$ so $P = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I$, $P^n = I\ \forall n$. This process is non-random – no transitions ever occur.

Case 2: $\delta = -1$. Here $a = b = 1$ so $P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \equiv J$. Since $J^2 = I$ it follows that $P^n = J$ for n odd and $P^n = I$ for n even, so $\lim_{n\to\infty} P^n$ does not exist. Thus the process simply alternates between states 0 and 1 forever; this behavior is called *periodic.* The following Case 3 is *aperiodic:*

Case 3: $-1 < \delta < 1$. Here $\lim_{n\to\infty} P^n$ exists:

$$\lim_{n\to\infty} P^n = I + \left(\tfrac{1}{1-\delta}\right) B = \begin{pmatrix} \frac{b}{a+b} & \frac{a}{a+b} \\ \frac{b}{a+b} & \frac{a}{a+b} \end{pmatrix} \equiv P^\infty. \tag{2.7}$$

Note that the two rows of P^∞ are *identical*; each is $(\frac{b}{a+b}, \frac{a}{a+b}) \equiv \pi$ which thus gives the long-term probabilities that $X_n = 1$ or 2 regardless of whether $X_0 = 1$ or 2. The distribution $\pi \equiv (\pi_1, \pi_2)$ is called the *limiting distribution* of the chain $\{X_n\}$. Note that π is also a *stationary distribution:*

$$\pi P = \pi, \quad \pi P^2 = \pi, \ \ldots, \ \pi P^n = \pi, \quad \text{so} \ \ \pi P^\infty = \pi, \tag{2.8}$$

which can be verified directly, since $P^\infty = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \pi$.

Subcase 3.1: If $a > 0$ and $b > 0$ then $\pi_1 > 0$ and $\pi_2 > 0$; this case is called *ergodic.* Both states 1 and 2 are recurrent – they occur infinitely often.

Subcase 3.2: If $a > 0$ and $b = 0$ then

$$P = \begin{array}{c c} & \begin{array}{cc} 1 & 2 \end{array} \\ \begin{array}{c} 1 \\ 2 \end{array} & \begin{pmatrix} 1-a & a \\ 0 & 1 \end{pmatrix} \end{array}, \tag{2.9}$$

so state 2 is *absorbing:* once the process enters state 2 it remains there forever. State 1 is *transient:* it can occur at most finitely many times before absorption into state 2. If we let N denote the time to absorption, what is $\mathrm{E}_1(N) \equiv \mathrm{E}(N|X_0 = 1)$? This can be answered by *first-step analysis:*

$$\begin{aligned} \mathrm{E}_1(N) &= \Pr_1[X_1 = 1]\,\mathrm{E}_1[N|X_1 = 1] + \Pr_1[X_1 = 2]\,\mathrm{E}_1[N|X_1 = 2] \\ &= (1-a)(1 + \mathrm{E}_1(N)) + a \cdot 1, \qquad \text{[why?]} \end{aligned}$$

$$\text{so} \ \ \mathrm{E}_1(N) = \frac{1}{a}. \tag{2.10}$$

In fact, first-step analysis can be used to determine the entire distribution of N via its *probability generating function (pgf)* $\psi_1(t) \equiv \mathrm{E}(t^N|X_0 = 1)$:

$$\begin{aligned} \psi_1(t) &= \Pr_1[X_1 = 1]\,\mathrm{E}_1[t^N|X_1 = 1] + \Pr_1[X_1 = 2]\,\mathrm{E}_1[t^N|X_1 = 2] \\ &= (1-a)\mathrm{E}_1(t^{1+N}) + at^1 \qquad \text{[why?]} \\ &= (1-a)t\psi_1(t) + at, \end{aligned}$$

so

$$\psi_1(t) = \frac{at}{1-(1-a)t} = at + (1-a)at^2 + (1-a)^2at^3 + \cdots. \tag{2.11}$$

which is the pgf of the geometric(p) distribution with $p = a$ [interpret]. □

2.1. Finite state space with only transient and absorbing states

Example 2.2 *(One transient state, several absorbing states).*
Extend Subcase 3.2 of Example 2.1a to the following tpm (compare to (2.9)):

$$P = \begin{array}{c} \\ 1 \\ 2 \\ \vdots \\ s \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & \cdots & s \end{array} \\ \begin{pmatrix} q & r_2 & \cdots & r_s \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \end{array} \equiv \begin{pmatrix} q & \mathbf{r} \\ 0 & I \end{pmatrix}, \tag{2.12}$$

where $\mathbf{r} = (r_2, \ldots, r_s)$ and $q + r_2 + \cdots + r_s = 1$. Assume that $q < 1$. Then state 1 is transient while states $2, \ldots, s$ are absorbing.

(a) *Find the absorption probabilities*

$$u_{1j} \equiv \Pr[\text{absortion in state } j \mid X_0 = 1] \equiv \Pr_1[\text{absorption in } j]. \tag{2.13}$$

Educated guess: $u_{1j} \propto r_j$ [why?], so $u_{1j} = r_j/(r_2 + \cdots + r_s) = r_j/(1-q)$.

Verification method 1: "powering P". Show that

$$P^n = \begin{pmatrix} q^n & (1 + q + \cdots + q^{n-1})\mathbf{r} \\ 0 & I \end{pmatrix} = \begin{pmatrix} q^n & \left(\frac{1-q^n}{1-q}\right)\mathbf{r} \\ 0 & I \end{pmatrix} \to \begin{pmatrix} 0 & \frac{\mathbf{r}}{1-q} \\ 0 & I \end{pmatrix}.$$

Verification method 2: "first-step analysis".

$$\begin{aligned} u_{1j} &= \sum\nolimits_{i=1}^{s} \Pr_1[X_1 = i]\Pr_1[\text{absorption in } j \mid X_1 = i] \\ &= qu_{1j} + r_j \cdot 1 + \sum\nolimits_{i \neq 1,j} r_i \cdot 0, \end{aligned} \tag{2.14}$$

so, as surmised,

$$u_{1j} = \frac{r_j}{1-q}. \tag{2.15}$$

(b) *Let N denote the time to absorption. Find the distribution of N and*

$$v_1 \equiv \mathrm{E}[N \mid X_0 = 1] \equiv \mathrm{E}_1(N). \tag{2.16}$$

Educated guess: $N \sim \text{geometric}(1-q)$ [why?], so $\mathrm{E}_1(N) = 1/(1-q)$.

Use first-step analysis to find v_1:

$$\begin{aligned} v_1 &= \sum\nolimits_{i=1}^{s} \Pr_1[X_1 = i]\, \mathrm{E}_1(N|X_1 = i) \\ &= q(1+v_1) + \sum\nolimits_{i\geq 2} r_i \cdot 1, \\ &= q(1+v_1) + (1-q), \end{aligned} \tag{2.17}$$

so, as surmised,

$$v_1 = \frac{1}{1-q}. \tag{2.18}$$

Use first-step analysis to find the pgf $\psi_1(t) \equiv \mathrm{E}_1(t^N)$:

$$\begin{aligned} \psi_1(t) &= \sum\nolimits_{i=1}^{s} \Pr_1[X_1 = i]\, \mathrm{E}_1[t^N|X_1 = i] \\ &= q\mathrm{E}_1(t^{1+N}) + \sum\nolimits_{i\geq 2} r_i \cdot t^1 \\ &= qt\psi_1(t) + (1-q)t, \end{aligned} \tag{2.19}$$

so, as surmised,

$$\psi_1(t) = \frac{(1-q)t}{1-qt} = (1-q)t + q(1-q)t^2 + q^2(1-q)t^3 + \cdots. \tag{2.20}$$

which is the pgf of the geometric$(1-q)$ distribution.

Special case: Suppose that $s = 3$, and relabel the states as $0, +1, -1$:

$$P = \begin{array}{c} \\ 0 \\ +1 \\ -1 \end{array} \begin{array}{c} \begin{array}{ccc} 0 & +1 & -1 \end{array} \\ \begin{pmatrix} q & r_+ & r_- \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \end{array}. \tag{2.21}$$

If $X_0 = 0$, then $\{X_n\}$ is an *asymmetric random walk with absorbing barriers at* ± 1. Then (2.15) and (2.18) imply that

$$\Pr_0[\text{absorption in } \pm 1] = \frac{r_\pm}{r_+ + r_-}, \tag{2.22}$$

$$\mathrm{E}_0(N) = \frac{1}{r_+ + r_-}; \tag{2.23}$$

Note that these agree with (1.14) (p.7) and (1.21) (p.9) for SSRW with $r_+ = r_- = \frac{1}{2}$ and $m_1 = m_2 = 1$. □

Example 2.3a *(Several transient states, several absorbing states).* Extend Example 2.2 to the following tpm (compare to (2.12)):

$$P = \begin{array}{c} \\ 1,\cdots,r \\ r+1,\cdots,s \end{array} \begin{array}{c} \begin{array}{cc} 1,\cdots,r & r+1,\cdots,s \end{array} \\ \begin{pmatrix} Q & R \\ 0 & I \end{pmatrix} \end{array}. \tag{2.24}$$

where $Q = \{q_{ij}\}$, $R = \{r_{ij}\}$, and I denotes the identity matrix. Clearly states $r+1,\ldots,s$ are absorbing, while states $1,\ldots,r$ are transient provided that each row $\mathbf{r}_i$ of R is nonzero.[2]

(a) *Find the matrix* $U \equiv \{u_{ij} \mid i = 1,\ldots,r,\ j = r+1,\ldots,s\}$, *where*

$$u_{ij} \equiv \Pr[\text{absorption in state } j \mid X_0 = i] \equiv \Pr_i[\text{absorption in } j].$$

Educated guess: $\mathbf{u}_i \equiv (u_{i,r+1},\ldots,u_{i,s}) \propto \mathbf{r}_i$? No! (related to Exercise 2.1* and Proposition 2.3.)

Solution method 1. "powering P": Because P is block-triangular,

$$P^n = \begin{pmatrix} Q^n & (I + Q + \cdots + Q^{n-1})R \\ 0 & I \end{pmatrix}. \tag{2.25}$$

Lemma 2.1. *The following two conditions are equivalent:*

(2.26) $\sum_{n=0}^{\infty} Q^n$ *is convergent (in which case* $Q^n \to 0$*);*

(2.27) $(I - Q)^{-1}$ *exists (in which case* $(I - Q)^{-1} = \sum_{n=0}^{\infty} Q^n$*).*

---

[2] By Proposition 3.7, this holds under the weaker condition $R \neq 0$, provided that $\{1,\ldots,r\}$ constitutes a single equivalence class.

Proof. Both implications follow from the identity

$$(I + Q + \cdots + Q^{n-1})(I - Q) = I - Q^n : \tag{2.28}$$

If (2.26) holds, let $n \to \infty$ in (2.28) to obtain $(\sum_{n=0}^{\infty} Q^n)(I - Q) = I$, thus (2.27) holds. Conversely, if (2.27) holds then

$$I + Q + \cdots + Q^{n-1} = (I - Q^n)(I - Q)^{-1} \equiv (I - Q^n)W. \tag{2.29}$$

Because $Q \geq 0$, each component $(I + Q + \cdots + Q^{n-1})_{ij}$ is nondecreasing in n, hence if $\sum_{n=0}^{\infty} Q^n$ fails to converge, its ijth component must diverge to infinity for some i, j. But this leads to a contradiction, since

$$\left|[(I - Q^n)W]_{ij}\right| = \left|\sum\nolimits_{k=1}^{r} (\delta_{ik} - p_{ik}^{(n)}) w_{kj}\right| \leq r \max_{k,j} |w_{kj}| < \infty. \qquad \square$$

Thus if we *assume* that $(I - Q)^{-1}$ exists (see Prop. 3.7), then by (2.25),

$$P^n \to \begin{pmatrix} 0 & (I - Q)^{-1} R \\ 0 & I \end{pmatrix},$$

so

$$U = (I - Q)^{-1} R. \tag{2.30}$$

Solution method 2. first-step analysis:

$$\begin{aligned} u_{ij} &= \sum\nolimits_{k=1}^{s} \Pr_i[X_1 = k] \Pr_i[\text{absorption in } j \mid X_1 = k] \\ &= \sum\nolimits_{k=1}^{r} q_{ik}\, u_{kj} + \sum\nolimits_{k=r+1}^{s} r_{ik}\, \delta_{kj} \qquad \left(\delta_{kj} = \begin{Bmatrix} 0, & \text{if } k \neq j \\ 1, & \text{if } k = j \end{Bmatrix}\right) \\ &= (QU)_{ij} + r_{ij}, \end{aligned}$$

which is equivalent to the matrix equation

$$U = QU + R. \tag{2.31}$$

Thus under assumption (2.27) we again have the solution $U = (I - Q)^{-1} R$.

(b) *Let N denote the time to absorption. Find the distribution of N and find the (column) vector $\mathbf{v} = (v_1, \ldots, v_r)'$, where*

$$v_i \equiv \mathrm{E}[N \mid X_0 = i] \equiv \mathrm{E}_i(N). \tag{2.32}$$

Guess: $\mathcal{L}(N|X_0 = i) = \text{geometric}(1 - q_{i1} - \cdots - q_{ir})$. (No! – Exercise 2.1*).

Finding $\mathbf{v}$. Use first-step analysis to find v_i: for $i = 1, \ldots, r$,

$$\begin{aligned} v_i &= \sum_{k=1}^{s} \Pr_i[X_1 = k] \, \mathrm{E}_i[N|X_1 = k] \\ &= \sum_{k=1}^{r} q_{ik} \, (1 + v_k) + \sum_{k=r+1}^{s} r_{ik} \cdot 1 \\ &= 1 + \sum_{k=1}^{r} q_{ik} \, v_k, \end{aligned} \tag{2.33}$$

which is equivalent to the vector equation

$$\mathbf{v} = \mathbf{e}_r + Q\mathbf{v}, \tag{2.34}$$

where $\mathbf{e}_r = (1, \ldots, 1)' : r \times 1$. Thus if $(I - Q)^{-1}$ exists (cf. (2.27)),

$$\mathbf{v} = (I - Q)^{-1}\mathbf{e}_r. \tag{2.35}$$

Finding the distribution of N. Use first-step analysis to find the vector of pgf's $\Psi(t) \equiv (\psi_1(t), \ldots, \psi_r(t))'$, where

$$\psi_i(t) \equiv \mathrm{E}[t^N \mid X_0 = i] \equiv \mathrm{E}_i(t^N) : \tag{2.36}$$

For $i = 1, \ldots, r$,

$$\begin{aligned} \psi_i(t) &= \sum_{k=1}^{s} \Pr_i[X_1 = k] \, \mathrm{E}_i[t^N|X_1 = k] \\ &= \sum_{k=1}^{r} q_{ik} \, \mathrm{E}_k(t^{1+N}) + \sum_{k=r+1}^{m} r_{ik} \cdot t^1 \\ &= t \left[\sum_{k=1}^{r} q_{ik} \, \psi_k(t) + \sum_{k=r+1}^{s} r_{ik} \right] \\ &= t \left[\mathbf{q}_i \Psi(t) + \mathbf{r}_i \mathbf{e}_{s-r} \right], \end{aligned} \tag{2.37}$$

where $\mathbf{q}_i$ and $\mathbf{r}_i$ are the ith row vectors of Q and R, respectively This is equivalent to the vector equation

$$\Psi(t) = t[Q\Psi(t) + R\mathbf{e}_{s-r}]. \tag{2.38}$$

From (2.24) $Q\mathbf{e}_r + R\mathbf{e}_{s-r} = \mathbf{e}_r$, so for sufficiently small t we have

$$\begin{aligned} \Psi(t) &= t(I - tQ)^{-1} R\mathbf{e}_{s-r} \\ &= t(I - tQ)^{-1}(I - Q)\mathbf{e}_r. \end{aligned} \tag{2.39}$$

□

(c) *Result (2.34)≡(2.35) can be generalized as follows. Let* $g(k)$, $k = 1, \ldots r$, *be a function on the transient states and let* S_g *be the random sum*

$$S_g = \sum_{n=0}^{N-1} g(X_n) \equiv g(X_0) + g(X_1) + \cdots + g(X_{N-1}). \tag{2.40}$$

Find the (column) vector $\mathbf{w}_g = (w_1, \ldots, w_r)'$, *where*

$$w_i = \mathrm{E}_i(S_g). \tag{2.41}$$

Solution. Use first-step analysis to find w_i: for $i = 1, \ldots, r$,

$$\begin{aligned} w_i &= \sum_{k=1}^{s} \Pr_i[X_1 = k]\, \mathrm{E}_i[S_g \mid X_1 = k] \\ &= \sum_{k=1}^{r} q_{ik}(g(i) + w_k) + \sum_{k=r+1}^{s} r_{ik} g(i) \\ &= g(i) + \sum_{k=1}^{r} q_{ik}\, w_k, \end{aligned} \tag{2.42}$$

which is equivalent to the vector equation

$$\mathbf{w}_g = \mathbf{g} + Q\mathbf{w}_g, \tag{2.43}$$

where $\mathbf{g} = (g(1), \ldots, g(r))'$. Thus if $(I - Q)^{-1}$ exists (cf. (2.27)),

$$\mathbf{w}_g = (I - Q)^{-1}\mathbf{g}. \tag{2.44}$$

Result (2.34)≡(2.35) follows by setting $\mathbf{g} = \mathbf{e}_r$, in which case $S_g = N$.

As another application, set $g_j(k) = \delta_{kj}$ for a fixed transient state j. Here S_{g_j} is the total number of visits to state j before termination. Then from (2.44), $\mathbf{w}_{g_j}$ is simply the jth column of $(I - Q)^{-1}$, so

$$(\mathbf{w}_{g_1}, \ldots, \mathbf{w}_{g_r}) = (I - Q)^{-1} \equiv W. \tag{2.45}$$

□

Exercise 2.1*. Criticize the following result and "proof":

Result: Let $A \equiv \{1, \ldots, r\}$ denote the set of transient states and set

$$\delta = \Pr[X_1 \in A | X_0 \in A]. \tag{2.46}$$

Show that for each initial state $i = 1, \ldots, r$, the conditional distribution $\mathcal{L}(N|X_0 = i)$ is "almost" geometric$(1-\delta)$, as in (2.48)-(2.50) below. Furthermore, show that $\delta < 1$, hence $v_i \equiv \mathrm{E}_i(N) < \infty$.

"Proof": For each $n = 1, 2, \ldots,$

$$\begin{aligned}
\Pr_i[N > n] &= \Pr_i[X_1 \in A, \cdots, X_n \in A] \\
&= \Pr_i[X_1 \in A] \Pr_i[X_2 \in A | X_1 \in A] \cdots \Pr_i[X_n \in A | X_{n-1} \in A] && (2.47) \\
&= \mathbf{q}_i \mathbf{e}_r \cdot \delta^{n-1} && (2.48)
\end{aligned}$$

by the Markov property and the stationarity assumption. Therefore [verify]

$$\Pr_i[N = 1] = 1 - \mathbf{q}_i \mathbf{e}_r, \tag{2.49}$$

$$\Pr_i[N = n] = (1-\delta)\delta^{n-2} \cdot \mathbf{q}_i \mathbf{e}_r, \quad n \geq 2. \tag{2.50}$$

Furthermore, with $p_i = \Pr[X_0 = i]$,

$$\begin{aligned}
\delta &= \Pr[X_0 \in A,\, X_1 \in A] / \Pr[X_0 \in A] \\
&= \sum_{i=1}^r \sum_{j=1}^r \Pr[X_0 = i,\, X_1 = j] / \sum_{i=1}^r \Pr[X_0 = i] \\
&= \sum_{i=1}^r \sum_{j=1}^r p_i q_{ij} / \sum_{i=1}^r p_i \\
&= \sum_{i=1}^r p_i \mathbf{q}_i \mathbf{e}_r / \sum_{i=1}^r p_i \\
&\leq \max_i \mathbf{q}_i \mathbf{e}_r < 1
\end{aligned}$$

by the assumption that each row $\mathbf{r}_i$ of R is nonzero. $\square$

Example 2.4 *(Success Runs: PK §3.6.3.)* At each trial $n = 1, 2, \ldots$, there are three possibilities: failure (F), continue (C), or success (S). Let X_n denote the *number of consecutive S's by the end of the nth trial.* The game terminates the first time $X_n = 0$ (an F occurs) or $X_n = m$ (m consecutive S's occur). Then $\{X_n\}$ is a Markov chain with state space $\mathcal{S} = \{0, 1, \ldots, m-1, m\}$ and tpm of the form

$$
(2.51) \qquad P = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ \vdots \\ m-2 \\ m-1 \\ m \end{array}
\begin{array}{c}
\begin{array}{ccccccc} 0 & 1 & 2 & 3 & \cdots & m-1 & m \end{array} \\
\begin{pmatrix}
1 & 0 & 0 & 0 & \cdots & 0 & 0 \\
p_1 & r_1 & q_1 & 0 & \cdots & 0 & 0 \\
p_2 & 0 & r_2 & q_2 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\
p_{m-2} & 0 & 0 & 0 & \cdots & q_{m-2} & 0 \\
p_{m-1} & 0 & 0 & 0 & \cdots & r_{m-1} & q_{m-1} \\
0 & 0 & 0 & 0 & \cdots & 0 & 1
\end{pmatrix}
\end{array}.
$$

Re-order the states to put P in the form (2.24) with $r = m-1$ and $s = 2$:

$$
(2.52) \qquad \begin{array}{c} \\ 1 \\ 2 \\ \vdots \\ m-2 \\ m-1 \\ m \\ 0 \end{array}
\begin{array}{c}
\begin{array}{ccccccc} 1 & 2 & 3 & \cdots & m-1 & m & 0 \end{array} \\
\begin{pmatrix}
r_1 & q_1 & 0 & \cdots & 0 & 0 & p_1 \\
0 & r_2 & q_2 & \cdots & 0 & 0 & p_2 \\
\vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & q_{m-2} & 0 & p_{m-2} \\
0 & 0 & 0 & \cdots & r_{m-1} & q_{m-1} & p_{m-1} \\
0 & 0 & 0 & \cdots & 0 & 1 & 0 \\
0 & 0 & 0 & \cdots & 0 & 0 & 1
\end{pmatrix}
\end{array}
\equiv \begin{pmatrix} Q & R \\ 0 & I \end{pmatrix}.
$$

If each $p_i > 0$ then states $1, \ldots, m-1$ are transient (also see Footnote 2) and m, 0 are absorbing. By (2.30) the matrix $U \equiv \{u_{ij} | 1 \le i \le m-1,\ j = m, 0\}$ of absorption probabilities is given by $U = (I - Q)^{-1} R$, where, by (2.52),

$$
I - Q = \begin{array}{c} \\ 1 \\ 2 \\ \vdots \\ m-2 \\ m-1 \end{array}
\begin{array}{c}
\begin{array}{ccccc} 1 & 2 & 3 & \cdots & m-1 \end{array} \\
\begin{pmatrix}
s_1 & -q_1 & 0 & \cdots & 0 \\
0 & s_2 & -q_2 & \cdots & 0 \\
\vdots & \vdots & \vdots & & \vdots \\
0 & 0 & 0 & \cdots & -q_{m-2} \\
0 & 0 & 0 & \cdots & s_{m-1}
\end{pmatrix}
\end{array}
$$

with $s_i = 1 - r_i = p_i + q_i$. Because $I - Q$ is upper triangular, its inverse exists provided that each diagonal entry $s_i > 0$, in which case $(I-Q)^{-1} \equiv W \equiv \{w_{ij}\}$ is also upper triangular. From the equations $W(I-Q) = (I-Q)W = I$ it is found that

$$\begin{cases} w_{ij} = 0, & \text{if } i > j; \\ w_{ii} = \frac{1}{s_i}; & \\ w_{ij} = \frac{q_i}{s_i}\frac{q_{i+1}}{s_{i+1}} \cdots \frac{q_{j-1}}{s_{j-1}}\frac{1}{s_j}, & \text{if } i < j; \end{cases}$$

hence

$$(2.53) \quad W \equiv (I-Q)^{-1} = \begin{pmatrix} \frac{1}{s_1} & \frac{q_1}{s_1}\frac{1}{s_2} & \frac{q_1}{s_1}\frac{q_2}{s_2}\frac{1}{s_3} & \cdots & \frac{q_1}{s_1} \cdots \frac{q_{m-2}}{s_{m-2}}\frac{1}{s_{m-1}} \\ 0 & \frac{1}{s_2} & \frac{q_2}{s_2}\frac{1}{s_3} & \cdots & \frac{q_2}{s_2} \cdots \frac{q_{m-2}}{s_{m-2}}\frac{1}{s_{m-1}} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & \frac{1}{s_{m-1}} \end{pmatrix}.$$

Thus by (2.30),

$$(2.54) \quad U = \begin{array}{c} \\ 1 \\ 2 \\ \vdots \\ m-1 \end{array} \begin{array}{c} \begin{array}{cc} m & 0 \end{array} \\ \begin{pmatrix} \frac{q_1}{s_1} \cdots \frac{q_{m-1}}{s_{m-1}} & \frac{p_1}{s_1} + \frac{q_1}{s_1}\frac{p_2}{s_2} + \cdots + \frac{q_1}{s_1} \cdots \frac{q_{m-2}}{s_{m-2}}\frac{p_{m-1}}{s_{m-1}} \\ \frac{q_2}{s_2} \cdots \frac{q_{m-1}}{s_{m-1}} & \frac{p_2}{s_2} + \frac{q_2}{s_2}\frac{p_3}{s_3} + \cdots + \frac{q_2}{s_2} \cdots \frac{q_{m-2}}{s_{m-2}}\frac{p_{m-1}}{s_{m-1}} \\ \vdots & \vdots \\ \frac{q_{m-1}}{s_{m-1}} & \frac{p_{m-1}}{s_{m-1}} \end{pmatrix} \end{array}.$$

Note that $\frac{q_i}{s_i}$ and $\frac{p_i}{s_i}$ depend only on the ratio $\frac{p_i}{q_i}$, not on the value of r_i, hence the absorption probabilities $\{u_{ij}\}$ depend only on these ratios. This is expected because a transition $i \to i$ does not affect absorption.

Finally, the vector of expected absorption times $\mathbf{v} = (v_1, \ldots, v_{m-1})'$ is given by $\mathbf{v} = (I-Q)^{-1}\mathbf{e}_{m-1}$ (cf. (2.35)), hence

$$(2.55) \qquad \mathbf{v} = \begin{array}{c} 1 \\ 2 \\ \vdots \\ m-1 \end{array} \begin{pmatrix} \frac{1}{s_1} + \frac{q_1}{s_1}\frac{1}{s_2} + \cdots + \frac{q_1}{s_1} \cdots \frac{q_{m-2}}{s_{m-2}}\frac{1}{s_{m-1}} \\ \frac{1}{s_2} + \frac{q_2}{s_2}\frac{1}{s_3} + \cdots + \frac{q_2}{s_2} \cdots \frac{q_{m-2}}{s_{m-2}}\frac{1}{s_{m-1}} \\ \vdots \\ \frac{1}{s_{m-1}} \end{pmatrix}.$$

Note that $v_{m-1} = \frac{1}{s_{m-1}}$. This is evident because, starting from $X_0 = m-1$, absorption into $\{m, 0\}$ occurs with probability $p_{m-1} + q_{m-1} = s_{m-1}$, while the process continues with probability $r_{m-1} = 1 - s_{m-1}$. Thus the absorption time N has the geometric distribution with $\mathrm{E}_{m-1}(N) = \frac{1}{s_{m-1}}$.

Special case: Suppose that for each $i = 1, \ldots, m-1$, $p_i = p$, $q_i = q$, and $r_i = 0$, so $s_i = 1$. Then from (2.52), (2.54), and (2.55),

$$
(2.56) \qquad P = \begin{array}{c} \\ 1 \\ 2 \\ \vdots \\ m-2 \\ m-1 \\ m \\ 0 \end{array}
\begin{array}{c}
\begin{array}{ccccccc} 1 & 2 & 3 & \cdots & m-1 & m & 0 \end{array} \\
\begin{pmatrix}
0 & q & 0 & \cdots & 0 & 0 & p \\
0 & 0 & q & \cdots & 0 & 0 & p \\
\vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & q & 0 & p \\
0 & 0 & 0 & \cdots & 0 & q & p \\
0 & 0 & 0 & \cdots & 0 & 1 & 0 \\
0 & 0 & 0 & \cdots & 0 & 0 & 1
\end{pmatrix}
\end{array},
$$

(2.57)

$$
U = \begin{array}{c} \\ 1 \\ 2 \\ \vdots \\ m-1 \end{array}
\begin{array}{c}
\begin{array}{cc} m & 0 \end{array} \\
\begin{pmatrix}
q^{m-1} & 1-q^{m-1} \\
q^{m-2} & 1-q^{m-2} \\
\vdots & \vdots \\
q & 1-q
\end{pmatrix}
\end{array}, \quad
\mathbf{v} = \begin{pmatrix} 1 + q + \cdots + q^{m-2} \\ 1 + q + \cdots + q^{m-3} \\ \vdots \\ 1 \end{pmatrix}.
$$

The absorption probabilities $u_{im} = q^{m-i}$ are evident for $i = 1, \ldots, m-1$: because 0 is an absorbing state, if the process starts at state $X_0 = i$ then absorption into state m occurs iff $m - i$ consecutive S's occur.

However, this analysis does *not* give us the value of

$$
(2.58) \qquad u_{0m} \equiv \Pr[m \text{ consecutive S's before first F} \mid X_0 = 0],
$$

because we have formulated the process so that 0 is an absorbing state: if the process starts at $X_0 = 0$ it will never leave the state 0. This difficulty can be overcome by adding one more state to the state space, as follows.

Consider the Markov chain $\{\tilde{X}_0, \tilde{X}_1, \tilde{X}_2, \ldots\}$, where

$$
\begin{aligned}
\tilde{X}_n &= (Y_n, Z_n), \\
Y_n &= \text{no. of consecutive S's by the end of the } n\text{th trial}, \\
Z_n &= \text{no. of consecutive F's by the end of the } n\text{th trial}.
\end{aligned}
$$

The state space for $\{\tilde{X}_n\}$ is $\tilde{\mathcal{S}} = (0,0),(0,1),(1,0),\ldots,(m-1,0),(m,0)$; note that $|\tilde{\mathcal{S}}| = m+2$ whereas $|\mathcal{S}| = m+1$. The process stops if either 1 F or m consecutive S's are obtained, so the states $(0,1)$ and $(m,0)$ are absorbing. To obtain u_{0m} in (2.58) we assume that the process starts in state $(0,0)$. Note that it can never return to $(0,0)$, in fact the tpm is

$$\tilde{P} = \begin{array}{c} \\ (0,0) \\ (1,0) \\ \vdots \\ (m-2,0) \\ (m-1,0) \\ (m,0) \\ (0,1) \end{array} \begin{array}{c} \begin{array}{ccccccc} (0,0) & (1,0) & (2,0) & \cdots & (m-1,0) & (m,0) & (0,1) \end{array} \\ \begin{pmatrix} 0 & q & 0 & \cdots & 0 & 0 & p \\ 0 & 0 & q & \cdots & 0 & 0 & p \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & q & 0 & p \\ 0 & 0 & 0 & \cdots & 0 & q & p \\ 0 & 0 & 0 & \cdots & 0 & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 1 \end{pmatrix} \end{array}$$

$$\equiv \qquad \begin{pmatrix} \tilde{Q} & \tilde{R} \\ 0 & I \end{pmatrix},$$

which has the same form as P in (2.56). Here

$$I - \tilde{Q} = \begin{array}{c} \\ (0,0) \\ (1,0) \\ \vdots \\ (m-2,0) \\ (m-1,0) \end{array} \begin{array}{c} \begin{array}{ccccc} (0,0) & (1,0) & (2,0) & \cdots & (m-1,0) \end{array} \\ \begin{pmatrix} 1 & -q & 0 & \cdots & 0 \\ 0 & 1 & -q & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & -q \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix} \end{array},$$

and [verify]

$$(I - \tilde{Q})^{-1} = \begin{pmatrix} 1 & q & q^2 & \cdots & q^{m-1} \\ 0 & 1 & q & \cdots & q^{m-2} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & q \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix},$$

hence (compare to (2.57))

$$\tilde{U} = (I - \tilde{Q})^{-1}\tilde{R} = \begin{array}{c} \\ (0,0) \\ (1,0) \\ \vdots \\ (m-1,0) \end{array} \begin{array}{c} \begin{array}{cc} (m,0) & (0,1) \end{array} \\ \begin{pmatrix} q^m & 1-q^m \\ q^{m-1} & 1-q^{m-1} \\ \vdots & \vdots \\ q & 1-q \equiv p \end{pmatrix} \end{array},$$

$$\tilde{\mathbf{v}} = (I - \tilde{Q})^{-1}\mathbf{e}_m = \begin{pmatrix} 1+q+\cdots+q^{m-1} \\ 1+q+\cdots+q^{m-2} \\ \vdots \\ 1 \end{pmatrix} \equiv \begin{pmatrix} v_{(0,0)} \\ v_{(1,0)} \\ \vdots \\ v_{(m-1,0)} \end{pmatrix}. \tag{2.59}$$

The *last* $m-1$ rows of $\tilde{U}$ and $\tilde{\mathbf{v}}$ agree with those of U and $\mathbf{v}$ in (2.57), while their *first* rows provide the information missing from the preceding analysis, namely, the absorption probabilities and expected waiting times starting at state $(0,0)$, i.e. starting fresh with no S's and no F's.

A Variant of Success Runs (TK §III.5.4). Instead of stopping after a failure (F), suppose that the process *renews* after each F and continues in this way until m consecutive successes (S) are obtained. As above, let X_n be the number of consecutive S's up to and including the nth trial. Here 0 is *not* an absorbing state; the game terminates the first time $X_n = m$. Then $\{X_n\}$ is a Markov chain [verify] with state space $\mathcal{S} = \{0, 1, \ldots, m-1, m\}$ and tpm of the form

$$P = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ \vdots \\ m-1 \\ m \end{array} \begin{array}{c} \begin{array}{ccccccc} 0 & 1 & 2 & 3 & \cdots & m-1 & m \end{array} \\ \begin{pmatrix} p & q & 0 & 0 & \cdots & 0 & 0 \\ p & 0 & q & 0 & \cdots & 0 & 0 \\ p & 0 & 0 & q & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ p & 0 & 0 & 0 & \cdots & 0 & q \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \end{pmatrix} \end{array}.$$

If $q > 0$, this essentially differs from the tpm in (2.51) only in the first row. The states $0, 1, \ldots, m-1$ are transient while state m is absorbing.

This tpm P is already in the form $\begin{pmatrix} Q & R \\ 0 & I \end{pmatrix}$. Because there is only one absorbing state, obviously $U = \mathbf{e}_m$ (compare to (2.57)). The vector of expected absorption times is given by $\mathbf{v} = (I - Q)^{-1}\mathbf{e}_m$, where

$$I - Q = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ \vdots \\ m-1 \end{array} \begin{array}{c} \begin{array}{cccccc} 0 & 1 & 2 & 3 & \cdots & m-1 \end{array} \\ \begin{pmatrix} q & -q & 0 & 0 & \cdots & 0 \\ -p & 1 & -q & 0 & \cdots & 0 \\ -p & 0 & 1 & -q & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ -p & 0 & 0 & 0 & \cdots & 1 \end{pmatrix} \end{array}.$$

Because $I - Q$ is not triangular, its inverse is not immediately apparent; instead, we can solve the system of linear equations $(I - Q)\mathbf{v} = \mathbf{e}_m$ directly for $\mathbf{v} \equiv (v_0, v_1, \ldots, v_{m-1})'$. This system is

$$\begin{aligned} qv_0 - qv_1 &= 1 \\ -pv_0 + v_1 - qv_2 &= 1 \\ -pv_0 + v_2 - qv_3 &= 1 \\ &\vdots \\ -pv_0 + v_{m-3} - qv_{m-2} &= 1 \\ -pv_0 + v_{m-2} - qv_{m-1} &= 1 \\ -pv_0 + v_{m-1} &= 1. \end{aligned}$$

By subtracting each equation from its predecessor we obtain

$$\begin{aligned} v_0 - v_1 &= \tfrac{1}{q} \\ v_1 - v_2 &= \tfrac{v_0 - v_1}{q} = \tfrac{1}{q^2} \\ v_2 - v_3 &= \tfrac{v_1 - v_2}{q} = \tfrac{1}{q^3} \\ &\vdots \\ v_{m-2} - v_{m-1} &= \tfrac{v_{m-3} - v_{m-2}}{q} = \tfrac{1}{q^{m-1}} \\ v_{m-1} &= \tfrac{v_{m-2} - v_{m-1}}{q} = \tfrac{1}{q^m}, \end{aligned}$$

so by successive addition starting at the bottom we obtain

$$\begin{aligned} v_0 &= \tfrac{1}{q} + \tfrac{1}{q^2} + \cdots + \tfrac{1}{q^m} \\ v_1 &= \tfrac{1}{q^2} + \cdots + \tfrac{1}{q^m} \\ &\vdots \\ v_{m-2} &= \tfrac{1}{q^{m-1}} + \tfrac{1}{q^m} \\ v_{m-1} &= \tfrac{1}{q^m}. \end{aligned}$$

Note: Because $\frac{1}{q} > 1$. these expected absorption times $v_0, v_1 \ldots, v_{m-1}$ are *greater* than $v_{(0,0)}, v_{(1,0)}, \ldots, v_{(m-1,0)}$ in (2.59). This is to be expected since here the process does *not* stop after an F. □

Example 2.5a *(Gambler's Ruin: simple (asymmetric) random walk on the integers* $\{0, 1, \ldots, m\}$ *with absorbing states* 0 *and* m*).*

At each time $n = 1, 2, \ldots$, a gambler wins \$1 with probability p and loses \$1 with pr. $q = 1 - p$, $0 < p < 1$. Thus his total capital at time n is

$$X_n = X_0 + Z_1 + Z_2 + \cdots + Z_n, \tag{2.60}$$

where X_0 is his initial capital and $Z_1, Z_2, \ldots$ are iid rv's with

$$\Pr[Z_k = +1] = p, \quad \Pr[Z_k = -1] = q.$$

The game terminates the first time $X_n = 0$ (the gambler is ruined) or $X_n = m$ (his opponent is ruined), where m is the total capital. Denote the (random) termination time by N. Then $\{X_n\}$ is a Markov chain [verify] with tpm

$$P = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ \vdots \\ m-1 \\ m \end{array} \begin{array}{c} \begin{array}{cccccccc} 0 & 1 & 2 & 3 & \cdots & m-2 & m-1 & m \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ q & 0 & p & 0 & \cdots & 0 & 0 & 0 \\ 0 & q & 0 & p & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & q & 0 & p \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 \end{pmatrix} \end{array}. \tag{2.61}$$

The states $1, \ldots, m-1$ are transient [verify via the method of Proposition 1.1 or apply Proposition 3.7] and 0, m are absorbing, so rewrite P as follows:
(2.62)

$$P = \begin{array}{c} \\ 1 \\ 2 \\ \vdots \\ m-2 \\ m-1 \\ m \\ 0 \end{array} \begin{array}{c} \begin{array}{cccccccc} 1 & 2 & 3 & \cdots & m-2 & m-1 & m & 0 \end{array} \\ \begin{pmatrix} 0 & p & 0 & \cdots & 0 & 0 & 0 & q \\ q & 0 & p & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & p & 0 & 0 \\ 0 & 0 & 0 & \cdots & q & 0 & p & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 1 \end{pmatrix} \end{array} \equiv \begin{pmatrix} Q & R \\ 0 & I \end{pmatrix}.$$

First we find the probability that the gambler eventually wins given that $X_0 = i$, $i = 1, \ldots, m-1$. Since the game must terminate eventually, this probability is

$$u_i \equiv \Pr_i[X_N = m]. \tag{2.63}$$

From (2.31) (p.30) and (2.62) we have the vector equation

$$\begin{pmatrix} u_1 \\ \vdots \\ u_{m-2} \\ u_{m-1} \end{pmatrix} - Q \begin{pmatrix} u_1 \\ \vdots \\ u_{m-2} \\ u_{m-1} \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ p \end{pmatrix}. \tag{2.64}$$

Because $I - Q$ is not triangular, its inverse cannot be written out immediately. However, $I - Q$ is tridiagonal, so a recursive solution of (5) for the u_i's is straightforward:

Write (2.64) as the system of linear equations

$$\begin{aligned} u_1 - pu_2 &= 0 \\ u_2 - qu_1 - pu_3 &= 0 \\ u_3 - qu_2 - pu_4 &= 0 \\ &\vdots \\ u_{m-2} - qu_{m-3} - pu_{m-1} &= 0 \\ u_{m-1} - qu_{m-2} &= p \end{aligned}$$

Now write each initial u_i as $pu_i + qu_i$ and set $u_0 = 0$, $u_m = 1$ to obtain

$$\begin{aligned}
q(u_1 - u_0) &= p(u_2 - u_1) \\
q(u_2 - u_1) &= p(u_3 - u_2) \\
q(u_3 - u_2) &= p(u_4 - u_3) \\
&\vdots \\
q(u_{m-2} - u_{m-3}) &= p(u_{m-1} - u_{m-2}) \\
q(u_{m-1} - u_{m-2}) &= p(u_m - u_{m-1}).
\end{aligned}$$

If we set $\delta = q/p$, it follows that

$$\begin{aligned}
\delta u_1 &= u_2 - u_1 \\
\delta^2 u_1 = \delta(u_2 - u_1) &= u_3 - u_2 \\
&\vdots \\
\delta^{m-2} u_1 = \delta(u_{m-2} - u_{m-3}) &= u_{m-1} - u_{m-2} \\
\delta^{m-1} u_1 = \delta(u_{m-1} - u_{m-2}) &= 1 - u_{m-1},
\end{aligned}$$

so

$$\begin{aligned}
\delta u_1 &= u_2 - u_1 \\
(\delta + \delta^2) u_1 &= u_3 - u_1 \\
&\vdots \\
(\delta + \cdots + \delta^{m-2}) u_1 &= u_{m-1} - u_1 \\
(\delta + \cdots + \delta^{m-1}) u_1 &= 1 - u_1,
\end{aligned}$$

thus

$$\begin{aligned}
(1 + \delta) u_1 &= u_2 \\
(1 + \delta + \delta^2) u_1 &= u_3 \\
&\vdots \\
(1 + \delta + \cdots + \delta^{m-2}) u_1 &= u_{m-1} \\
(1 + \delta + \cdots + \delta^{m-1}) u_1 &= 1,
\end{aligned}$$

hence (start at the end of the system and work back up)

$$u_i = \frac{1+\delta+\cdots+\delta^{i-1}}{1+\delta+\cdots+\delta^{m-1}} \tag{2.65}$$

$$= \begin{cases} \frac{1-\delta^i}{1-\delta^m}, & \text{if } \delta \neq 1; \\ \frac{i}{m}, & \text{if } \delta = 1 \end{cases} \tag{2.66}$$

for $i = 1, \ldots, m-1$. Note that the final result $u_i = i/m$ when $\delta = 1$ (i.e., when $p = q = 1/2$) agrees with our previous result (1.14) for SSRW: the probability that a SSRW hits m_1 before $-m_2$ is $\frac{m_2}{m_1+m_2}$, given that it starts at 0.

Next we find $v_i \equiv \mathrm{E}_i(N)$, the expected time to termination given that $X_0 = i$ $(1 \leq i \leq m-1)$. From (2.34) and (2.62) and $v_i = pv_i + qv_i$,

$$\begin{aligned}
1 = v_1 - pv_2 &= qv_1 - p(v_2 - v_1) \\
1 = v_2 - qv_1 - pv_3 &= q(v_2 - v_1) - p(v_3 - v_2) \\
1 = v_3 - qv_2 - pv_4 &= q(v_3 - v_2) - p(v_4 - v_3) \\
&\vdots \\
1 = v_{m-2} - qv_{m-3} - pv_{m-1} &= q(v_{m-2} - v_{m-3}) - p(v_{m-1} - v_{m-2}) \\
1 = v_{m-1} - qv_{m-2} &= q(v_{m-1} - v_{m-2}) + pv_{m-1},
\end{aligned}$$

so

$$\begin{aligned}
p(v_2 - v_1) &= qv_1 - 1 \\
p(v_3 - v_2) &= q(v_2 - v_1) - 1 \\
p(v_4 - v_3) &= q(v_3 - v_2) - 1 \\
&\vdots \\
p(v_{m-1} - v_{m-2}) &= q(v_{m-2} - v_{m-3}) - 1 \\
-pv_{m-1} &= q(v_{m-1} - v_{m-2}) - 1.
\end{aligned}$$

Now add from the top down to obtain

$$\begin{aligned}
p(v_2 - v_1) &= qv_1 - 1 \\
p(v_3 - v_1) &= qv_2 - 2 \\
p(v_4 - v_1) &= qv_3 - 3 \\
&\vdots \\
p(v_{m-1} - v_1) &= qv_{m-2} - (m-2) \\
-pv_1 &= qv_{m-1} - (m-1),
\end{aligned}$$

hence, with $\delta = q/p$,

$$\begin{aligned}
v_2 &= (1+\delta)v_1 - \tfrac{1}{p} \\
v_3 &= v_1 + \delta v_2 - \tfrac{2}{p} = (1+\delta+\delta^2)v_1 - \tfrac{1}{p}(\delta+2) \\
v_4 &= v_1 + \delta v_3 - \tfrac{3}{p} = (1+\delta+\delta^2+\delta^3)v_1 - \tfrac{1}{p}(\delta^2+2\delta+3) \\
&\vdots \\
v_{m-1} &= v_1 + \delta v_{m-2} - \tfrac{m-2}{p} = (1+\delta+\cdots+\delta^{m-2})v_1 \\
&\quad - \tfrac{1}{p}[\delta^{m-3} + 2\delta^{m-2} + \cdots + (m-3)\delta + (m-2)] \\
0 &= v_1 + \delta v_{m-1} - \tfrac{m-1}{p} = (1+\delta+\cdots+\delta^{m-1})v_1 \\
&\quad - \tfrac{1}{p}[\delta^{m-2} + 2\delta^{m-3} + \cdots + (m-2)\delta + (m-1)].
\end{aligned}$$

The last equation implies that

$$v_1 = \frac{[\delta^{m-2} + 2\delta^{m-3} + \cdots + (m-2)\delta + (m-1)]}{p(1+\delta+\cdots+\delta^{m-1})}, \tag{2.67}$$

which yields $v_2, \ldots, v_{m-1}$ by substitution in the remaining equations.

Some algebra yields the following simplified expressions [verify!]:

$$v_i = \begin{cases} \frac{1}{p-q}\left[\frac{m(1-\delta^i)}{1-\delta^m} - i\right], & \text{if } \delta \neq 1, \\ i(m-i), & \text{if } \delta = 1. \end{cases} \tag{2.68}$$

for $i = 1, \ldots, m-1$. Note again that the result $v_i = i(m-i)$ when $\delta = 1$ agrees with our previous result $\mathrm{E}(N_{m_1,m_2}) = m_1 m_2$ in Proposition 1.3 for

SSRW, where N_{m_1,m_2} is the first time that the SSRW hits m_1 or $-m_2$, given that it starts at 0. $\square$

2.2. Finite state space: limiting distributions

As above, let $\{X_n\}$ be a Markov chain with finite state space $\mathcal{S} = \{1, \ldots, s\}$ and tpm P. Let $\pi \equiv (\pi_1, \ldots, \pi_s)$ be an s-dimensional row vector and let $\mathbf{e} \equiv \mathbf{e}_s \equiv (1, \ldots, 1)'$ denote the s-dimensional column vector of 1's.

Definition 2.1. π is a *limiting distribution* for $\{X_n\}$ if

$$\lim_{n\to\infty} P^n = \begin{pmatrix} \pi \\ \vdots \\ \pi \end{pmatrix} \equiv \mathbf{e}\,\pi. \tag{2.69}$$

$\square$

Note that (2.69) requires that π is the limiting distribution regardless of the initial state. It is possible that $\lim P^n$ exists yet no limiting distribution exists according to Definition 2.1: simply take $P = I$ (the identity matrix).

Proposition 2.1. *If a limiting distribution π exists, then:*

(i) π is a probability distribution, i.e., each $\pi_j \geq 0$ and $\sum \pi_j = 1$.

(ii) π is the unique solution to the system of linear equations

$$\pi = \pi P, \qquad \pi\,\mathbf{e} = 1. \tag{2.71}$$

Proof. (i) $\pi_j = \lim_{n\to\infty}(P^n)_{ij} \geq 0$. Also P is a stochastic matrix, so

$$\mathbf{e} = P\mathbf{e} = P^2\mathbf{e} = \cdots = P^n\mathbf{e} \to (\mathbf{e}\,\pi)\mathbf{e} = \mathbf{e}(\pi\,\mathbf{e}) = \mathbf{e}(\textstyle\sum \pi_j), \tag{2.70}$$

hence $\sum \pi_j = 1$.

(ii) $\mathbf{e}\,\pi = \lim P^{n+1} = (\lim P^n)\,P = (\mathbf{e}\,\pi)P$, so

$$s\pi = \mathbf{e}'\mathbf{e}\,\pi = \mathbf{e}'\mathbf{e}\,\pi\,P = s\pi P,$$

hence π satisfies $\pi = \pi\,P$. To show uniqueness, suppose that $\nu = \nu P$ and $\nu\,\mathbf{e} = 1$. Then

$$\nu = \nu\,P = \nu\,P^2 = \cdots = \nu\,P^n \to \nu(\mathbf{e}\,\pi) = (\nu\mathbf{e})\pi = \pi. \qquad \square$$

Example 2.6 *(doubly stochastic matrix).* The tpm P is *doubly stochastic* if its rows sum to 1 *and* its columns sum to 1. The second condition can be expressed as $\mathbf{e}'P = \mathbf{e}'$, hence $\pi \equiv \frac{1}{s}\mathbf{e}' = (\frac{1}{s}, \ldots, \frac{1}{s})$ satisfies (2.71). Thus *if a limiting distribution exists it must be* π, so all states are equally likely as $n \to \infty$. □

Interpretation of π (when it exists):

(a) If we take the initial distribution $\mathbf{p}_0$ of X_0 to be π then the unconditional distribution of X_1 is $\pi P = \pi$, hence by induction, *the unconditional distribution of X_n is π for all X_n.* Thus *the limiting distribution π is the unique stationary distribution for the chain.*

(b) For *any* initial distribution $\mathbf{p}_0 \equiv (p_1, \ldots, p_s)$,

$$\mathcal{L}(X_n) = \mathbf{p}_0 P^n \to \mathbf{p}_0(\mathbf{e}\,\pi) = (\mathbf{p}_0\mathbf{e})\pi = \pi \quad \text{as } n \to \infty. \tag{2.72}$$

Thus, regardless of the initial distribution, for large n, $X_n, X_{n+1}, \ldots$ behaves like a sequence of *identically distributed* (but not necessarily independent) observations of a random variable distributed over $(1, \ldots, s)$ with probabilities $(\pi_1, \ldots, \pi_s)$. This suggests that $\{X_n\}$ may satisfy the (Weak) Law of Large Numbers: for any function g on $\mathcal{S} \equiv (1, \ldots, s)$,

$$\frac{1}{n}\sum_{k=1}^{n} g(X_k) \xrightarrow{p} \mathrm{E}\left[g(X_\pi)\right] \equiv \sum_{j=1}^{s} g(j)\pi_j. \tag{2.73}$$

This is true! – see Theorem 2.2 below for a special case.

Example 2.1b *(the two-state Markov chain, continued).* Here $\mathcal{S} = \{1, 2\}$,

$$P = \begin{array}{c} \\ 1 \\ 2 \end{array}\!\!\begin{array}{c} \begin{array}{cc} 1 & 2 \end{array} \\ \begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix} \end{array} \qquad (0 \le a,\, b \le 1), \tag{2.4$'$}$$

and

$$P^n = I + \left(\tfrac{1-\delta^n}{1-\delta}\right)\begin{pmatrix} -a & a \\ b & -b \end{pmatrix}, \tag{2.6$'$}$$

where $\delta = 1 - a - b$ $(-1 \le \delta \le 1)$.

Case 1: $\delta = 1$. Here $a = b = 0$ so $P = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I$, $P^n = I\ \forall n$. Since both states are absorbing, there are two (trivial) "limiting distributions": $(1,0)$ (entered only from state 1) and $(0,1)$ (entered only from state 2).

Case 2: $\delta = -1$. Here $a = b = 1$ so $P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \equiv J$, $P^n = J$ for n odd and $P^n = I$ for n even, so $\lim_{n\to\infty} P^n$ does not exist. (This behavior is called *periodic*). The following Case 3 is *aperiodic:*

Case 3: $-1 < \delta < 1$. A limiting distribution π exists (necessarily unique):

$$\lim_{n\to\infty} P^n = \begin{pmatrix} \frac{b}{a+b} & \frac{a}{a+b} \\ \frac{b}{a+b} & \frac{a}{a+b} \end{pmatrix} \equiv \mathbf{e}\,\pi, \tag{2.7$'$}$$

where $\pi = (\frac{b}{a+b}, \frac{a}{a+b})$.

Subcase 3.2: If $a > 0$ and $b = 0$ then state 1 is transient and state 2 is absorbing:

$$P = \begin{pmatrix} 1-a & a \\ 0 & 1 \end{pmatrix}. \tag{2.9$'$}$$

Here $\pi_1 = 0$ because $b = 0$, so $\pi = (0,1)$ and $\mathbf{e}\,\pi = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}$. We have already studied this case so will not consider it further.

Subcase 3.1: If $a > 0$ and $b > 0$ then both $\pi_1 > 0$ and $\pi_2 > 0$. There are no transient or absorbing states – both states 1 and 2 are recurrent, i.e., they occur infinitely often. This case is called *ergodic* and will be the case of most interest to us henceforth. □

There are two main approaches to the study of the limiting distribution in the ergodic case: *algebraic* and *geometric.*

2.3. Algebraic approach via the eigenvectors and eigenvalues of P

The *eigenvalues* $\lambda_1, \ldots, \lambda_s$ of P are the roots (real or complex) of the sth order polynomial equation $|P - \lambda I| = 0$. For each root λ_i the matrix $P - \lambda_i I$

is singular, so there exists a nonzero row vector x_i and a nonzero column vector y_i (both possibly complex) such that

$$x_i(P - \lambda_i I) = 0 \quad \text{and} \quad (P - \lambda_i I)y_i = 0,$$

hence

$$x_i P = \lambda_i x_i \quad \text{and} \quad P y_i = \lambda_i y_i. \tag{2.74}$$

Thus x_i (resp., y_i) is called a *left (right) eigenvector* with eigenvalue λ_i.

Because P is stochastic, $P\mathbf{e} = \mathbf{e}$, so $\mathbf{e}$ is a right eigenvector of P with eigenvalue $\lambda_1 \equiv 1$. Also, if P has a limiting distribution π then π is stationary, i.e., $\pi P = \pi$, so π is a left eigenvector of P with eigenvalue $\lambda_1 \equiv 1$.

Lemma 2.2. *Each eigenvalue λ of P satisfies $|\lambda| \le 1$. If $\lim_{n\to\infty} P^n$ exists then either $\lambda = 1$ or $|\lambda| < 1$. (That is, $\lambda \ne e^{i\theta}$ for $\theta \ne 2k\pi$.)*

Proof. From (2.74), $\lambda z = Pz$ for some nonzero $z \equiv (z_1, \ldots, z_s)'$. Let $|z_i| = \max(|z_1|, \ldots, |z_s|) > 0$. Then

$$|\lambda|\,|z_i| = |\lambda z_i| = |(Pz)_i| = \left|\sum\nolimits_j p_{ij} z_j\right| \le \sum\nolimits_j p_{ij}|z_j| \le |z_i|,$$

hence $|\lambda| \le 1$. Furthermore, $\lambda^n z = P^n z$ which is convergent by assumption, hence λ^n must be convergent, so either $\lambda = 1$ or $|\lambda| < 1$. □

We now show that under the following diagonalizability assumptions D1 and D2, P has a limiting distribution.

Assumption D1. *P is similar to a diagonal matrix.* That is,

$$P = SDS^{-1} \tag{2.75}$$

for some nonsingular matrix S and some diagonal matrix D. □

Assumption D2. $|\lambda_2| < 1, \ldots, |\lambda_s| < 1$. □

If D1 holds then $|P - \lambda I| = |D - \lambda I|$, so the diagonal elements of D are the eigenvalues $\lambda_1 (\equiv 1), \ldots, \lambda_s$ of P, that is

$$D = D_\lambda \equiv \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_s \end{pmatrix}. \tag{2.76}$$

Furthermore, $PS = SD_\lambda$ and $S^{-1}P = D_\lambda S^{-1}$, so the column vectors $\mathbf{s}_1, \ldots, \mathbf{s}_s$ of S (row vectors $\mathbf{t}_1, \ldots, \mathbf{t}_s$ of S^{-1}) are right (left) eigenvectors of P. (Note that we can take $\mathbf{s}_1 = \mathbf{e}$.) Then for each $n = 1, 2, \ldots$,

$$P^n = SD_\lambda^n S^{-1} = S \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & \lambda_2^n & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_s^n \end{pmatrix} S^{-1}. \tag{2.77}$$

Thus if D2 also holds, then as $n \to \infty$,

$$P^n \to S \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 \end{pmatrix} S^{-1} = \mathbf{e}\,\mathbf{t}_1. \tag{2.78}$$

Thus, *if assumptions D1 and D2 hold, P has limiting distribution $\pi \equiv \mathbf{t}_1$.*

Example 2.1c *(the two-state Markov chain, continued).* From (2.4)′,

$$P = \begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix} = \begin{pmatrix} 1 & a \\ 1 & -b \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1-a-b \end{pmatrix} \begin{pmatrix} \frac{b}{a+b} & \frac{a}{a+b} \\ \frac{1}{a+b} & \frac{-1}{a+b} \end{pmatrix},$$

which is of the form $SD_\lambda S^{-1}$ [verify!] with $\lambda_2 = 1 - a - b \equiv \delta$. Thus, when $|\delta| < 1$, P has limiting distribution $\pi = \left(\frac{b}{a+b}, \frac{a}{a+b}\right)$ (recall (2.7)′).

Exercise 2.2. (i) Show directly that the eigenvalues of P are 1 and δ by solving the quadratic equation $|P - \lambda I| = 0$.

(ii) Find a right eigenvector $\mathbf{s}_2$ of P for the eigenvalue δ by solving the vector equation $P\mathbf{s}_2 = \delta\mathbf{s}_2$.

(iii) Verify that $\pi \equiv \left(\frac{b}{a+b}, \frac{a}{a+b}\right)$ is the unique solution to the system of equations $\pi = \pi P$, $\pi\,\mathbf{e} = 1$ (recall (2.71)). □

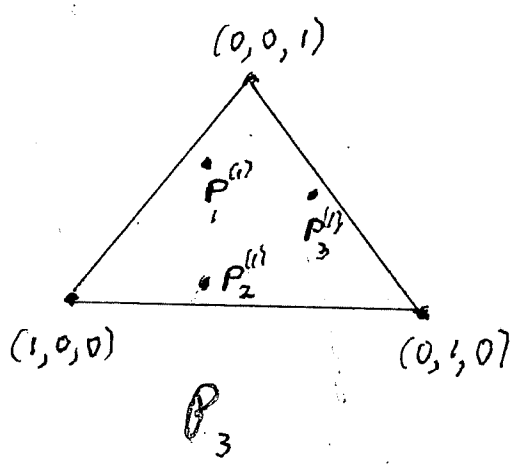
Exercise 2.3. For a general $s \times s$ tpm P show that a *sufficient* condition that D1 hold is that the s eigenvalues of P are distinct. Show by a counterexample that this condition is not necessary for D1 to hold. □

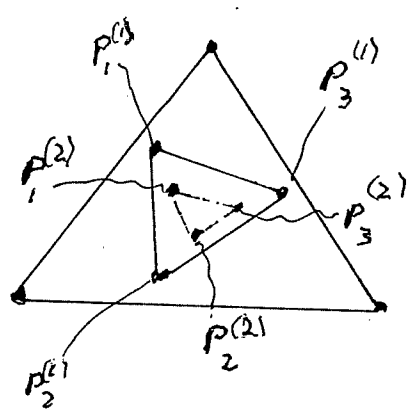
2.4. Geometric approach for a regular tpm

The geometric approach to establishing the existence of a limiting distribution is often easier to apply than the algebraic approach. Denote the row vectors of $P \equiv P^{(1)}$ by $\mathbf{p}_1 \equiv \mathbf{p}_1^{(1)}, \ldots, \mathbf{p}_s \equiv \mathbf{p}_s^{(1)}$. Each $\mathbf{p}_i^{(1)}$ lies in the s-dimensional *probability simplex*

$$\mathcal{P}_s := \{\mathbf{p} \equiv (p_1, \ldots, p_s) \mid p_i \geq 0, \sum p_i = 1\}. \tag{2.79}$$

Thus $P^{(1)}$ can be viewed as an array of vectors $(\mathbf{p}_1^{(1)}, \ldots, \mathbf{p}_s^{(1)}) \in \mathcal{P}_s$:





Then

$$P^{(2)} = P^2 = P\begin{pmatrix} \mathbf{p}_1^{(1)} \\ \vdots \\ \mathbf{p}_s^{(1)} \end{pmatrix} = \begin{pmatrix} \sum p_{1j}\mathbf{p}_j^{(1)} \\ \vdots \\ \sum p_{sj}\mathbf{p}_j^{(1)} \end{pmatrix} \equiv \begin{pmatrix} \mathbf{p}_1^{(2)} \\ \vdots \\ \mathbf{p}_s^{(2)} \end{pmatrix},$$

so *each row* $\mathbf{p}_i^{(2)}$ *of* $P^{(2)}$ *is a weighted average of the rows of* $P^{(1)}$, *that is,* $\mathbf{p}_i^{(2)}$ *lies in the simplex with vertices* $\mathbf{p}_1^{(1)}, \ldots, \mathbf{p}_s^{(1)}$. By induction, *each row* $\mathbf{p}_i^{(n+1)}$ *of* $P^{(n+1)}$ *lies in the simplex with vertices* $\mathbf{p}_1^{(n)}, \ldots, \mathbf{p}_s^{(n)}$.

Since these simplices are nested (decreasing with n), this suggests that they converge to a single point π in the interior $\mathcal{P}_s^o$ of $\mathcal{P}_s$, that is, $\mathbf{p}_i^{(n)} \to \pi$ for $i = 1, \ldots, s$. This holds for the important case where P is *regular*.

Definition 2.2. The $s \times s$ tpm P is *regular* if $P^{(\nu)} \equiv P^\nu > 0$ for some $\nu = \nu(P) \geq 1$, that is, if $p_{ij}^{(\nu)} > 0$ for all $1 \leq i, j \leq s$. □

Example 2.7. A 3×3 tpm P of the form

$$P = \begin{pmatrix} 0 & + & + \\ 0 & 0 & + \\ + & 0 & 0 \end{pmatrix} \tag{2.80}$$

is regular: $P^5 > 0$ but not P, P^2, P^3, or P^4 [verify]. However, a tpm of the form $\begin{pmatrix} Q & R \\ 0 & I \end{pmatrix}$ is not regular. In fact, a regular tpm can have no absorbing states [obvious] or $(*)$ transient states [apply Prop. 3.3 and Thm. 2.1]. □

Exercise 2.4. Criticize the following "proof" of $(*)$: If state i is transient, $\Pr[N < \infty] = 1$, where N is the *last* time n that $X_n = i$. By regularity, $p_{ii}^{(\nu)} > 0$ for some $\nu \geq 1$, so $\Pr[X_{N+\nu} = i | X_N = i] > 0$ by the Markov property. However $X_{N+\nu} \neq i$ by the definition of N, a contradiction. □

Theorem 2.1. *P is regular $\Leftrightarrow$ it has a limiting distribution $\pi \equiv (\pi_1, \ldots, \pi_s)$ in $\mathcal{P}_s^o$, that is, each $\pi_j > 0$. In fact, each $\mathbf{p}_i^{(n)} \to \pi$ at a geometric rate: there exist $c > 0$ and $0 < \rho < 1$ such that for all $i, j = 1, \ldots, s$ and all $n \geq 1$,*

$$|p_{ij}^{(n)} - \pi_j| \leq c\,\rho^n. \tag{2.81}$$

Proof. ($\Rightarrow$) Define $m_j^{(n)} = \min_i p_{ij}^{(n)}$ and $M_j^{(n)} = \max_i p_{ij}^{(n)}$, the column minima and maxima of $P^{(n)}$, respectively. Because each row of $P^{(n+1)}$ is a weighted average of the rows of $P^{(n)}$,

$$m_j^{(n)} \leq m_j^{(n+1)} \leq M_j^{(n+1)} \leq M_j^{(n)}. \tag{2.82}$$

It follows from the relation $P^{\nu+n} = P^\nu P^n$ that

$$p_{ij}^{(\nu+n)} = \sum_{l=1}^{s} p_{il}^{(\nu)} p_{lj}^{(n)}, \tag{2.83}$$

a weighted average of $p_{1j}^{(n)}, \ldots, p_{sj}^{(n)}$ with weights $p_{i1}^{(\nu)}, \ldots, p_{is}^{(\nu)}$. Because P is regular, $m \equiv \min_{i,j} p_{ij}^{(\nu)} > 0$ for some $\nu \geq 1$ (necessarily $m \leq 1/s$). Thus

$$m_j^{(\nu+n)} \geq \left\{[1-(s-1)m]m_j^{(n)} + m\sum_{l=2}^{s} p_{(l)j}^{(n)}\right\}, \tag{2.84}$$

$$M_j^{(\nu+n)} \leq \left\{m\sum_{l=1}^{s-1} p_{(l)j}^{(n)} + [1-(s-1)m]M_j^{(n)}\right\}, \tag{2.85}$$

where $m_j^{(n)} \equiv p_{(1)j}^{(n)} \leq \cdots \leq p_{(s)j}^{(n)} \equiv M_j^{(n)}$ are the ordered values of $p_{1j}^{(n)}, \ldots, p_{sj}^{(n)}$. Thus for all $n = 1, 2, \ldots$,

$$\begin{aligned} M_j^{(\nu+n)} - m_j^{(\nu+n)} &\leq [1-(s-1)m]\big(M_j^{(n)} - m_j^{(n)}\big) + m\big(p_{(1)j}^{(n)} - p_{(s)j}^{(n)}\big) \\ &= (1-sm)(M_j^{(n)} - m_j^{(n)}). \end{aligned} \tag{2.86}$$

Replace n by $\nu(r-1)+q$ for $r = 1, 2, \ldots$ and $q = 0, 1, \ldots, \nu-1$, then iterate to obtain

$$\begin{aligned} M_j^{(\nu r+q)} - m_j^{(\nu r+q)} &\leq (M_j^{(\nu(r-1)+q)} - m_j^{(\nu(r-1)+q)})(1-sm) \\ &\leq (M_j^{(\nu(r-2)+q)} - m_j^{(\nu(r-2)+q)})(1-sm)^2 \\ &\vdots \\ &\leq (M_j^{(q)} - m_j^{(q)})(1-sm)^r \\ &\leq (1-sm)^r. \end{aligned} \tag{2.87}$$

Thus $M_j^{(\nu r+q)} - m_j^{(\nu r+q)} \to 0$ as $r \to \infty$, hence by (2.82),

$$\lim_{n\to\infty} m_j^{(n)} = \lim_{n\to\infty} M_j^{(n)} \equiv \pi_j. \tag{2.88}$$

Because $m \leq m_j^{(n)} \leq p_{ij}^{(n)} \leq M_j^{(n)}$ for $n \geq \nu$, it follows that

$$\lim_{n\to\infty} p_{ij}^{(n)} = \pi_j \geq m > 0. \tag{2.89}$$

In fact, set $n = \nu r + q$ in (2.87) to obtain that for all $n > \nu$,

$$|p_{ij}^{(n)} - \pi_j| \le M_j^{(n)} - m_j^{(n)} \le (1 - sm)^{\frac{n-q}{\nu}} \le (1 - sm)^{\frac{n-\nu}{\nu}} \equiv c'\,\rho^n,$$

where $\rho = (1 - sm)^{\frac{1}{\nu}} < 1$ and $c' = (1 - m)^{-1}$. By increasing c' if necessary we can insure that (2.81) holds for all $n \ge 1$.

($\Leftarrow$) This follows from (2.89) and the assumption that each $\pi_j > 0$. $\square$

Theorem 2.2 *(The Weak Law of Large Numbers for a regular Markov chain with finite state space). If P is regular with limiting distribution π, then for any function g on $\mathcal{S} \equiv (1, \ldots, s)$ and any initial state i,*

$$\text{(2.73)}' \qquad \frac{1}{n}\sum_{k=1}^{n} g(X_k) \xrightarrow{p} \mathrm{E}\left[g(X_\pi)\right] \equiv \sum_{j=1}^{s} g(j)\pi_j.$$

Proof. Begin with the special case where $g(x) = 1_j(x)$, the indicator function of the single state j. Here

$$\hat{\pi}_{n,j} \equiv \tfrac{1}{n}\sum_{k=1}^{n} 1_j(X_k)$$

is the proportion of times $k = 1, \ldots, n$ such that $X_k = j$, and $\mathrm{E}[1_j(X_\pi)] = \pi_j$. We wish to show that for every $\delta > 0$,

$$\text{(2.90)} \qquad \Pr_i[\,|\hat{\pi}_{n,j} - \pi_j| > \delta\,] \to 0 \quad \text{as } n \to \infty.$$

As with the WLLN for iid random variables, it suffices to show that

$$\text{(2.91)} \qquad \mathrm{E}_i[(\hat{\pi}_{n,j} - \pi_j)^2] \to 0 \quad \text{as } n \to \infty$$

and then apply Chebyshev's inequality to obtain (2.90). Now

$$\begin{aligned}
&\mathrm{E}_i[(\hat{\pi}_{n,j} - \pi_j)^2] \\
&= \mathrm{E}_i\left[\left(\tfrac{1}{n}\sum_{k=1}^{n}\{1_j(X_k) - \pi_j\}\right)^2\right] \\
&= \frac{1}{n^2}\mathrm{E}_i\left[\sum_{k=1}^{n}\sum_{l=1}^{n} 1_j(X_k)1_j(X_l) - \{1_j(X_k) + 1_j(X_l)\}\pi_j + \pi_j^2\right]
\end{aligned}$$

$$= \frac{1}{n^2}\sum_{k=1}^{n}\sum_{l=1}^{n}\left[\Pr_i\{X_k = j, X_l = j\} - (\Pr_i\{X_k = j\} + \Pr_i\{X_l = j\})\pi_j + \pi_j^2\right]$$
$$= \frac{1}{n^2}\sum_{k=1}^{n}\sum_{l=1}^{n}\left[p_{ij}^{(s)}p_{jj}^{(t)} - (p_{ij}^{(k)} + p_{ij}^{(l)})\pi_j + \pi_j^2\right]$$
$$\equiv \frac{1}{n^2}\sum_{k=1}^{n}\sum_{l=1}^{n} b_{ij}^{(k,l)},$$

where $s = \min(k,l)$ and $t = \max(k,l) - \min(k,l)$. By (2.81), however,

$$\begin{aligned}|b_{ij}^{(k,l)}| &= |(p_{ij}^{(s)} - \pi_j)p_{jj}^{(t)} + \pi_j(p_{jj}^{(t)} - \pi_j) + \pi_j(\pi_j - p_{ij}^{(k)}) + \pi_j(\pi_j - p_{ij}^{(l)})| \\ &\le c\,(\rho^s + \rho^t + \rho^k + \rho^l) \\ &\le c\,[(\rho^k + \rho^l) + \rho^t + \rho^k + \rho^l],\end{aligned}$$

so

$$\mathrm{E}_i[(\hat{\pi}_{n,j} - \pi_j)^2] \le \frac{c}{n^2}\sum_{k=1}^{n}\sum_{l=1}^{n}[2(\rho^k + \rho^l) + \rho^t].$$

But

$$\frac{1}{n^2}\sum_{k=1}^{n}\sum_{l=1}^{n}(\rho^k + \rho^l) = \frac{2n\rho(1-\rho^n)}{n^2(1-\rho)} \le \frac{2}{n(1-\rho)} \to 0 \quad \text{as } n \to \infty,$$
$$\frac{1}{n^2}\sum_{k=1}^{n}\sum_{l=1}^{n}\rho^t \overset{(?)}{\le} \frac{2n(1-\rho^n)}{n^2(1-\rho)} \le \frac{2}{n(1-\rho)} \to 0 \quad \text{as } n \to \infty,$$

hence (2.91) holds, which implies (2.90).

Finally consider (2.73)′:

$$\begin{aligned}\frac{1}{n}\sum_{k=1}^{n} g(X_k) &= \frac{1}{n}\sum_{k=1}^{n}\sum_{j=1}^{s} g(j)1_j(X_k)\\ &= \sum_{j=1}^{s} g(j)\left[\frac{1}{n}\sum_{k=1}^{n} 1_j(X_k)\right]\\ &\equiv \sum_{j=1}^{s} g(j)\,\hat{\pi}_{n,j}\\ &\xrightarrow{p} \sum_{j=1}^{s} g(j)\,\pi_j \equiv \mathrm{E}\left[g(X_\pi)\right]\end{aligned}$$

by (2.90), so (2.73)′ is valid. □

Exercise 2.5. Verify the inequality "(?)".

Exercise 2.6. (i) P is regular only if (but not if) $P + P^2 + \cdots + P^s > 0$.
(ii)** P is regular if and only if $P^{(s^2)} > 0$.

Remark 2.1. Since $[\mathrm{E}_i(\hat{\pi}_{n,j} - \pi_j)]^2 \le \mathrm{E}_i[(\hat{\pi}_{n,j} - \pi_j)^2] \to 0$ as $n \to \infty$,

$$\mathrm{E}_i(\hat{\pi}_{n,j}) \to \pi_j \quad \text{for each } j = 1, \ldots, s.$$

(This is the result given on page 173 of PK.) □

Remark 2.2. The Strong Law of Large Numbers also holds for a regular Markov chain [Reference??]: If P is regular with limiting distribution π, then for any function g on $\mathcal{S} \equiv (1, \ldots, s)$ and any initial state i,

$$\frac{1}{n}\sum_{k=1}^{n} g(X_k) \overset{w.pr.1}{\to} \mathrm{E}\left[g(X_\pi)\right] \equiv \sum_{j=1}^{s} g(j)\pi_j. \qquad \square$$

Let $N_j \ge 1$ denote *the time of the first visit to state j after time* 0 and let $\mu_{ij} = \mathrm{E}_i(N_j) \ge 1$. Thus $\mu_{jj} \equiv \mathrm{E}_j(N_j)$ is *the mean return time to state j*. There is a simple relation between μ_{jj} and the limiting probability π_j.

Proposition 2.2. *Let $\pi \equiv (\pi_1, \pi_2, \ldots)$ be a stationary distribution for the Markov chain $\{X_n\}$ with finite or countable state space $\mathcal{S}$. For a fixed state $j \in \mathcal{S}$ assume that $\sum_i \pi_i \mu_{ij} < \infty$, e.g., if $\mu_{ij} \le B_j < \infty$ $\forall i \in \mathcal{S}$. Then*

$$\pi_j \mu_{jj} = 1, \quad \text{so} \quad \pi_j > 0 \iff \mu_{jj} < \infty. \tag{2.92}$$

Proof. Condition on the outcome of the first step to obtain

$$\begin{aligned}\mu_{ij} = \mathrm{E}_i(N_j) &= \sum_k \Pr_i[X_1 = k]\,\mathrm{E}[N_j \mid X_1 = k] \\ &= p_{ij} \cdot 1 + \sum_{k \neq j} p_{ik}\,\mathrm{E}[N_j \mid X_1 = k] \\ &= p_{ij} + \sum_{k \neq j} p_{ik}\,\{1 + \mathrm{E}_k(N_j)\} \\ &= 1 + \sum_{k \neq j} p_{ik}\,\mu_{kj}.\end{aligned}$$

Therefore

$$\begin{aligned}\infty > \sum_i \pi_i \mu_{ij} &= \sum_i \pi_i + \sum_i \pi_i \sum_{k \neq j} p_{ik}\,\mu_{kj} \\ &= 1 + \sum_{k \neq j} \Big(\sum_i \pi_i p_{ik} \Big) \mu_{kj} \\ &= 1 + \sum_{k \neq j} \pi_k \mu_{kj}\end{aligned}$$

by the stationarity of π, hence $\pi_j \mu_{jj} = 1$, yielding (2.92). □

A sufficient condition for the boundedness of μ_{ij} over i is now given:

Proposition 2.3. *(i) Let $\{X_n\}$ be a Markov chain with finite or countable state space $\mathcal{S}$. Let A be a subset of $\mathcal{S}$ such that $m_A \equiv \min_{i \notin A} \Pr_i^{(\nu)}(A) > 0$ for some integer $\nu \equiv \nu(A) \geq 1$. Then for each initial state $i \in \mathcal{S}$,*

$$\mathrm{E}_i(N_A) \leq \nu\Big(\frac{1 + m_A}{m_A}\Big) < \infty, \tag{2.93}$$

where N_A is the time of the first visit to A after time 0.

(ii) If $m_j \equiv \min_{i \neq j} p_{ij}^{(\nu)} > 0$ for some integer $\nu \equiv \nu(j) \geq 1$, then

$$\mu_{ij} \leq \nu\Big(\frac{1 + m_j}{m_j}\Big) < \infty \quad \forall i \in \mathcal{S}. \tag{2.94}$$

In particular, this holds if $\mathcal{S}$ is finite and $\{X_n\}$ is regular.

Proof. (i) (Compare to Exercise 2.1*.) $\Pr_i[N_A \geq k]$ is nonincreasing in k, so

$$\begin{aligned}
\mathrm{E}_i(N_A) &= \sum_{k\geq 0} \Pr_i[N_A > k] = \sum_{n\geq 0}\sum_{j=0}^{\nu-1} \Pr_i[N_A > n\nu + j] \\
&\leq \nu \sum_{n\geq 0} \Pr_i[N_A > n\nu] \\
&\leq \nu\Big(2 + \sum_{n\geq 2} \Pr_i[X_\nu \notin A, \ldots, X_{n\nu} \notin A]\Big).
\end{aligned}$$

But for $n \geq 2$,

$$\begin{aligned}
&\Pr_i[X_\nu \notin A, \ldots, X_{n\nu} \notin A] \\
&\leq \prod_{k=2}^{n} \Pr_i[X_{k\nu} \notin A | X_\nu \notin A, \ldots, X_{(k-1)\nu} \notin A] \\
&= \prod_{k=2}^{n} \sum_{i_k \notin A} \Pr_i[X_{k\nu} = i_k | X_\nu \notin A, \ldots, X_{(k-1)\nu} \notin A] \\
&= \prod_{k=2}^{n} \sum_{i_k \notin A} \frac{\Pr_i[X_\nu \notin A, \ldots, X_{(k-1)\nu} \notin A, X_{k\nu} = i_k]}{\Pr_i[X_\nu \notin A, \ldots, X_{(k-1)\nu} \notin A]} \\
&= \prod_{k=2}^{n} \sum_{i_k \notin A} \frac{\sum_{i_1 \notin A, \ldots, i_{k-1} \notin A} \Pr_i[X_\nu = i_1, \ldots, X_{(k-1)\nu} = i_{k-1}, X_{k\nu} = i_k]}{\Pr_i[X_\nu \notin A, \ldots, X_{(k-1)\nu} \notin A]} \\
&= \prod_{k=2}^{n} \sum_{i_k \notin A} \frac{\sum_{i_1 \notin A, \ldots, i_{k-1} \notin A} \Pr_i[X_\nu = i_1, \ldots, X_{(k-1)\nu} = i_{k-1}]\, p^{(\nu)}_{i_{k-1}, i_k}}{\Pr_i[X_\nu \notin A, \ldots, X_{(k-1)\nu} \notin A]} \\
&= \prod_{k=2}^{n} \frac{\sum_{i_1 \notin A, \ldots, i_{k-1} \notin A} \Pr_i[X_\nu = i_1, \ldots, X_{(k-1)\nu} = i_{k-1}] \sum_{i_k \notin A} p^{(\nu)}_{i_{k-1}, i_k}}{\Pr_i[X_\nu \notin A, \ldots, X_{(k-1)\nu} \notin A]} \\
&= \prod_{k=2}^{n} \frac{\sum_{i_1 \notin A, \ldots, i_{k-1} \notin A} \Pr_i[X_\nu = i_1, \ldots, X_{(k-1)\nu} = i_{k-1}](1 - \Pr^{(\nu)}_{i_{k-1}}(A))}{\Pr_i[X_\nu \notin A, \ldots, X_{(k-1)\nu} \notin A]} \\
&\leq (1 - m_A)^{n-1}.
\end{aligned}$$

Thus $\mathrm{E}_i(N_A) < \nu\big(2 + \sum_{n\geq 2}(1 - m_A)^{n-1}\big) = \nu\big(\frac{1+m_A}{m_A}\big)$.

(ii) This follows from (i) by setting $A = \{j\}$. $\square$

Example 2.8a *(Simple random walk on a circle).* Consider the tpm

$$(2.95) \qquad P = \begin{array}{c} \\ 0 \\ 1 \\ \vdots \\ s-2 \\ s-1 \end{array} \begin{array}{c} \begin{array}{ccccccc} 0 & 1 & 2 & \cdots & s-3 & s-2 & s-1 \end{array} \\ \begin{pmatrix} r_0 & p_0 & 0 & \cdots & 0 & 0 & q_0 \\ q_1 & r_1 & p_1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & q_{s-2} & r_{s-2} & p_{s-2} \\ p_{s-1} & 0 & 0 & \cdots & 0 & q_{s-1} & r_{s-1} \end{pmatrix} \end{array}.$$

This represents a simple random walk on the circle: the states $0, 1, \ldots, s-1$ are arranged clockwise around the circle; from state i we move one step clockwise (counterclockwise) with probability p_i (q_i) and stay in state i with probability r_i. If P is regular then a unique limiting $\equiv$ stationary distribution $\pi \equiv (\pi_0, \ldots, \pi_{s-1}) > 0$ exists by Theorem 2.1 and satisfies $\pi = \pi P$, $\pi\,\mathbf{e} = 1$.

Exercise 2.7. Assume that all $p_i > 0$ and either
(i) some $r_i > 0$, or
(ii)* s is odd and some $q_i > 0$.
Show that P is regular with $\nu(P) = 2s - 2$ in (i) and $\nu(P) =$??** in (ii)*. (Regularity also follows from Prop. 3.6(ii).) $\square$

For the special case where the chain is *rotationally invariant* $\equiv$ *circularly symmetric*, i.e., $(p_i, q_i, r_i) = (p, q, r)$ for each $i = 0, \ldots, s-1$, then

$$(2.96) \qquad P = \begin{pmatrix} r & p & 0 & \cdots & 0 & 0 & q \\ q & r & p & \cdots & 0 & 0 & 0 \\ 0 & q & r & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & q & r & p \\ p & 0 & 0 & \cdots & 0 & q & r \end{pmatrix}.$$

Assume that $0 < p < 1$ and *either* $r > 0$, *or* s *is odd and* $q > 0$, so P is regular and doubly stochastic (Exercise 2.7). Thus the unique stationary distribution is $\pi = (\frac{1}{s}, \ldots, \frac{1}{s})$, so all states are equally likely in the limit. Furthermore, by (2.92), the mean return time $\mu_{jj} = s$ for each j. $\square$

Example 2.9 *(Rotationally invariant ≡ circularly symmetric random walk on the circle).* Consider the *cyclic* tpm

$$P = \begin{pmatrix} p_0 & p_1 & p_2 & \cdots & p_{s-2} & p_{s-1} \\ p_{s-1} & p_0 & p_1 & \cdots & p_{s-3} & p_{s-2} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ p_2 & p_3 & p_4 & \cdots & p_0 & p_1 \\ p_1 & p_2 & p_3 & \cdots & p_{s-1} & p_0 \end{pmatrix}, \tag{2.97}$$

where $p_0 + \ldots + p_{s-1} = 1$; clearly P is doubly stochastic. To avoid periodicity (see Example 2.10), assume that $(*)$ all $p_i > 0$. Then P is regular, so again the unique limiting distribution exists and is given by $\pi = (\frac{1}{s}, \ldots, \frac{1}{s})$. The mean return time is $\mu_{jj} = s$ for each j. □

Exercise 2.8. Show that $(*)$ can be weakened to $(**)$: $p_i > 0$ and $p_j > 0$ for some $0 \le i < j \le s-1$ such that $j - i$ and s are relatively prime. □

Example 2.10 *(An asymmetric simple random walk on the circle).* Consider the 4-state Markov chain with tpm

$$P = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{pmatrix} 0 & .50 & 0 & .50 \\ .75 & 0 & .25 & 0 \\ 0 & .50 & 0 & .50 \\ .75 & 0 & .25 & 0 \end{pmatrix} \end{array}. \tag{2.98}$$

This represents a random walk on the states $0, 1, 2, 3$ arranged clockwise around a circle: from state 0 we move clockwise to 1 and counterclockwise to 3, each with probability .5; from state 1 we move clockwise to 2 with probability .25 and counterclockwise to 0 with probability .75, etc. From the figure [draw it] we expect that if a limiting distribution $\pi \equiv (\pi_0, \pi_1, \pi_2, \pi_3)$ exists, it will satisfy

$$\pi_0 > \pi_1 = \pi_3 > \pi_2. \qquad \text{[why?]} \tag{2.99}$$

To verify this, solve the stationary equations $\pi = \pi P$ with $\pi\,\mathbf{e} = 1$:

$$\begin{aligned} \pi_0 &= .75\pi_1 + .75\pi_3 \\ \pi_1 &= .50\pi_0 + .50\pi_2 \\ \pi_2 &= .25\pi_1 + .25\pi_3 \\ \pi_3 &= .50\pi_0 + .50\pi_2. \end{aligned}$$

Because $\pi_1 = \pi_3$ by symmetry [see figure], this reduces to the system

$$\begin{aligned}\pi_0 &= 1.50\pi_1\\ \pi_1 &= .50\pi_0 + .50\pi_2\\ \pi_2 &= .50\pi_1,\end{aligned}$$

so $\pi_0 = 3\pi_2$, hence $\pi_1 = 2\pi_2$, thus $(\pi_0, \pi_1, \pi_2, \pi_3) = (3\pi_2, 2\pi_2, \pi_2, 2\pi_2)$. Therefore $\pi_2 = 1/8$ and $\pi = (3/8, 2/8, 1/8, 2/8)$, which confirms (2.99). The mean return times are $\mu_{00} = 8/3$, $\mu_{11} = 4$, $\mu_{22} = 8$, $\mu_{33} = 4$.

However, no limiting distribution exists according to Definition 2.1: if we rearrange P in the form

$$\tilde{P} = \begin{array}{c} \\ 0\\ 2\\ 1\\ 3 \end{array}\begin{array}{c} \begin{array}{cccc} 0 & 2 & 1 & 3 \end{array}\\ \begin{pmatrix} 0 & 0 & .50 & .50\\ 0 & 0 & .50 & .50\\ .75 & .25 & 0 & 0\\ .75 & .25 & 0 & 0 \end{pmatrix}\end{array} \tag{2.100}$$

then $\tilde{P}$ *is periodic with period 2.* Specifically,

$$\tilde{P}^2 = \begin{array}{c} \\ 0\\ 2\\ 1\\ 3 \end{array}\begin{array}{c} \begin{array}{cccc} 0 & 2 & 1 & 3 \end{array}\\ \begin{pmatrix} .75 & .25 & 0 & 0\\ .75 & .25 & 0 & 0\\ 0 & 0 & .50 & .50\\ 0 & 0 & .50 & .50 \end{pmatrix}\end{array}, \tag{2.101}$$

$\tilde{P}^3 = \tilde{P}$, $\tilde{P}^4 = \tilde{P}^2$, etc. Thus $\tilde{P}^n$ does not converge, but rather alternates between the two forms (2.100) and (2.101). Nonetheless, note that

$$\frac{1}{2}\left(\tilde{P} + \tilde{P}^2\right) = \begin{array}{c} \\ 0\\ 2\\ 1\\ 3 \end{array}\begin{array}{c} \begin{array}{cccc} 0 & 2 & 1 & 3 \end{array}\\ \begin{pmatrix} 3/8 & 1/8 & 2/8 & 2/8\\ 3/8 & 1/8 & 2/8 & 2/8\\ 3/8 & 1/8 & 2/8 & 2/8\\ 3/8 & 1/8 & 2/8 & 2/8 \end{pmatrix}\end{array} \equiv \mathbf{e}\,\tilde{\pi}, \tag{2.102}$$

where $\tilde{\pi}$ is just the corresponding rearrangement $(\pi_0, \pi_2, \pi_1, \pi_3)$ of π. Thus $\tilde{\pi}$ *does* represent the long-term distribution if averaged over odd and even times n. (Also see my solution to PK Problem 4.1.2.) $\square$.

2.5. Converting a non-Markovian chain to a Markov chain

Suppose that for some integer $r \geq 2$, X_n depends on not only its immediate predecessor X_{n-1} but on its r predecessors $X_{n-r}, \ldots, X_{n-1}$. That is,

$$\mathcal{L}(X_n \mid \ldots, X_{n-r-1}, X_{n-r}, \ldots, X_{n-1}) = \mathcal{L}(X_n \mid X_{n-r}, \ldots, X_{n-1}) \tag{2.103}$$

for $n \geq r$. The chain $\{X_n\}$, though not Markovian, can be converted to a Markov chain $\{Y_n\}$ by grouping the observations in groups of r: if we define

$$Y_n = (X_{n-r+1}, \ldots, X_n), \tag{2.104}$$

then it is easily shown that $\{Y_n \mid n = r, r+1, \ldots\}$ is Markovian.

Exercise 2.9. Show that $\{Y_n\}$ in (2.104) satisfies the Markov property. □

Example 2.11a. Consider four coins labeled 1, 2, 3, 4 with Pr[Heads] = p_1, p_2, p_3, p_4, respectively. Denote the outcome of the nth toss by X_n, where

$$\text{coin} \begin{cases} 1 \text{ is tossed if} & (X_{n-2}, X_{n-1}) = (H, H), \\ 2 \text{ is tossed if} & (X_{n-2}, X_{n-1}) = (H, T), \\ 3 \text{ is tossed if} & (X_{n-2}, X_{n-1}) = (T, H), \\ 4 \text{ is tossed if} & (X_{n-2}, X_{n-1}) = (T, T). \end{cases}$$

Clearly $\{X_n\}$ satisfies (2.103) with $r = 2$, so $\{Y_n\} \equiv \{(X_{n-1}, X_n)\}$ is a Markov chain with tpm

$$P = \begin{array}{c} \\ HH \\ HT \\ TH \\ TT \end{array} \begin{array}{c} \begin{array}{cccc} HH & HT & TH & TT \end{array} \\ \begin{pmatrix} p_1 & q_1 & 0 & 0 \\ 0 & 0 & p_2 & q_2 \\ p_3 & q_3 & 0 & 0 \\ 0 & 0 & p_4 & q_4 \end{pmatrix} \end{array}. \tag{2.105}$$

(Compare this example to the note following (2.4).) □

Exercise 2.10. In (2.105), assume that $0 < p_i < 1$ for $i = 1, 2, 3, 4$.

(i) Show that P is regular; find $\nu(P)$, the limiting = stationary distribution $\pi \equiv (\pi_{HH}, \pi_{HT}, \pi_{TH}, \pi_{TT})$, and the mean return times $\mu_{HH,HH}$, $\mu_{HT,HT}$, $\mu_{TH,TH}$, $\mu_{TT,TT}$ for $\{Y_n\}$.

(ii) Find $\lim_{n\to\infty} \Pr[X_n = H]$. (The answer does not depend on the initial state.) $\square$

Example 2.12 *(Self-avoiding random walk in $\mathbf{Z}^d$).* For $d \geq 2$, consider the SSRW $\{\tilde{\mathbf{S}}_n\}$ in $\mathbf{Z}^d$ as defined in (1.36); assume that $\tilde{\mathbf{S}}_0 = 0$. Let N denote the first time $n \geq 2$ that $\tilde{\mathbf{S}}_n$ repeats a state already visited and set

$$X_n = \begin{cases} \tilde{\mathbf{S}}_n & \text{for } n \leq N, \\ \tilde{\mathbf{S}}_N & \text{for } n > N. \end{cases}$$

The distribution of X_n does not satisfy (2.103) for any integer r [verify!]; instead it depends on the *entire* history $X_0, \ldots, X_{n-1}$. Thus $\{X_n\}$ is not Markovian and cannot be converted into a Markov chain by grouping. $\square$

3. Classification of the States of a Markov Chain

Let $\{X_n \mid n = 0, 1, 2, \ldots\}$ be a discrete Markov chain with finite or countable state space $\mathcal{S}$ and tpm $P \equiv \{p_{ij}\}$. As already seen, the n-step tpm $P^{(n)} \equiv \{p_{ij}^{(n)}\}$ is simply P^n. By convention we set $P^{(0)} = I$.

3.1. Equivalence classes; periodicity, recurrence, transience

State j is *accessible* from state i, indicated by $i \to j$, if $p_{ij}^{(n)} > 0$ for some $n \geq 0$. If $i \neq j$, this occurs iff there exists a *path* from i to j (also indicated by $i \to j$), that is, a sequence of states $i \equiv k_0$, k_1, $\ldots$, k_{n-1}, $k_n \equiv j$ such that $p_{k_{l-1},k_l} > 0$ for $l = 1, \ldots, n$ [why?].

If $i \to j$ and $j \to i$ we say that i and j *communicate* and write $i \leftrightarrow j$. Clearly "$\leftrightarrow$" is an equivalence relation (reflexive, symmetric, and transitive), hence partitions $\mathcal{S}$ into a set of *equivalence classes.*

For any pair $i, j \in \mathcal{S}$, $i \leftrightarrow j$ iff i and j belong to the same equivalence class. Thus, if i and j belong to different equivalence classes, then possibly $i \to j$ or $j \to i$, but not both.

If all states communicate then there is only one equivalence class, $\mathcal{S}$ itself. Here the Markov chain $\{X_n\}$ and its tpm P are called *irreducible.* Any (finite) regular chain is irreducible. The converse is false (e.g. Example 2.10.) but is true if the irreducible chain is aperiodic (see Proposition 3.6(ii)).

The *period* $d(i)$ of state i is the greatest common divisor of all $n \geq 1$ such that $p_{ii}^{(n)} > 0$. (If no such n exists, set $d(i) = 0$.) If $d(i) = 1$ then i is called *aperiodic.* Clearly any state i with $p_{ii} > 0$ is aperiodic.

Proposition 3.1. *Periodicity is a class property: if $i \leftrightarrow j$ then $d(i) = d(j)$.*

Proof. (Shiryaev, *Probability* (1984), pp.530-1.) Since $i \to j$ and $j \to i$, there are integers $k \geq 1$ and $l \geq 1$ such that $p_{ij}^{(k)} > 0$ and $p_{ji}^{(l)} > 0$. Thus $p_{ii}^{(k+l)} \geq p_{ij}^{(k)} p_{ji}^{(l)} > 0$, so $k + l$ is divisible by $d(i)$. Consider any $n \geq 1$ that is not divisible by $d(i)$. Then $k + n + l$ is also not divisible by $d(i)$, so $p_{ii}^{(k+n+l)} = 0$. But $p_{ii}^{(k+n+l)} \geq p_{ij}^{(k)} p_{jj}^{(n)} p_{ji}^{(l)}$, so $p_{jj}^{(n)} = 0$. So if $p_{jj}^{(n)} > 0$ then n is divisible by $d(i)$, hence $d(i) \leq d(j)$. Similarly $d(j) \leq d(i)$, so $d(i) = d(j)$.$\square$

Thus the period $d(\mathcal{C})$ is well-defined for any equivalence class $\mathcal{C}$. If $d(\mathcal{C}) = 1$ the class $\mathcal{C}$ is called *aperiodic.*

Example 3.1. Consider a tpm of the form

$$P \equiv \begin{pmatrix} 0 & + & + \\ + & 0 & + \\ + & + & 0 \end{pmatrix}.$$

The three states communicate so the chain is irreducible, and eac h state i has period 1 [verify] so the chain is aperiodic, even though each $p_{ii} = 0$. □

Definition 3.1. The state i is *recurrent* if, starting in state i, the chain returns to i with probability 1; otherwise i is called *transient.* □

The probability that the *first* visit to j occurs at time n is denoted by

$$f_{ij}^{(n)} = \Pr_i[X_n = j \text{ but } X_r \neq j \text{ for } 1 \leq r \leq n-1]. \tag{3.1}$$

The probability that the chain visits j at least once starting from i is

$$f_{ij} := \sum\nolimits_{n=1}^{\infty} f_{ij}^{(n)}. \tag{3.2}$$

Thus the state i is recurrent (transient) iff $f_{ii} = 1$ ($f_{ii} < 1$).

Note that $f_{ii}^{(1)} = p_{ii}$; by convention we set $f_{ii}^{(0)} = 0$. The recursion

$$p_{ij}^{(n)} = \sum\nolimits_{k=1}^{n} f_{ij}^{(k)} p_{jj}^{(n-k)} \qquad \text{[verify]} \tag{3.3}$$

obtained from the Law of Total Probability, will be useful. Define

$$Z_j = \sum\nolimits_{n=1}^{\infty} 1_j(X_n), \tag{3.4}$$

$$N_j = \min\nolimits_{n \geq 1}\{X_n \mid X_n = j\}, \tag{3.5}$$

respectively the total number of visits to state j and the first time $n \geq 1$ that the chain visits state j. From (3.1) and (3.2),

$$\begin{aligned} \Pr_i[N_j = n] &= f_{ij}^{(n)}, \\ \Pr_i[N_j < \infty] &= f_{ij} \quad . \end{aligned} \tag{3.6}$$

Note that

$$\Pr_i[Z_j \geq 1] = \Pr_i[N_j < \infty] = f_{ij}, \tag{3.7}$$

$$\Pr_i[\text{state } j \text{ occurs i.o.}] = \Pr_i[Z_j = \infty] = \lim_{k\to\infty} \Pr_i[Z_j \geq k], \tag{3.8}$$

$$\mathrm{E}_i(Z_j) = \mathrm{E}_i\left[\sum\nolimits_{n=1}^{\infty} 1_j(X_n)\right] = \sum\nolimits_{n=1}^{\infty} p_{ij}^{(n)}. \tag{3.9}$$

Proposition 3.2. *$Z_i \sim$ Geometric$(p = 1 - f_{ii})$: For $k = 0, 1, 2, \ldots$,*

$$\Pr_i[Z_i \geq k] = (f_{ii})^k. \tag{3.10}$$

Proof. Clearly (3.10) holds for $k = 0$ trivially and for $k = 1$ by (3.7), so we need consider $k \geq 2$ only.

Non-rigorous argument: By the Markov property, the chain renews itself each time it returns to state i, so

(3.11)

$$\Pr_i[\text{at least } k \text{ returns to } i] = (\Pr_i[\text{at least one return to } i])^k = (f_{ii})^k.$$

Criticism: The Markov property states that for each *fixed* time n, the future $X_{n+1}, X_{n+2} \ldots$ depends only on the present value of X_n, not on the past values. However, the above argument for (3.11) rests on the assertion that $X_{N_i+1}, X_{N_i+2} \ldots$ depends only on the present value of X_{N_i}, where N_i is the *random* time of the first return to i (and similarly for the second return and all succeeding returns to i). Thus, for a rigorous proof we would need to show that the Markov property extends to the *Strong Markov Property (SMP)*, where the "present" is represented by X_N where N is any *stopping time* as defined before Theorem 1.2 (Wald's Lemma). This is in fact true – the following argument illustrates how the SMP can be proved.

Rigorous argument for (3.10):

$$\begin{aligned}
&\Pr_i[Z_i \geq k] \\
&= \sum_{n=1}^{\infty} \Pr_i[Z_i \geq k,\, N_i = n] + \overbrace{\Pr_i[Z_i \geq k,\, N_i = \infty]}^{=0} \\
&= \sum_{n=1}^{\infty} \Pr_i[\{X_{n+r}\}_{r=1}^{\infty} \text{ returns to } i \text{ at least } k-1 \text{ times},\, N_i = n]
\end{aligned}$$

$$= \sum_{n=1}^{\infty} \Pr_i[\{X_{n+r}\}_{r=1}^{\infty} \text{ returns to } i \text{ at least } k-1 \text{ times} | N_i = n] \Pr_i[N_i = n]$$

$$\overset{*}{=} \sum_{n=1}^{\infty} \Pr_i[\{X_{n+r}\}_{r=1}^{\infty} \text{ returns to } i \text{ at least } k-1 \text{ times} | X_n = i] \Pr_i[N_i = n]$$

$$\overset{**}{=} f_{ii}^{k-1} \sum_{n=1}^{\infty} \Pr_i[N_i = n] \overset{***}{=} f_{ii}^{k}.$$

Because $\{N_i = n\} = \{X_1 \neq i, \ldots, X_{n-1} \neq i, X_n = i\}$, * follows from the Markov property, ** follows by induction, and *** follows from (3.6). □

Proposition 3.3. *(i) State i is*

$$\left\{ \begin{matrix} \text{recurrent} \\ \text{transient} \end{matrix} \right\} \overset{*}{\Longleftrightarrow} \Pr_i[\text{state } i \text{ recurs i.o.}] = \left\{ \begin{matrix} 1 \\ 0 \end{matrix} \right\} \overset{**}{\Longleftrightarrow} \sum_{n=1}^{\infty} p_{ii}^{(n)} \left\{ \begin{matrix} = \infty \\ < \infty \end{matrix} \right\}.$$

(ii) If j is transient then $\sum_{n=1}^{\infty} p_{ij}^{(n)} < \infty$ for all i. Thus $\lim_{n\to\infty} p_{ij}^{(n)} = 0$.

Proof. (i) * is immediate from the definition of transience:

$$\begin{aligned} \Pr_i[\text{state } i \text{ recurs i.o.}] &\overset{(3.8)}{=} \lim_{k\to\infty} \Pr_i[Z_i \geq k] \\ &\overset{(3.10)}{=} \lim_{k\to\infty} f_{ii}^{k} \quad = \left\{ \begin{matrix} 1 & \text{if } f_{ii} = 1, \\ 0 & \text{if } f_{ii} < 1 \end{matrix} \right\}, \end{aligned}$$

while ** follows from (3.9) and (3.10):

$$\begin{aligned} \sum_{n=1}^{\infty} p_{ii}^{(n)} \overset{(3.9)}{=} \mathrm{E}_i(Z_i) &= \sum_{k=1}^{\infty} \Pr_i[Z_i \geq k] \\ &\overset{(3.10)}{=} \sum_{k=1}^{\infty} f_{ii}^{k} \quad \left\{ \begin{matrix} = \infty & \text{if } f_{ii} = 1, \\ < \infty & \text{if } f_{ii} < 1 \end{matrix} \right\}. \end{aligned} \tag{3.12}$$

□

(ii) Assume that j is transient. Then from (3.3),

$$\begin{aligned} \sum_{n=1}^{\infty} p_{ij}^{(n)} &= \sum_{n=1}^{\infty} \left(\sum_{k=1}^{n} f_{ij}^{(k)} p_{jj}^{(n-k)} \right) \\ &= \sum_{k=1}^{\infty} \left(\sum_{n=k}^{\infty} f_{ij}^{(k)} p_{jj}^{(n-k)} \right) \\ &= \left(\sum_{k=1}^{\infty} f_{ij}^{(k)} \right) \left(\sum_{n=0}^{\infty} p_{jj}^{(n)} \right) < \infty, \end{aligned}$$

since $\sum_{k=1}^{\infty} f_{ij}^{(k)} = f_{ij} \leq 1$ by (3.2) and $\sum_{n=0}^{\infty} p_{jj}^{(n)} < \infty$ by (i). □

Remark 3.1. (i) If i is transient ($f_{ii} < 1$), then by (3.10) and (3.12),

$$Z_i \sim \text{ Geometric}\,(1 - f_{ii}), \tag{3.13}$$

$$\mathrm{E}_i(Z_i) = \frac{f_{ii}}{1 - f_{ii}} < \infty. \tag{3.14}$$

(ii) Since $\Pr_i[X_n = i] = p_{ii}^{(n)}$, the implication $\Pr_i[i \text{ i.o.}] \overset{**}{\Longleftarrow} \sum_{n=1}^{\infty} p_{ii}^{(n)} < \infty$ also follows from the Borel-Cantelli Lemma 1.1. □

Proposition 3.4. *Recurrence (transience) is a class property: if $i \leftrightarrow j$ and i is recurrent (transient) then j is recurrent (transient).*

Proof. For integers $k \geq 1$ and $l \geq 1$ such that $p_{ij}^{(k)} > 0$ and $p_{ji}^{(l)} > 0$,

$$\sum_{n=1}^{\infty} p_{ii}^{(n)} \geq \sum_{n=1}^{\infty} p_{ii}^{(k+n+l)} \geq p_{ij}^{(k)} \left(\sum_{n=1}^{\infty} p_{jj}^{(n)} \right) p_{ji}^{(l)}. \tag{3.15}$$

A similar inequality holds if i and j are interchanged, hence

$$\sum_{n=1}^{\infty} p_{ii}^{(n)} < \infty \iff \sum_{n=1}^{\infty} p_{jj}^{(n)} < \infty.$$

The result (i) now follows from Proposition 3.3. □

Proposition 3.5. *A Markov chain can never leave a recurrent class.*

Proof. Suppose $i \in \mathcal{C}$, $j \notin \mathcal{C}$, and $i \to j$, where $\mathcal{C}$ is a recurrent class. Since $\Pr_i[i \text{ recurs i.o.}] = 1$, necessarily $j \to i$, so $i \leftrightarrow j$, a contradiction. □

Proposition 3.6. *(i) An irreducible Markov chain with finite $\mathcal{S}$ is recurrent. (ii) An irreducible chain with finite $\mathcal{S}$ is regular iff it is aperiodic. Aperiodicity holds if $p_{jj} > 0$ for some j (but not only if: see Example 3.1).*

Proof. (i) Because there are only finitely many states, at least one state must recur infinitely often with positive probability, hence by Proposition 3.3 must be recurrent. Thus by Proposition 3.4 all states are recurrent.

(ii) *(if):* We'll prove this under the assumption, stronger than aperiodicity, that there exists a state j for which $p_{jj} > 0$.[3] For this j consider all triples $(i, j, k) \in \mathcal{S} \times \{j\} \times \mathcal{S}$ (possibly $i = j$, $j = k$, and/or $i = j$). Since any pair of states communicate, there exist paths $i \to j$ and $j \to k$, hence there exists a path $i \xrightarrow{j} k$ that passes through j. Denote its length by n_{ik} and set $\nu = \max_{i,k} n_{ik}$. Because $p_{jj} > 0$, if $n_{ik} < \nu$ we can extend the path $i \xrightarrow{j} k$ to a path of length ν by adding a loop $j \to j$ of length $\nu - n_{ik}$. Thus the tpm P satisfies $P^\nu > 0$, hence the chain is regular.

(only if): If $P^\nu > 0$ then $P^{\nu+n} > 0$ for all $n \geq 1$, so P is aperiodic. □

3.2. Examples: equivalence, periodicity, recurrence/transience

Example 2.1d *(the two-state Markov chain, continued).*
Consider the four specific cases P_1, P_2, P_3, P_4 of the general two-state tpm P in (2.4) with state space $\mathcal{S} = \{1, 2\}$:

$$P_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad P_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$
$$P_3 = \begin{pmatrix} 1-a & a \\ 0 & 1 \end{pmatrix}, \qquad P_4 = \begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix},$$

where $0 < a, b < 1$. Then [verify]:

- P_1 has two classes $\{1\}$ and $\{2\}$; each is aperiodic.
- P_2 has one class $\{1, 2\}$; it is irreducible, hence recurrent by Proposition 3.6(i), and has period 2.
- P_3 has two classes: $\{1\}$ is transient by Proposition 3.5, $\{2\}$ is absorbing hence trivially recurrent and aperiodic.
- P_4 has one class $\{1, 2\}$; P_4 is regular, hence irreducible and aperiodic.

Example 2.8b *(Simple random walk on a circle, continued).*
The state space is $\mathcal{S} = \{0, \ldots, s-1\}$ and the tpm P is given in (2.95), p.58.

- If either all $p_i > 0$ or all $q_i > 0$ (or both), then all states communicate, so P is irreducible, hence recurrent by Proposition 3.6(i).

[3] For the general case see Shiryaev, *Probability* (1984), pp.539-40.

• If, in addition, at least one $r_i > 0$ then this chain has period 1, i.e., is aperiodic, hence regular by Proposition 3.6(ii).

• If all $p_i > 0$, all $q_i > 0$, and all $r_i = 0$, this chain has period 2 if s is even (Example 2.10) but period 1 (aperiodic) if s is odd (Example 3.1, p.64). □

Example 2.11b. The tpm P in (2.105), p.61, is irreducible and aperiodic if $0 < p_i < 1$ for $i = 1, 2, 3, 4$, hence regular by Proposition 3.6(ii). □

Example 3.2a *(Simple random walk (SRW) on* $\mathbf{Z}^1$*).* Here the state space $\mathcal{S} \equiv \{0, \pm 1, \pm 2, \ldots\}$ is the *countable* set of all integers. Starting at $S_0 = 0$, the chain $\{S_n\}$ takes one step to the right (left) with probability p (q), where $p + q = 1$ and $0 < p,\ q < 1$. All states communicate [verify], so the chain is irreducible, and each state has period 2.

To determine the recurrence/transience of this chain we can apply Proposition 3.3 for any state i: state i is recurrent (transient) iff

$$\sum\nolimits_{n=1}^{\infty} p_{ii}^{(n)} = \infty \quad (< \infty). \tag{3.16}$$

But $p_{ii}^{(n)} = 0$ for odd n, while by the binomial distribution and Stirling's approximation,

$$\begin{aligned} p_{ii}^{(2n)} &= \Pr[n \text{ steps right, } n \text{ steps left}] \\ &= \binom{2n}{n} p^n q^n = \frac{(2n)!}{(n!)^2} p^n q^n \\ &\sim \frac{(2n)^{2n} e^{-2n} \sqrt{2\pi 2n}}{(n^n e^{-n} \sqrt{2\pi n})^2} p^n q^n = \frac{(4pq)^n}{\sqrt{\pi n}}. \end{aligned} \tag{3.17}$$

Since $4pq = 1$ (< 1) iff $p = 1/2$ ($\neq 1/2$), it follows from (3.16) and (3.17) that *the chain is recurrent (transient) iff* $p = 1/2$ ($\neq 1/2$).

Note 1: Recall that in Proposition 1.4(i) we established the recurrence when $p = 1/2$ (SSRW) by applying Wald's Lemma.

Note 2: Because $S_n = Z_1 + \cdots + Z_n$ where each $Z_i = \pm 1$ with probabilities p and q, the transience of SRW when $p \neq 1/2$ also follows from the SLLN: $S_n/n \to \mathrm{E}(Z_i) = p - q = 2p - 1$, so $S_n \to \infty$ ($-\infty$) with probability 1 when $p > 1/2$ ($p < 1/2$). □

Example 3.3 *(Simple symmetric random walk (SSRW) on $\mathbf{Z}^2$).* We now apply Proposition 3.3 to show that the "pure" SSRW $\{\tilde{\mathbf{S}}_n\}$ defined in (1.36) remains recurrent in $\mathbf{Z}^2$. (By Proposition 1.7 (Polya's Theorem) it is transient in $\mathbf{Z}^d$ for $d \geq 3$.)

This chain is irreducible and periodic with period 2 [verify]. Assume that $\tilde{\mathbf{S}}_0 = \mathbf{0}$. Again $\mathrm{Pr}_{\mathbf{0}}[\tilde{\mathbf{S}}_n = \mathbf{0}] = 0$ for odd n, while by (1.38) with $d = 2$,

$$\begin{aligned}
\mathrm{Pr}_{\mathbf{0}}[\tilde{\mathbf{S}}_{2n} = \mathbf{0}] &= \sum_{k_1+k_2=n} \frac{(2n)!}{(k_1!)^2(k_2!)^2}\frac{1}{4^{2n}} && [\mathrm{Multinomial}_4(2n; \tfrac{1}{4}, \tfrac{1}{4}, \tfrac{1}{4}, \tfrac{1}{4})] \\
&= \frac{(2n)!}{(n!)^2 4^{2n}} \sum_{k_1+k_2=n} \left(\frac{n!}{k_1!k_2!}\right)^2 \\
&= \binom{2n}{n}\frac{1}{4^{2n}} \sum_{k_1+k_2=n} \binom{n}{k_1}\binom{n}{k_2} \\
&= \binom{2n}{n}^2 \frac{1}{4^{2n}} && [\text{the hypergeometric formula}] \\
&= \left[\frac{(2n)!}{(n!)^2 4^n}\right]^2 \\
&\sim \left[\frac{(2n)^{2n}e^{-2n}\sqrt{2\pi 2n}}{(n^n e^{-n}\sqrt{2\pi n})^2 4^n}\right]^2 && [\text{Stirling's approximation}] \\
&= \frac{1}{\pi n}. && (3.18)
\end{aligned}$$

Thus $\sum_{n=0}^{\infty} \mathrm{Pr}_{\mathbf{0}}[\tilde{\mathbf{S}}_{2n} = \mathbf{0}] = \infty$ so the chain is recurrent (Proposition 3.3).□

Example 3.4 *(Gambler's Ruin with an infinitely rich opponent).* As in Example 2.5a, at each time $n = 1, 2, \ldots$, a gambler wins \$1 with probability p and loses \$1 with pr. $q = 1 - p$, $0 < p, q < 1$; X_n denotes his total capital after the nth game. If his opponent's capital is unlimited, the state space is $\{0, 1, 2, \ldots\}$, which is countable, and the tpm is

$$P = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ \vdots \end{array} \begin{array}{c} \begin{array}{cccccc} 0 & 1 & 2 & 3 & 4 & \cdots \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & \cdots \\ q & 0 & p & 0 & 0 & \cdots \\ 0 & q & 0 & p & 0 & \cdots \\ 0 & 0 & q & 0 & p & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{pmatrix} \end{array}. \tag{3.19}$$

There are two equivalence classes: $\mathcal{C}_0 \equiv \{0\}$ is recurrent since it is absorbing, while $\mathcal{C}_1 \equiv \{1, 2, \ldots\}$ cannot be recurrent [why?] hence is transient.

It follows from Example 3.2a, p.69 that if $p \leq 1/2$ then with probability 1 the gambler is eventually ruined[4] from any starting point $i \geq 1$, i.e.,

$$\Pr_i[X_n = 0 \text{ for some } n] = 1. \tag{3.20}$$

What if $p > 1/2$? In Example 2.5a eqn. (2.66) we found that when the opponent's capital is $m \geq 1$, the probability of the gambler's eventual ruin ($X_n = 0$) given $X_0 = i$ ($1 \leq i \leq m-1$) is $1 - u_i$, where

$$u_i = \frac{1 + \delta + \cdots + \delta^{i-1}}{1 + \delta + \cdots + \delta^{m-1}} \tag{2.65$'$}$$

$$= \frac{1 - \delta^i}{1 - \delta^m} \tag{2.66$'$}$$

with $\delta = q/p < 1$. Since $u_i \to 1 - \delta^i$ as $m \to \infty$, the probability of ruin is

$$\Pr_i[X_n = 0 \text{ for some } n] = \delta^i < 1 \tag{3.21}$$

so

$$\Pr_i[X_n \geq 1 \text{ for all } n] = 1 - \delta^i > 0. \tag{3.22}$$

Thus, when $p > 1/2$, with probability $1 - \delta^i$ the game never terminates.

What happens in this case? Because $\mathcal{C}_1$ is transient X_n cannot remain bounded as $n \to \infty$, i.e. $\limsup_{n\to\infty} X_n = \infty$. There are two possibilities:

[4] Also recall Note 2 on p.69: the SLLN applies when $p < 1/2$.

A: $\liminf_{n\to\infty} X_n = \infty$, so $\lim_{n\to\infty} X_n = \infty$;

B: $\liminf_{n\to\infty} X_n < \infty$, so X_n exhibits increasingly large oscillations.

Conditional on the event $\{X_n \geq 1 \text{ for all } n\}$, does A or B occur?

Exercise 3.1. For any initial state $i \geq 1$, show that when $p > 1/2$,

$$\Pr_i[\mathrm{C} \mid X_n \geq 1 \text{ for all } n] = 1,$$

and determine whether C = A or C = B. □

Exercise 3.2 *(Simple random walk on $\mathbf{Z}^1$ with reflecting boundary at 0).* Change the game in Example 3.4 so that 0 is reflecting rather than absorbing, that is, the tpm is changed from (3.19) to

$$P^* = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ \vdots \end{array} \begin{array}{c} \begin{array}{cccccc} 0 & 1 & 2 & 3 & 4 & \cdots \end{array} \\ \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & \cdots \\ q & 0 & p & 0 & 0 & \cdots \\ 0 & q & 0 & p & 0 & \cdots \\ 0 & 0 & q & 0 & p & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{pmatrix} \end{array}. \tag{3.23}$$

This chain is irreducible with period 2. For which values of p is the chain recurrent? transient? Determine the values of $\liminf_{n\to\infty} X_n$ and $\limsup_{n\to\infty} X_n$ in each case. (Despite irreducibility, transience is possible because the state space $\{0, 1, 2, \ldots\}$ is infinite.) □

Hint: The following lemma may helpful.

Lemma 3.1. *If i is recurrent and $i \to j$ then $f_{ij} = 1$.*

Proof. By Proposition 3.5, $i \leftrightarrow j$, hence $p_{ji}^{(k)} > 0$ for some $k \geq 1$. Because j is also recurrent, $\Pr_j[X_n = j \text{ i.o.}] = 1$, hence

$$\begin{aligned} p_{ji}^{(k)} &= \Pr_j[X_k = i,\ X_n = j \text{ for some } n \geq k+1] \\ &= \sum\nolimits_{n\geq k+1} \Pr_j[X_k = i,\ \ X_{k+1} \neq j, \ldots, X_{n-1} \neq j,\ X_n = j] \\ &= \sum\nolimits_{n\geq k+1} p_{ji}^{(k)} f_{ij}^{(n-k)} \\ &= p_{ji}^{(k)} \sum\nolimits_{n-k\geq 1} f_{ij}^{(n-k)} \\ &= p_{ji}^{(k)} f_{ij}. \end{aligned} \tag{3.26}$$

Therefore $f_{ij} = 1$ as asserted. □

Example 2.3b *(several transient and several absorbing states, continued).* Recall the finite tpm

$$(2.24)' \quad P = \begin{pmatrix} Q & R \\ 0 & I \end{pmatrix}, \quad Q : r \times r, \quad R : r \times (s-r), \quad I : (s-r) \times (s-r).$$

In this example we stated that if *each* row $\mathbf{r}_i$ of R is nonzero then states $1, \ldots, r$ are transient. In the following proposition this assumption is weakened as follows: states $1, \ldots, r$ form a single class and *at least one* row of R is nonzero, i.e, $R \neq 0$.

Proposition 3.7. *Assume that P has the form (2.24)′ and that $\mathcal{C} \equiv \{1, \ldots, r\}$ constitutes a single class. If $R \neq 0$ then $\mathcal{C}$ is a transient class. Furthermore $(I - Q)^{-1}$ exists, hence (2.30), (2.35), and (2.44) are valid.*

Proof. Since $R \neq 0$ the chain can leave $\mathcal{C}$, hence $\mathcal{C}$ cannot be recurrent by Proposition 3.5, thus must be transient.

Next, by Lemma 2.1, to show that $(I - Q)^{-1}$ exists it suffices to show that $\sum_{n=1}^{\infty} Q^n$ is convergent. For $1 \leq i, j \leq r$, it follows from (2.25) that

$$(Q^n)_{ij} = (P^n)_{ij} \equiv p_{ij}^{(n)},$$

so

$$\left(\sum\nolimits_{n=1}^{\infty} Q^n\right)_{ij} = \sum\nolimits_{n=1}^{\infty} p_{ij}^{(n)} < \infty$$

by Proposition 3.3(ii), hence $\sum_{n=1}^{\infty} Q^n$ is convergent. □

Remark 3.2. Recall (3.4) and (3.5) for the definitions of Z_j and N_j. Then

$$\begin{aligned} \sum\nolimits_{n=1}^{\infty} p_{ij}^{(n)} &= \mathrm{E}_i(Z_j) \\ &= \mathrm{E}_i(Z_j | N_j < \infty) \Pr_i[N_j < \infty] + \mathrm{E}_i(Z_j | N_j = \infty) \Pr_i[N_j = \infty] \\ &\overset{*}{=} \mathrm{E}_i[1 + \mathrm{E}_j(Z_j)] \Pr_i[N_j < \infty] + 0 \cdot \Pr_i[N_j = \infty] \\ &= [1 + \mathrm{E}_j(Z_j)] \Pr_i[N_j < \infty]. \end{aligned} \tag{3.29}$$

(Relation * requires the Strong Markov Property – see the "rigorous argument" for (3.10).) By (3.14), (3.29) can be rewritten as

$$\mathrm{E}_i(Z_j) = [1 + \mathrm{E}_j(Z_j)] f_{ij} = \frac{f_{ij}}{1 - f_{jj}}, \tag{3.30}$$

where (see (3.6) and (3.24)), $f_{jj} \equiv \Pr_j[N_j < \infty] < 1$ and $f_{ij} \equiv \Pr_i[N_j < \infty]$ are the probabilities that the chain visits j at least once starting from j (resp., from i). (When $i = j$, (3.30) reduces to (3.14).) □

Example 2.5b *(Gambler's Ruin, continued).* The tpm is (2.62)′

$$P = \begin{array}{c} \\ 1 \\ 2 \\ \vdots \\ m-1 \\ m \\ 0 \end{array} \begin{array}{c} \begin{array}{cccccccc} 1 & 2 & 3 & \cdots & m-2 & m-1 & m & 0 \end{array} \\ \begin{pmatrix} 0 & p & 0 & \cdots & 0 & 0 & 0 & q \\ q & 0 & p & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & q & 0 & p & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 1 \end{pmatrix} \end{array} \equiv \begin{pmatrix} Q & R \\ 0 & I \end{pmatrix}.$$

Since $0 < p, q < 1$, $\{1, \ldots, m-1\}$ forms a single class [verify] and $R \neq 0$, hence by Proposition 3.7 this class is transient and $(I - Q)^{-1}$ exists. □

3.3. Countable chains: limiting and stationary distributions, positive recurrent and null recurrent classes

For the remainer of Section 3 we assume that the state space $\mathcal{S}$ of the Markov chain $\{X_n\}$ is *countably infinite*, say $\mathcal{S} = \{1, 2, \ldots\}$. Here the existence and determinination of limiting distributions for $\{X_n\}$ requires extensions of the definitions and results in §2.2 as follows.

Definition 3.1. Let $\pi \equiv (\pi_1, \pi_2, \ldots)$ be a countable row vector and let $\mathbf{e} \equiv (1, 1 \ldots,)'$ denote the countable column vector of 1's. Then π is a *limiting array* for $\{X_n\}$ if for all $i, j \in \mathcal{S}$,

$$\exists \lim_{n\to\infty} p_{ij}^{(n)} = \pi_j. \tag{3.31}$$

Clearly $0 \leq \pi_j \leq 1$. π is a *limiting (sub)distribution* for $\{X_n\}$ if in addition to (3.31), $(\sum_j \pi_j \leq 1)$ $\sum_j \pi_j = 1$. □

Note that (3.31) requires that π_j is the limit regardless of the initial state i. It is possible that $\lim_{n\to\infty} p_{ij}^{(n)}$ exists for all i, j yet no limiting array exists according to Definition 3.1: take $P = I$.

If j is transient then $\pi_j = 0$ by Proposition 3.3(ii).

Remark 3.3. Definition 3.1 should be compared to Definition 2.1 for finite state spaces. In particular, we do not write $\lim_{n\to\infty} P^n = \mathbf{e}\,\pi$ in (3.31), for this would require that the convergence in (3.31) be uniform in i, j.

*Questions**: when is the convergence (3.31) uniform in i? in j? in i, j?*□

Proposition 3.8. *If a limiting array π exists, then:*

(i) π is a subdistribution, i.e., $\sum_j \pi_j \le 1$, and π is stationary, i.e., $\pi P = \pi$.

(ii) Either: $\sum_j \pi_j = 1$, in which case π is the unique limiting distribution, in fact π is the unique stationary distribution: $\pi P = \pi$;

Or: $\sum_j \pi_j = 0$, i.e., all $\pi_j = 0$, and no stationary distribution exists.

(iii) If $\pi_i > 0$ and $i \to j$, then $\pi_j > 0$. Thus positivity of π_i is a class property: if $i \leftrightarrow j$ then $\pi_i > 0 \iff \pi_j > 0$.

Proof. (i) For each integer $r \ge 1$ and each i,

$$\sum_{j=1}^{r} \pi_j = \sum_{j=1}^{r} \lim_{n\to\infty} p_{ij}^{(n)} = \lim_{n\to\infty} \sum_{j=1}^{r} p_{ij}^{(n)} \le 1.$$

Now let $r \to \infty$ to obtain $\sum_j \pi_j \le 1$. [Or just apply Fatou's Lemma.]

Next, for any $j, k \ge 1$,

$$\begin{aligned}\sum_{i=1}^{r} \pi_i p_{ij} &= \sum_{i=1}^{r} \lim_{n\to\infty} p_{ki}^{(n)}\, p_{ij} = \lim_{n\to\infty} \sum_{i=1}^{r} p_{ki}^{(n)}\, p_{ij} \le \lim_{n\to\infty} \sum_{i=1}^{\infty} p_{ki}^{(n)}\, p_{ij} \\ &= \lim_{n\to\infty} p_{kj}^{(n+1)} = \pi_j,\end{aligned}$$

so $\sum_i \pi_i p_{ij} \le \pi_j$. If strict inequality holds, that is if $\sum_i \pi_i p_{ij} < \pi_j$ for some j, then

$$\sum_j \pi_j > \sum_j \sum_i \pi_i p_{ij} = \sum_i \pi_i \sum_j p_{ij} = \sum_i \pi_i,$$

a contradiction. Thus $\sum_i \pi_i p_{ij} = \pi_j$ for all j, i.e., $\pi P = \pi$.

(ii) By (i), $\pi = \pi P = \cdots = \pi P^n$. [We can interchange the order of infinite summations when all summands are nonnegative.] Therefore for each $r \geq 1$,

$$\pi_j = \lim_{n\to\infty} \sum_i \pi_i p_{ij}^{(n)} = \lim_{n\to\infty} \sum_{i=1}^{r} \pi_i p_{ij}^{(n)} + \lim_{n\to\infty} \sum_{i=r+1}^{\infty} \pi_i p_{ij}^{(n)}$$

$$\equiv \Big(\sum_{i=1}^{r} \pi_i\Big)\pi_j + \epsilon_{r,j,n}. \tag{3.32}$$

But $\epsilon_{r,j,n} \leq \sum_{i=r+1}^{\infty} \pi_i \to 0$ as $r \to \infty$, so $\pi_j = \Big(\sum_i \pi_i\Big)\pi_j$. [Or just apply the Bounded Convergence Theorem.] Therefore for every j,

$$\pi_j\Big(1 - \sum\nolimits_i \pi_i\Big) = 0, \tag{3.33}$$

so either $\sum_i \pi_i = 1$ or else all $\pi_j = 0$, in which case $\sum_j \pi_j = 0$.

In the first case, π is a both a limiting distribution and a stationary distribution. To show that such π is unique, suppose that $\lambda \equiv (\lambda_1, \lambda_2, \ldots)$ is another stationary distribution, i.e., all $\lambda_j \geq 0$, $\sum_j \lambda_j = 1$, $\lambda = \lambda P$. Then $\lambda = \lambda P^n$ so as in (3.32),

$$\lambda_j = \lim_{n\to\infty} \sum_i \lambda_i p_{ij}^{(n)} = \Big(\sum_i \lambda_i\Big)\pi_j = \pi_j, \tag{3.34}$$

hence $\lambda = \pi$. In the second case, if a stationary distribution λ exists then $\lambda_j = \pi_j$ for all j by (3.34), contradicting all $\pi_j = 0$.

(iii) Choose $k \geq 1$ such that $p_{ij}^{(k)} > 0$. Then

$$\pi_j = \lim_{n\to\infty} p_{ij}^{(n+k)} \geq \lim_{n\to\infty} p_{ii}^{(n)} p_{ij}^{(k)} = \pi_i p_{ij}^{(k)},$$

so $\pi_i > 0 \Rightarrow \pi_j > 0$. □

We are ready to address the principal question in the theory of Markov chains: when does a limiting distribution exist? The following deep result provides the answer: a limiting distribution exists if and only if a stationary distribution exists. This turns the question of determining probabilistic

limits into an algebraic question of solving an (infinite) system of linear equations. For simplicity we will restrict attention to the aperiodic case.

Theorem 3.1 *(Billingsley, "Probability and Measure", Theorems 8.6, 8.7). Suppose that the chain is irreducible and aperiodic.*

(i) A limiting distribution exists if and only if a stationary distribution $\pi \equiv (\pi_j)$ exists, that is,

$$\pi = \pi P, \quad \text{all } \pi_j \geq 0, \quad \text{and } \sum\nolimits_j \pi_j = 1.$$

Here π is the unique stationary distribution, the unique limiting distribution, and all $\pi_j > 0$.

(ii) If no stationary distribution exists then

$$\lim_{n\to\infty} p_{ij}^{(n)} = 0$$

for all pairs i, j, that is, $\pi \equiv (0, 0, \ldots)$ is the unique limiting array.

Proof. The proof uses the remarkable technique of "coupling" due to Doeblin. A pair of independent copies of the Markov chain are started from different states then followed until they first coincide. Once this occurs their further probabilistic evolutions are identical, by the Markov property, hence the limiting behavior of the original chain cannot depend on its initial state. This can be utilized to produce a unique limiting distribution. See Billingsley for details. □

If $\lim_{n\to\infty} p_{jj}^{(n)} \equiv \pi_j$ exists for a state j then it is inversely related to the mean return time $\mu_{jj} \equiv \mathrm{E}_j(N_j)$ to j as follows (compare to Proposition 2.2):

Proposition 3.9. *If $\lim_{n\to\infty} p_{jj}^{(n)} \equiv \pi_j$ exists then*

$$\pi_j = \begin{cases} \frac{1}{\mu_{jj}} > 0 & \text{if } \mu_{jj} < \infty, \\ 0 & \text{if } \mu_{jj} = \infty. \end{cases} \tag{3.35}$$

Proof. Recall that

$$\mu_{jj} = \mathrm{E}_j(N_j) = \sum_{n\geq 0} \mathrm{Pr}_j[N_j > n].$$

Next,

$$\begin{aligned}
1 &= \Pr_j[j \notin \{X_1, \ldots, X_n\}] + \Pr_j[j \in \{X_1, \ldots, X_n\}] \\
&= \Pr_j[N_j > n] + \sum_{k=1}^{n-1} \Pr_j[X_k = j, X_{k+1} \neq j, \ldots, X_n \neq j] + \Pr_j[X_n = j] \\
&= \Pr_j[N_j > n] + \sum_{k=1}^{n-1} \Pr_j[X_{k+1} \neq j, \ldots, X_n \neq j \mid X_k = j] \Pr_j[X_k = j] \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad + \Pr_j[X_n = j] \\
&= \sum_{k=0}^{n} \Pr_j[N_j > n-k]\, p_{jj}^{(k)} \\
(3.36) &= \sum_{l=0}^{n} \Pr_j[N_j > l]\, p_{jj}^{(n-l)}.
\end{aligned}$$

(The relations $p_{jj}^{(0)} = 1$ and $\Pr_j[N_j > 0] = 1$ are used for the first and last terms of the sum, respectively.) Thus for any fixed integer r,

$$1 \geq \sum\nolimits_{l=0}^{r} \Pr_j[N_j > l]\, p_{jj}^{(n-l)} \to \Big(\sum\nolimits_{l=0}^{r} \Pr_j[N_j > l]\Big)\pi_j \quad \text{as } n \to \infty.$$

Finally let $r \to \infty$ to obtain $1 \geq \mu_{jj}\pi_j$. Therefore $\mu_{jj} = \infty \Rightarrow \pi_j = 0$.

If $\mu_{jj} < \infty$ then

$$\begin{aligned}
\sum\nolimits_{l=0}^{n} \Pr_j[N_j > l]\, p_{jj}^{(n-l)} &= \sum\nolimits_{l=0}^{\infty} \Pr_j[N_j > l]\, p_{jj}^{(n-l)}\, I_{[0,n]}(l) \\
(3.37) \qquad &\leq \sum\nolimits_{l=0}^{\infty} \Pr_j[N_j > l] < \infty,
\end{aligned}$$

so the Dominated Convergence Theorem ($\equiv$ Weierstrass's M-test) can be applied in (3.36) (see *) as follows:

$$\begin{aligned}
1 &= \lim_{n\to\infty} \sum\nolimits_{l=0}^{n} \Pr_j[N_j > l]\, p_{jj}^{(n-l)} \\
&= \lim_{n\to\infty} \sum\nolimits_{l=0}^{\infty} \Pr_j[N_j > l]\, p_{jj}^{(n-l)}\, I_{[0,n]}(l) \\
&\overset{*}{=} \sum\nolimits_{l=0}^{n} \Pr_j[N_j > l] \lim_{n\to\infty} p_{jj}^{(n-l)} = \mu_{jj}\pi_j,
\end{aligned}$$

which yields the second part of (3.35). $\square$

Proposition 3.10. *Consider an irreducible, aperiodic Markov chain. By Theorem 3.1,* $\lim_{n\to\infty} p_{ij}^{(n)} \equiv \pi_j$ *exists for every pair of states* i, j*. Only two mutually exclusive cases can occur:*

(a) All $\pi_j > 0$*, all* $\mu_{jj} < \infty$*,* $\pi_j = 1/\mu_{jj}$*. The chain is called "positive recurrent" and* $\pi \equiv (\pi_j)$ *is the limiting distribution.*

(b) All $\pi_j = 0$*, all* $\mu_{jj} = \infty$*, and no limiting distribution exists. The chain is either transient or recurrent; if the latter, it is called "null" recurrent.*

Proof. This follows directly from Theorem 3.1 and Proposition 3.9. □

Example 3.2b *(Simple symmetric random walk (SSRW) on* $\mathbf{Z}^1$*).* We encountered the case $\mu_{jj} = \infty$ in a SSRW. In Section 1 and Example 3.2a we saw that SSRW $\{S_n\}$ on $\mathbf{Z}^1$ is irreducible and recurrent. By (1.26),

$$\begin{aligned}\mu_{00} = \mathrm{E}_0(N_0) &= \mathrm{E}_0[N_0 \mid S_1 = 1]\mathrm{Pr}_0[S_1 = 1] + \mathrm{E}_0[N_0 \mid S_1 = -1]\mathrm{Pr}_0[S_1 = -1] \\ &= \tfrac{1}{2}\mathrm{E}_1(N_0) + \tfrac{1}{2}\mathrm{E}_{-1}(N_0) = \infty;\end{aligned}$$

similarly, $\mu_{jj} = \infty$ for any starting point j. This suggests that the chain meanders back and forth over $\mathbf{Z}^1$ but at a very slow rate.

The chain $\{S_n\}$ has period 2, so to apply Theorem 3.1 to SSRW, consider the process at even times only. That is, consider the chain $\{X_n\} = \{S_{2n}\}$ with state space $\mathcal{S} = \{0, \pm 2, \pm 4, \ldots\}$. Then $\{X_n\}$ is irreducible, aperiodic, and recurrent. By the symmetry of the process, if a limiting distribution $\pi = (\pi_0, \pi_{\pm 2}, \pi_{\pm 4}, \ldots)$ were to exist on $\mathbf{Z}^1$ it would be uniform, that is, $\pi_0 = \pi_{\pm 2} = \pi_{\pm 4} = \cdots$, but this would contradict $\sum \pi_{2i} = 1$. Thus no limiting distribution exists, so $\lim_{n\to\infty} p_{2i,2j}^{(n)} = 0$ by Theorem 3.1(ii), hence $\pi_{2j} = 0$ for all j. By Proposition 3.9, $\mu_{2j,2j} = \infty$; this behavior is called *null recurrence.* □

Remark 3.4. The distinction between positive recurrence and null recurrence is relevant only for a countable state space, because if the state space is finite then an irreducible aperiodic chain is always positive recurrent. This follows from Proposition 3.6(ii), which implies that the chain is regular so the limiting distribution π exists with all $\pi_j > 0$. □

Example 3.5 *(Remaining lifetime).* The lifetime α of a component is a discrete random variable with probability distribution $\alpha_1, \alpha_2, \ldots$ on $1, 2, \ldots$.

If a component dies during time period n, at the end of that period it is replaced by a new, independent component. Let X_n denote the remaining lifetime of the component present during period n. Then $\{X_n\}$ is a Markov chain with tpm given by [verify]

$$P = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{ccccc} 0 & 1 & 2 & 3 & \cdots \end{array} \\ \begin{pmatrix} \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 & \cdots \\ 1 & 0 & 0 & 0 & \cdots \\ 0 & 1 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 0 & \cdots \\ 0 & 0 & 0 & 1 & \cdots \end{pmatrix} \end{array}. \tag{3.38}$$

Assume that all $\alpha_i > 0$ so this chain is irreducible and aperiodic [verify]. Clearly $\Pr_0[X_n = 0 \text{ i.o.}] = 1$ [verify], so the chain is recurrent. We shall show that a limiting distribution exists iff the mean lifetime $\mathrm{E}(\alpha) < \infty$, in which case the chain is positive recurrent; it is null recurrent if $\mathrm{E}(\alpha) = \infty$.

By Theorem 3.1, $\{X_n\}$ is positive recurrent iff a stationary distribution $\pi \equiv (\pi_0, \pi_1, \ldots)$ exists, i.e, $\pi = \pi P$ and $\sum_j \pi_j = 1$. The stationary equations are

$$\begin{aligned} \pi_0 &= \pi_0\alpha_1 + \pi_1 &\Rightarrow\quad \pi_1 &= \pi_0(1-\alpha_1) \\ \pi_1 &= \pi_0\alpha_2 + \pi_2 &\Rightarrow\quad \pi_2 &= \pi_0(1-\alpha_1-\alpha_2) \\ \pi_2 &= \pi_0\alpha_3 + \pi_3 &\Rightarrow\quad \pi_3 &= \pi_0(1-\alpha_1-\alpha_2-\alpha_3) \\ &\vdots \end{aligned}$$

so, since $\sum_{k\geq 1} \alpha_k = 1$, equivalently,

$$\begin{aligned} &\left(\pi_0 \sum\nolimits_{k\geq 1} \alpha_k = \pi_0,\right) \\ &\pi_0 \sum\nolimits_{k\geq 2} \alpha_k = \pi_1 \\ &\pi_0 \sum\nolimits_{k\geq 3} \alpha_k = \pi_2 \\ &\pi_0 \sum\nolimits_{k\geq 4} \alpha_k = \pi_3 \\ &\qquad\vdots \end{aligned} \tag{3.39}$$

Sum these equations to obtain

$$\pi_0 \, \mathrm{E}(\alpha) = \sum\nolimits_j \pi_j. \tag{3.40}$$

Thus, if a stationary distribution π exists then[5] $\pi_0 > 0$ and $\pi_0 \mathrm{E}(\alpha) = 1$, hence $\mathrm{E}(\alpha) < \infty$. Conversely, if $\mathrm{E}(\alpha) < \infty$ then a stationary distribution exists, namely

$$\pi_0 = \frac{1}{\mathrm{E}(\alpha)}, \tag{3.41}$$

$$\pi_j = \frac{\sum_{k \geq j} \alpha_k}{\mathrm{E}(\alpha)}, \quad j \geq 1, \tag{3.42}$$

since these probabilities sum to 1 and satisfy (3.39) [verify]. By Theorem 3.1, this is the unique limiting distribution.

Note: Here the mean return time to state 0 is $\mu_{00} = \mathrm{E}(\alpha)$, , so (3.41) satisfies the general relation $\pi_j = \frac{1}{\mu_{jj}}$ for $j = 0$. □

Exercise 3.3. In Example 3.5 find the mean return times μ_{ij}, $i, j \geq 0$. □

Example 3.6 *(Simple random walk with partially reflecting boundary at 0; compare to Example 3.4 (Gambler's Ruin) and Exercise 3.2).* The tpm is

$$P^* = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ \vdots \end{array} \begin{array}{c} \begin{array}{cccccc} 0 & 1 & 2 & 3 & 4 & \cdots \end{array} \\ \begin{pmatrix} q & p & 0 & 0 & 0 & \cdots \\ q & 0 & p & 0 & 0 & \cdots \\ 0 & q & 0 & p & 0 & \cdots \\ 0 & 0 & q & 0 & p & \cdots \\ 0 & 0 & 0 & q & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{pmatrix} \end{array}. \tag{3.23$'$}$$

where $0 < p < 1$ and q=$1 - p$. This chain is irreducible and aperiodic, transient for $p > 1/2$, and recurrent for $p \leq 1/2$ [compare to SRW]. In the recurrent case we now show that a limiting distribution exists iff $p < 1/2$, in which case the chain is positive recurrent; it is null recurrent if $p = 1/2$.

[5] If $\pi_0 = 0$ then $\pi_j = 0$ for all j by (3.39), which is impossible.

The stationary equations are

$$
\begin{aligned}
\pi_0 &= q\pi_o + q\pi_1 &&\Rightarrow p\pi_0 = q\pi_1 &&\Rightarrow \pi_1 = \delta\pi_0 \\
\pi_1 &= p\pi_0 + q\pi_2 &&\Rightarrow (\delta - p)\pi_0 = q\pi_2 &&\Rightarrow \pi_2 = \delta^2\pi_0 \\
\pi_2 &= p\pi_1 + q\pi_3 &&\Rightarrow (\delta^2 - \delta p)\pi_0 = q\pi_3 &&\Rightarrow \pi_3 = \delta^3\pi_0 \\
\pi_3 &= p\pi_2 + q\pi_4 &&\Rightarrow (\delta^3 - \delta^2 p)\pi_0 = q\pi_4 &&\Rightarrow \pi_4 = \delta^4\pi_0 \\
&\vdots
\end{aligned}
\tag{3.43}
$$

where $\delta = p/q \leq 1$. Sum these equations together with $\pi_0 = \pi_0$ to obtain

$$\sum\nolimits_{j\geq 0} \pi_j = \pi_0 \sum\nolimits_{j\geq 0} \delta^j. \tag{3.44}$$

Thus, if a stationary distribution π exists then[6] $\pi_0 > 0$ and $\sum_{j\geq 0} \delta^j < \infty$, hence $\delta < 1$, i.e., $p < 1/2$.

Conversely, if $p < 1/2$ then $\delta < 1$, so (3.44) yields

$$\sum\nolimits_{j\geq 0} \pi_j = \frac{\pi_0}{1-\delta}. \tag{3.45}$$

Thus a stationary distribution exists, given by

$$\pi_0 = 1 - \delta, \tag{3.46}$$

$$\pi_j = (1-\delta)\delta^j, \quad j \geq 1, \tag{3.47}$$

since these probabilities sum to 1 and satisfy (3.43). By Theorem 3.1, this is the unique limiting distribution, the geometric$(1-\delta)$ distribution.

Note: From (3.46) and (3.35), the mean return time to state 0 is given by $\mu_{00} = \frac{1}{1-\delta}$, the mean of this geometric distribution. □

Exercise 3.4*. In Example 3.6 find the mean return times μ_{ij}, $i, j \geq 0$. □

Example 3.7 *(Variant of Success Runs; no upper bound).* As in Example 2.4, let X_n be the number of consecutive successes (S) up to and including

[6] If $\pi_0 = 0$ then $\pi_j = 0$ for all j by (3.43), which is impossible.

the nth trial. The process renews after each failure (F) and continues indefinitely. The state space is $\{0, 1, 2, \ldots\}$ and 0 is not an absorbing state, so the game never terminates. For added interest and applicability, assume that the success/failure probability depends on the current length of the run, so the tpm has the form

$$
\begin{array}{cc}
 & \begin{array}{cccccc} 0 & 1 & 2 & 3 & 4 & \cdots \end{array} \\
P = \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \\ \vdots \end{array} & \begin{pmatrix} p_0 & q_0 & 0 & 0 & 0 & \cdots \\ p_1 & 0 & q_1 & 0 & 0 & \cdots \\ p_2 & 0 & 0 & q_2 & 0 & \cdots \\ p_3 & 0 & 0 & 0 & q_3 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{pmatrix},
\end{array}
\tag{3.48}
$$

where $q_i = 1 - p_i$. For simplicity assume that $0 < p_i < 1$ for all i, so the chain is irreducible and aperiodic.

To determine transience/recurrence, note that for each $n \geq 0$,

$$\Pr_0[N_0 > n] = \prod_{i=0}^{n} (1 - p_i),$$

where N_0 is the time of the first return to 0. Thus [PK Lemma 4.4.1 p.207],

$$\Pr_0[N_0 = \infty] = \prod_{i=0}^{\infty} (1 - p_i) \begin{cases} > 0 & \text{if } \sum p_i < \infty; \\ = 0 & \text{if } \sum p_i = \infty, \end{cases} \tag{3.49}$$

so the chain is transient (recurrent) iff $\sum p_i$ is convergent (divergent). (In particular, if all $p_i = p > 0$ then the chain is recurrent.)

To determine positive vs. null recurrence, consider the stationary equations, beginnining with π_1:

$$
\begin{aligned}
\pi_1 &= q_0 \pi_0 = (1 - p_0)\pi_0 \\
\pi_2 &= q_1 \pi_1 = (1 - p_0)(1 - p_1)\pi_0 \\
\pi_3 &= q_2 \pi_2 = (1 - p_0)(1 - p_1)(1 - p_2)\pi_0 \\
&\vdots
\end{aligned}
\tag{3.50}
$$

Proceed as in Examples 3.4 and 3.5 to conclude that a stationary distribution exists iff[7]

$$\sum_{k\geq 0}\prod_{i=0}^{k}(1-p_i)<\infty, \tag{3.51}$$

in which case the unique stationary and limiting distribution is given by

$$\pi_0=\frac{1}{1+\sum_{j\geq 0}\prod_{i=0}^{j}(1-p_i)}, \tag{3.52}$$

$$\pi_j=\frac{\prod_{i=0}^{j-1}(1-p_i)}{1+\sum_{j\geq 0}\prod_{i=0}^{j}(1-p_i)},\quad j\geq 1. \tag{3.53}$$

If $p_i=p$ for all i then (3.51) holds [verify] and (3.52)-(3.53) become

$$\begin{aligned}\pi_0&=p,\\ \pi_j&=p(1-p)^j,\quad j\geq 1.\end{aligned}$$

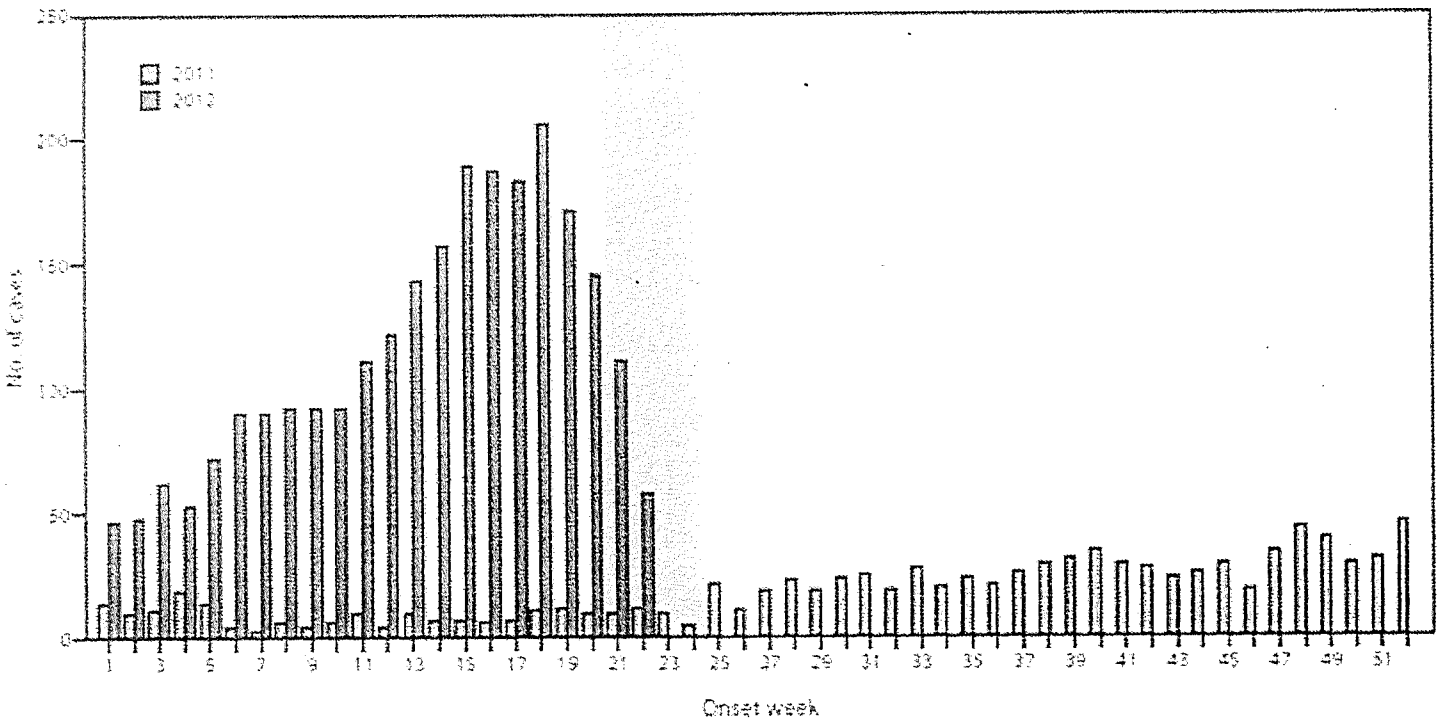
again a geometric distribution, with the mean return time to 0 given by $\mu_{00}=\frac{1}{\pi_0}=\frac{1}{p}$.

Note: Suppose $p_i=q$ for all i. (We have interchanged p and q to conform to the notation in Example 3.6.) Then this chain is "driven back" to 0 faster than the RW in Example 3.6, so its mean return time $\mu_{00}=\frac{1}{\pi_0}=\frac{1}{q}$ should be smaller than that in Example 3.6, namely $\frac{1}{\pi_0}=\frac{1}{1-\delta}$. Indeed $\frac{1}{q}=\frac{1}{1-p}<\frac{1}{1-\delta}$, which is true because $p<\frac{p}{q}=\delta$. □

3.4. Time-reversible Markov chains. [To be added...]

[7] Note that (3.51) $\Rightarrow \prod_{i=0}^{j}(1-p_i)\to 0$ as $j\to\infty$, i.e., $\prod_{i=0}^{\infty}(1-p_i)=0$, so (3.51) is a stronger condition than the condition for recurrence in (3.49).

FIGURE 1. Number of confirmed and probable pertussis cases reported, by week of onset — Washington, January 1, 2011 2012*



* Reports for 2012 as of June 16, 2012. The shaded area represents a lag during which additional cases likely occurred during 2012, b been reported to the Washington State Department of Health.

6. Examples

Example 6.1: Whooping cranes in North America. Miller et al. (1974) give the annual counts of rare migrating whooping cranes arriving in Texas from 1938 through 1972; see Guttorp (1991, p.47, p.190). Here $x_0 = 14$, $n = 34$ $(= 1972 - 1938)$, and $\{x_\nu \mid 1 \leq \nu \leq 34\}$ is observed, which we assume to arise from a semicritical GW process.

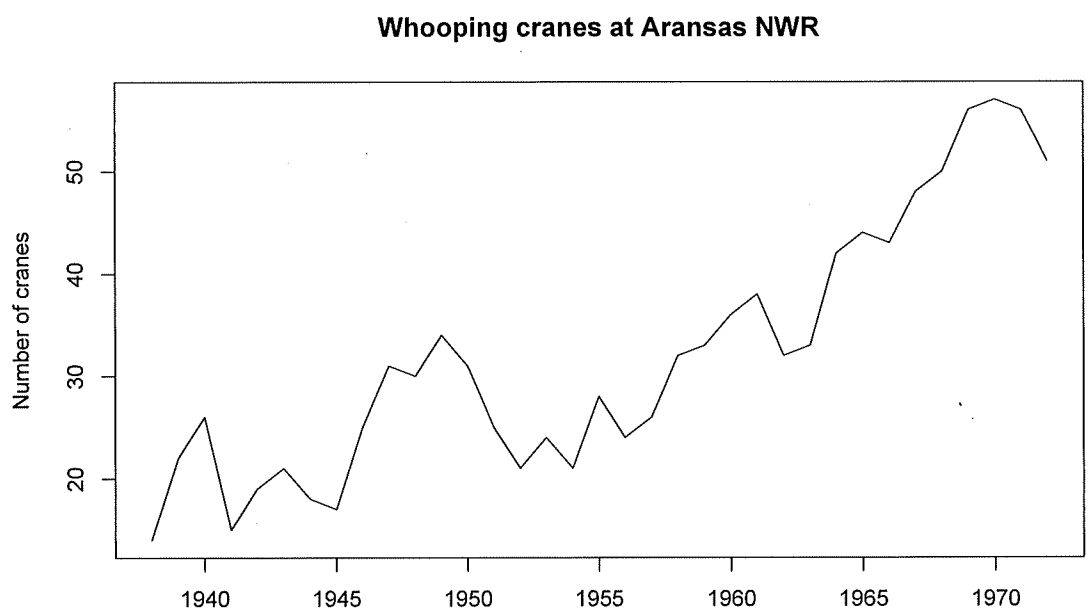


Figure 1: North American whooping crane population counts 1938-1972.

(i) Assume that extinction is considered catastrophic, so Case 1 obtains and the formulation and discussion in Section 3 are appropriate. Suppose first that u is unknown. Because $x_{34} = 51$ was observed in 1972, the universal upper bounds (45) and (46) for the p-value $\pi^0(x_n; x_0)$ against extinction (cf. (43)) are given by $\pi_3(51; 14) = .27$ and $\pi_4(51; 14) = .12$ respectively. In 1970 $x_{32} = 57$, so these bounds for that year reduce to $\pi_3(57; 14) = .25$ and $\pi_4(57; 14) = .09$ resp. These provide moderate but not conclusive evidence against eventual extinction and thus moderate evidence against the need for further intervention.

4. Discrete-time Branching Processes

A *discrete-time branching process* $\equiv$ *Galton-Watson (GW) process* is a Markov chain that describes the growth or decline of a population that reproduces by simple branching, or splitting. Applications include modeling population growth, the spread of epidemics, and nuclear chain reactions.

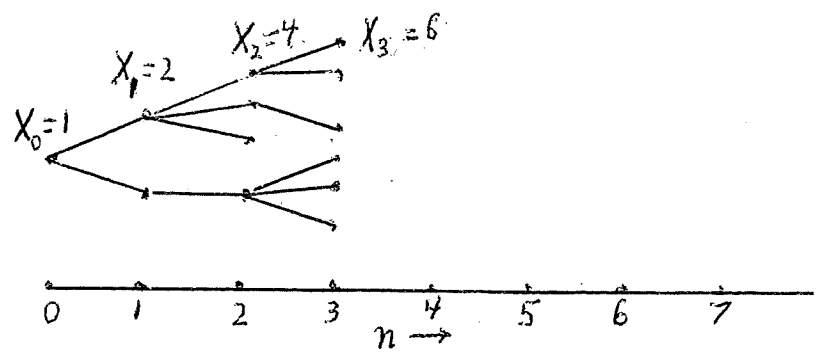
For each time $n = 0, 1, 2...$ let X_n denote the population size at time n; assume $X_0 = 1$. At time $n = 0$ this single individual splits into a random number $X_1 = \xi_1^{(0)} \sim \xi$ of first-generation offspring, where ξ has probability distribution $\mathbf{p} \equiv (p_0, p_1, p_2, \ldots)$ on $\{0, 1, 2, \ldots\}$. At time $n = 1$ each of these offspring independently splits into second-generation offspring according to the same distribution, that is,

$$X_2 = \xi_1^{(1)} + \cdots + \xi_{X_1}^{(1)}, \tag{4.1}$$

where $\xi_1^{(2)}, \ldots, \xi_{X_1}^{(2)}$ are iid rvs each with distribution $\mathbf{p}$. In general,

$$X_{n+1} = \xi_1^{(n)} + \cdots + \xi_{X_n}^{(n)}, \tag{4.2}$$

where $\xi_1^{(n)}, \ldots, \xi_{X_n}^{(n)}$ are iid rvs each with distribution $\mathbf{p}$.



By (4.2), $\mathcal{L}(X_{n+1}) \mid \{X_1, \ldots, X_n\}$ depends only on X_n, hence $\{X_n\}$ is a Markov chain with state space $\{0, 1, 2, \ldots\}$. Major questions include:

- What is the long-term behavior of X_n? Does $X_n \to 0$ (extinction), $X_n \to \infty$ (explosion), or might $\{X_n\}$ remain bounded?
- If $X_n \to 0$, what is the distribution of $T \equiv$ time to extinction? $\mathrm{E}(T)$?
- If $X_n \to \infty$, i.e. the population explodes, at what rate does it grow?

If $p_0 = 1$ then $X_n \equiv 0$ for all $n \geq 1$; if $p_1 = 1$ then $X_n \equiv 1$ for all $n \geq 0$. Thus to avoid these trivial cases we assume that $p_0 < 1$ and $p_1 < 1$.

4.1. The mean and variance of X_n

Some information about the long-term behavior of X_n can be obtained from its first and second moments. Set $\mathrm{E}(\xi) = \mu$ and $\mathrm{Var}(\xi) = \sigma^2$. From (4.1),

$$\begin{aligned}
\mathrm{E}(X_2) &= \mathrm{E}\left[\mathrm{E}\left(X_2 \mid X_1\right)\right] \\
&= \mathrm{E}\left[\mathrm{E}\left(\xi_1^{(1)} + \cdots + \xi_{X_1}^{(1)} \mid X_1\right)\right] \\
&= \mathrm{E}\left[\mu X_1\right] \\
&= \mu^2.
\end{aligned}$$

Similarly, from (4.2) and induction,

$$\mathrm{E}\left(X_n\right) = \mu^n \begin{cases} \to 0 & \text{if } \mu < 1; \\ = 1 & \text{if } \mu = 1; \\ \to \infty & \text{if } \mu > 1. \end{cases} \tag{4.3}$$

Next, again from (4.1), (4.2), and induction,

$$\begin{aligned}
\mathrm{Var}(X_2) &= \mathrm{E}\left[\mathrm{Var}\left(X_2 \mid X_1\right)\right] + \mathrm{Var}\left[\mathrm{E}\left(X_2 \mid X_1\right)\right] \\
&= \mathrm{E}\left[\sigma^2 X_1\right] + \mathrm{Var}\left[\mu X_1\right] \\
&= \sigma^2\mu + \sigma^2\mu^2 \\
&= \sigma^2\mu(1+\mu),
\end{aligned}$$

$$\mathrm{Var}\left(X_n\right) = \sigma^2\mu^{n-1}(1 + \mu + \cdots + \mu^{n-1}) \tag{4.4}$$

$$= \begin{cases} \sigma^2\mu^{n-1}\left(\frac{\mu^n - 1}{\mu - 1}\right) & \text{if } \mu \neq 1; \\ \sigma^2 n & \text{if } \mu = 1. \end{cases} \tag{4.5}$$

Thus if $\mu < 1$ then $\mathrm{E}(X_n) \to 0$ and $\mathrm{Var}(X_n) \to 0$, hence $X_n \xrightarrow{p} 0$ by Chebyshev's inequality. If $\mu > 1$ then $\mathrm{E}(X_n) \to \infty$ and $\mathrm{Var}(X_n) \to \infty$ at geometric rates, while if $\mu = 1$ then $\mathrm{E}(X_n)$ remains constant while $\mathrm{Var}(X_n) \to \infty$ at a linear rate. However, for more precise information about the limiting behavior of X_n we need to consider its probability generating function.

4.2. The probability generating function (pgf) of X_n

For $k = 1, 2, \ldots$ let $\mathbf{p}_k \equiv (p_{k0}, p_{k1}, p_{k2}, \ldots)$ denote the probability distribution of $\xi_1 + \cdots + \xi_k$, the sum of k iid copies of ξ; thus $\mathbf{p}_1 = \mathbf{p}$. By (4.2), the tpm of the Markov chain $\{X_n\}$ is

$$
(4.6) \qquad P = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ \vdots \\ k \\ \vdots \end{array} \begin{array}{c} \begin{array}{ccccc} 0 & 1 & 2 & 3 & \cdots \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots \\ p_{10} & p_{11} & p_{12} & p_{13} & \cdots \\ p_{20} & p_{21} & p_{22} & p_{23} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \\ p_{k0} & p_{k1} & p_{k2} & p_{k3} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \end{pmatrix} \end{array}.
$$

Clearly $\{0\}$ is an absorbing state and $p_{k0} = (p_0)^k$. If $p_0 > 0$ then each state $k \geq 1$ is transient, so $\Pr[1 \leq X_n \leq m \text{ i.o.}] = 0$ for any $m < \infty$, hence

$$
(4.7) \qquad \Pr\left[X_n \to 0 \text{ or } X_n \to \infty\right] = 1.
$$

If $p_0 = 0$ then death is not possible and, since $p_1 < 1$,

$$
p_{kl} \begin{cases} = 0 & \text{if } l < k; \\ = (p_1)^k < 1 & \text{if } l = k; \\ > 0 & \text{for at least one } l > k \end{cases}
$$

for each state $k \geq 1$. Here $\sum_n p_{kk}^{(n)} = \sum_n (p_1^k)^n < \infty$ so k is transient and

$$
(4.8) \qquad \Pr\left[X_n \to \infty\right] = 1,
$$

In general, however, it is not easy to obtain explicit expressions for p_{kl} to determine the probabilities of the two possibilities in (4.7), as well as the rates of convergence. For these purposes we utilize the *probability generating function (pgf)* $\phi \equiv \phi_\xi$ of the rv ξ:

$$
(4.9) \qquad \phi(s) = \mathrm{E}(s^\xi) = p_0 + p_1 s + p_2 s^2 + p_3 s^3 + \cdots,
$$

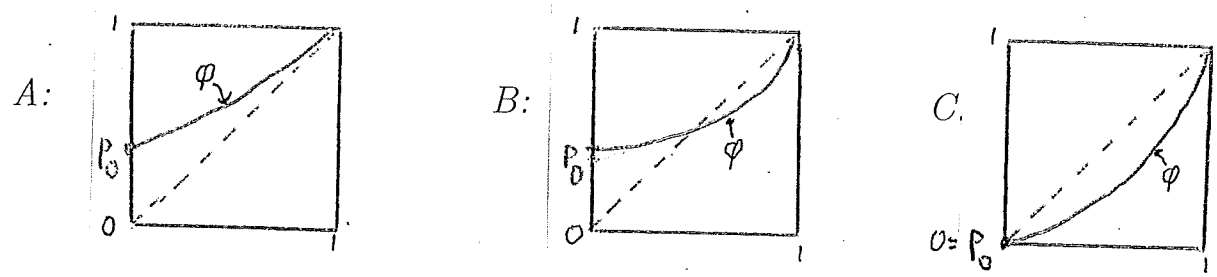
defined for $0 \leq s \leq 1$. Note that $\phi_\xi(s) = m_\xi(\log s)$ where m_ξ is the mgf of ξ.

Some properties of ϕ :

(a) $\phi(0) = p_0, \quad \phi(1) = 1, \quad 0 \le \phi(s) \le 1$ for $0 \le s \le 1$.

(b) $\phi'(s) = p_1 + 2p_2 s + 3p_3 s^2 + \cdots > 0$ for $0 < s \le 1$, so ϕ is strictly increasing on $[0,1]$. Also $\phi'(1) = \mathrm{E}(\xi) \equiv \mu$, the mean of the offspring distribution.

(c) $\phi''(s) = 2\cdot 1\, p_2 + 3\cdot 2\, p_3 s + \cdots \ge 0$, so ϕ is convex. In fact, unless $p_0 + p_1 = 1$ (no growth possible), $\phi''(s) > 0$ so ϕ is strictly convex. Three cases occur:



$\phi'(1) \le 1$, (so $p_0 > 0$) $\qquad$ $\phi'(1) > 1$, $p_0 > 0$ $\qquad$ $\phi'(1) > 1$, $p_0 = 0$

(d) $\phi'(1) = p_1 + 2p_2 + 3p_3 + \cdots = \mathrm{E}(\xi) \equiv \mu$.

$\phi''(1) = 2\cdot 1\, p_2 + 3\cdot 2\, p_3 + 4\cdot 3\, p_4 + \cdots = \mathrm{E}\,[\,\xi\,(\xi - 1)\,]$.

$\phi^{(r)}(1) = \mathrm{E}\,[\,\xi\,(\xi-1)\cdot(\xi - r + 1)\,]$ $\qquad$ [the rth "factorial" moment].

(e) $\phi(0) = p_0, \ \phi'(0) = p_1, \ \phi''(0) = 2p_2, \ \phi'''(0) = 3\cdot 2p_3, \ldots,$ so

$$p_r = \frac{1}{r!}\phi^{(r)}(0), \qquad r = 0, 1, 2, \ldots \ . \tag{4.10}$$

Thus the entire probability distribution can be recovered from ϕ, hence ϕ uniquely determines the distribution of ξ.

(f) Suppose that γ and η are independent integer-valued rvs with range $\{0, 1, 2, \ldots\}$. Then *the pgf of* $\gamma + \eta$ *is the product of the pgfs of* γ *and* η*:*

$$\phi_{\gamma+\eta}(s) = \mathrm{E}(s^{\gamma+\eta}) = \mathrm{E}(s^\gamma s^\eta) = \mathrm{E}(s^\gamma)\mathrm{E}(s^\eta) = \phi_\gamma(s)\phi_\eta(s). \tag{4.11}$$

As a consequence, if $\xi_1, \ldots, \xi_k$ are iid copies of ξ then

$$\phi_{\xi_1+\cdots+\xi_k}(s) = (\phi(s))^k\,. \tag{4.12}$$

Lemma 4.1. *Suppose that $\xi_1, \xi_2, \ldots$ are iid copies of ξ and are independent of the non-negative integer-valued rv N having pgf ϕ_N. Then*

$$\phi_{\xi_1+\cdots+\xi_N}(s) = \phi_N(\phi_\xi(s)). \tag{4.13}$$

Proof.
$$\begin{aligned}\phi_{\xi_1+\cdots+\xi_N}(s) &= \mathrm{E}\left(s^{\xi_1+\cdots+\xi_N}\right)\\ &= \mathrm{E}\left[\mathrm{E}\left(s^{\xi_1+\cdots+\xi_N}\right) \mid N\right]\\ &= \mathrm{E}[\,(\phi_\xi(s))^N\,] \qquad\qquad [\text{by } (4.12)]\\ &= \phi_N(\phi_\xi(s)).\end{aligned}$$
□

Proposition 4.1. *(i) For the GW process $\{X_n\}$, the pgf of X_n is the nth functional iterate ϕ_n of ϕ, that is,*

$$\phi_{X_n}(s) \equiv \phi_n(s) = \underbrace{\phi(\cdots\phi(\phi(s))\cdots)}_{n \text{ times}}. \tag{4.14}$$

(ii) The joint pgf $\phi_{X_1,\ldots,X_n}(s_1,\ldots,s_n) \equiv \mathrm{E}(s_1^{X_1}\cdots s_n^{X_n})$ of $(X_1,\ldots,X_n)$ is

$$\phi_{X_1,\ldots,X_n}(s_1,\ldots,s_n) = \phi(s_1\phi(s_2\phi(\ldots s_{n-1}\phi(s_n)\ldots))). \tag{4.15}$$

Note: Set $s_1 = \cdots = s_{n-1} = 1$ and $s_n = s$ in (4.15) to obtain (4.14).

Proof. (i) By (4.1) and (4.13) with $N = X_1$,

$$\phi_{X_2}(s) = \phi_{\xi_1^{(1)}+\cdots+\xi_{X_1}^{(1)}}(s) = \phi(\phi(s)).$$

The general case of (4.14) follows by induction using (4.2). □

Exercise 4.1. Prove (4.15). □

Thus, because the (limiting) pgf determines the (limiting) distribution, $\lim_{n\to\infty}\phi_n(s) \equiv \phi_\infty(s)$ determines the limiting distribution of X_n. Note that this limit always exists: if $\phi(s) \geq s$ $(< s)$ then since ϕ is increasing, $\phi_{n+1}(s) \geq \phi_n(s)$ $(< \phi_n(s))$ so $\{\phi_n(s)\}$ is monotone in n. In fact, for all $0 \leq s \leq 1$ it follows from the continuity of ϕ that

$$\phi_\infty(s) = \lim_{n\to\infty}\phi_{n+1}(s) = \lim_{n\to\infty}\phi(\phi_n(s)) = \phi\left(\lim_{n\to\infty}\phi_n(s)\right) = \phi(\phi_\infty(s)),$$

so $\phi_\infty(s)$ must be a solution of the equation

$$\phi(s) = s. \tag{4.16}$$

Because $\phi(s)$ is strictly increasing and convex on $[0,1]$ with $\phi(0) \geq 0$, $\phi(1) = 1$, and $p_1 < 1$, the equation (4.16) can have at most two solutions in $[0,1]$, the larger of which is always 1. Let $u \leq 1$ denote the *smaller solution.*

Example 4.1a *(The linear case: no growth possible, extinction certain).* Here $p_0 + p_1 = 1$, $p_2 = 0$, $p_3 = 0$, ... and $\phi(s) \equiv p_0 + p_1 s$ is linear in s; the generating rv ξ is a 0-1 Bernoulli(p_1) rv. The equation (4.16) is $p_0 + p_1 s = s$, which, since $p_1 = 1 - p_0$, has the single solution $u = 1$. Thus $\phi_\infty(s) = 1$ for all $0 \leq s \leq 1$, which is the pgf of the distribution degenerate at 0.

Here the distribution of X_n can be obtained directly from its pgf ϕ_n:

$$\phi_n(s) = (1 - p_1^n) + p_1^n s, \tag{4.17}$$

which is again linear in s. Because $p_1 < 1$,

$$\Pr[X_n = 0] = \phi_n(0) = 1 - p_1^n \to 1 \quad \text{as } n \to \infty,$$

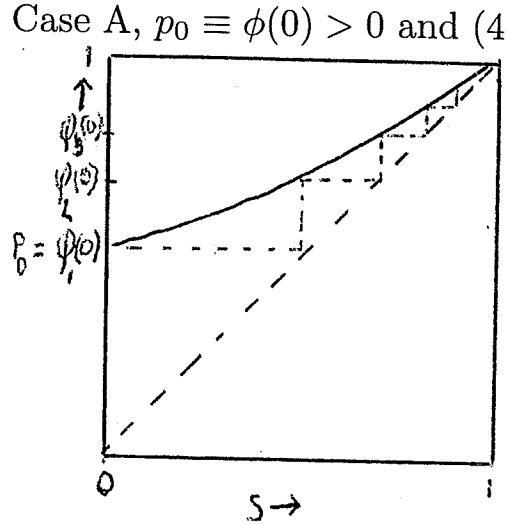
so $X_n \xrightarrow{p} 0$ at a geometric rate. Note that $\phi'(1) \equiv \mu = p_1 < 1$, so Case *A* ($\mu \leq 1$) holds; in fact Subcase *A1* ($\mu < 1$) holds (see below). □

4.3. The limiting behavior of X_n via its pgf ϕ_n

Since $X_n = 0 \Rightarrow X_{n+1} = 0$, $\{X_n \to 0\} = \cup_{n=1}^\infty \{X_n = 0\} = \lim_{n\to\infty} \{X_n = 0\}$, so

$$\Pr[X_n \to 0] = \lim_{n\to\infty} \Pr[X_n = 0] = \lim_{n\to\infty} \phi_n(0) = \phi_\infty(0). \tag{4.18}$$

- In Case A, $p_0 \equiv \phi(0) > 0$ and (4.16) has no solutions in $[0,1)$, so $u = 1$:



A1: $\mu < 1$: geometric rate;

A2: $\mu = 1$: linear rate.

Here it is easy to see from the geometry that this occurs if and only if $\phi'(1) \equiv \mu \leq 1$. Also in this case,

$$\Pr[X_n = 0] \equiv \phi_n(0) \uparrow 1, \tag{4.19}$$

so $\Pr[X_n \to 0] = 1$. Thus *extinction occurs with probability 1.* Note that

$$\phi_n(s) \uparrow 1_{[0,1]}(s) = 1 \ (\equiv u) \tag{4.20}$$

for *each* $s \in [0,1]$, the pgf of the distribution degenerate at 0.

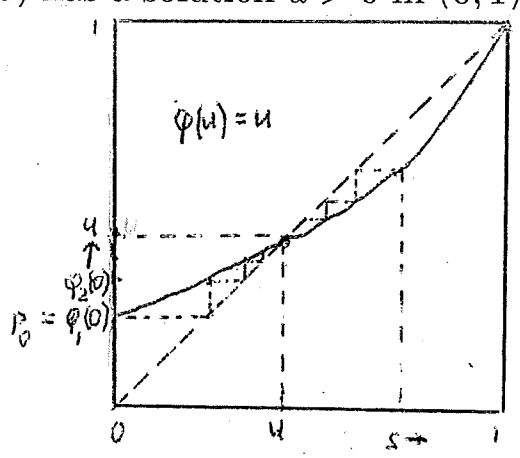
Subcase A1: $\phi'(1) \equiv \mu < 1$.

Here the convergence in (4.19) occurs at a *geometric rate* and $\mathrm{E}(T) < \infty$.

Subcase A2: $\phi'(1) \equiv \mu = 1$.

Here the convergence in (4.19) occurs at a *linear rate* and $\mathrm{E}(T) = \infty$ - see Proposition 4.3. Another distinction between A1 and A2 appears in (4.35).

- In Case B, (4.16) has a solution $u > 0$ in $(0,1)$; again $p_0 \equiv \phi(0) > 0$:



Here it is easy to see from the geometry that this occurs if and only if $\phi'(1) \equiv \mu > 1$ and $p_0 > 0$. In this case,

$$\Pr[X_n = 0] \equiv \phi_n(0) \uparrow u < 1. \tag{4.21}$$

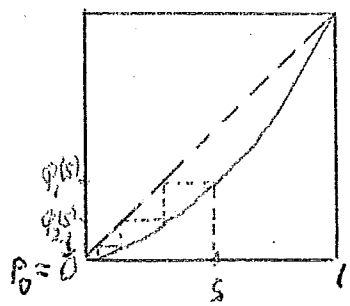
so $\Pr[X_n \to 0] = u$ and $\Pr[X_n \to \infty] = 1 - u$. Thus *extinction occurs with probability* u, *explosion occurs with probability* $1 - u$. (Note that $u > p_0$.)

From the geometry, note that $\phi'(u) < 1$ so the convergence in (4.21) also occurs at a geometric rate. Also note that

$$\phi_n(s) \to u1_{[0,1]}(s) + (1-u)1_{\{1\}}(s) \equiv \begin{cases} u & \text{if } 0 \le s < 1; \\ 1 & \text{if } s = 1, \end{cases} \tag{4.22}$$

which is the pgf of the distribution with mass u at 0 and mass $1-u$ at ∞.

- In Case C, (4.16) has the solution $u = 0$ in $[0,1)$; necessarily $p_0 \equiv \phi(0) = 0$.



This is the no-death case, so $\Pr[X_n \to 0] = 0$. What about $\Pr[X_n \to \infty]$? From the geometry, this case occurs if and only if $\phi'(1) \equiv \mu > 1$ and $p_0 = 0$, so $\phi_n(0) = 0$ for every n. Also in this case,

$$\phi_n(s) \to 1_{\{1\}}(s) \equiv \begin{cases} 0 & \text{if } 0 \le s < 1; \\ 1 & \text{if } s = 1, \end{cases} \tag{4.23}$$

which is the pgf of the distribution degenerate at $+\infty$. This indicates that $\Pr[X_n \to \infty] = 1$, which agrees with (4.8). Thus, *explosion occurs with probability* 1. [For $j \ge 1$, $\sum_n p_{jj}^{(n)} = \sum_n (p_1)^n < \infty$, so j is transient.]

Remark 4.1. Notice that in all three cases, $\Pr[\text{extinction}] = u$. Also, observe from (4.20), (4.22), and (4.23) that $\phi_\infty(s) \equiv \lim_{n\to\infty} \phi_n(s)$ exists for every $s \in [0,1]$ and is constant in s except possibly for a jump to 1 at $s = 1$. That is,

$$\phi_\infty(s) = \begin{cases} u, & 0 \le s < 1; \\ 1, & s = 1. \end{cases} \tag{4.24}$$

[See 3 preceding figures (in margins).]

Example 4.2 *(Binomial offspring distribution).*
If $\xi \sim \text{Binomial}(r, p)$, then $\mu \equiv \text{E}(\xi) = rp$, so

$$\mu \begin{cases} \leq 1 & \text{iff } p \leq \frac{1}{r} \quad \textit{(Case A)} \\ > 1 & \text{iff } p > \frac{1}{r} \quad \textit{(Case B)}. \end{cases}$$

In Case A (B), extinction occurs with probability 1 (probability $u < 1$).

Suppose that $p > \frac{1}{r}$. To find the extinction probability u we must solve the equation (4.16), i.e. $\phi(s) = s$. Since $\phi(s) = (q + ps)^r$ [verify], this is a polynomial equation in s of degree r. When $r = 2$, this becomes

$$(q + ps)^2 = s, \tag{4.25}$$

where $q = 1 - p$. This is equivalent to the quadratic equation,

$$p^2 s^2 + (2pq - 1)s + q^2 = 0,$$

which has the solutions

$$\frac{-(2pq-1) \pm \sqrt{1-4pq}}{2p^2} = \frac{-(2pq-1) \pm (2p-1)}{2p^2} = \left\{1, \left(\frac{q}{p}\right)^2\right\}.$$

Thus the extinction probability is $u = (\frac{q}{p})^2$ when $p > 1/2$, so $0 < u < 1$. □

Exercise 4.2a *(Geometric offspring distribution).*
Suppose that $\xi \sim \text{geometric}(p)$, that is, $p_k = pq^k$, $k = 0, 1, 2, \ldots$, $0 < p < 1$. Find the pgf $\phi \equiv \phi_p$ and use this to find $\mu = \text{E}(\xi)$. For which values of p does Case A1 hold? A2? B? (Case C cannot hold since $p_0 \equiv p > 0$.) Find the extinction probability u in all cases. □

Example 4.3a *(Poisson offspring distribution).*
Suppose that $\xi \sim \text{Poisson}(\lambda)$, that is, $p_k = e^{-\lambda}\lambda^k/k!$ for $k = 0, 1, \ldots$. Here $\phi(s) \equiv \phi_\lambda(s) = e^{\lambda(s-1)}$ and $\mu \equiv \text{E}(\xi) = \lambda$, so extinction occurs with probability 1 iff $\lambda \leq 1$ (Case A). If $\lambda > 1$ (Case B), then extinction occurs with probability u, where by (4.16) u is the unique value in $(0, 1)$ that satisfies

$$e^{\lambda(u-1)} = u. \tag{4.26}$$

An explicit solution is not available so this must be solved numerically. (Upper and lower bounds for u are given in Exercise 4.4d.) □

4.4. The total population size and time to extinction

Let T denote the time to extinction and Y the total population size, so

$$T = \min\{n \geq 1 \mid X_n = 0\}, \tag{4.27}$$

$$Y = \sum_{n=0}^{\infty} X_n = \sum_{n=0}^{T-1} X_n. \tag{4.28}$$

Thus $T, Y < \infty$ $(= \infty)$ $\iff$ extinction (explosion) occurs. From §4.3,

$$\Pr[T, Y < \infty] = u \begin{cases} = 1 & \text{if } \mu \leq 1 & \textit{(Case A)}; \\ \in (0,1) & \text{if } \mu > 1 \text{ and } p_0 > 0 & \textit{(Case B)}; \\ = 0 & \text{if } \mu > 1 \text{ and } p_0 = 0 & \textit{(Case C)}. \end{cases} \tag{4.29}$$

We seek the distributions of T and Y; for the latter, consider the "pgf"

$$\Phi(s) \equiv \Phi_Y(s) = \mathrm{E}(s^Y), \qquad 0 \leq s \leq 1. \tag{4.30}$$

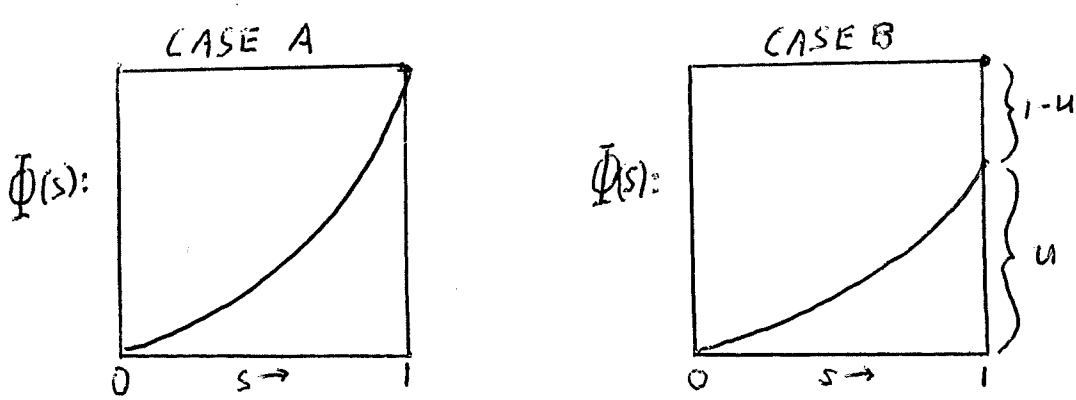
Since $Y \geq X_0 = 1$, Y has range $\{1, 2. \ldots, \infty\}$, so

$$\begin{aligned} \Phi(s) &= \sum_{k=1}^{\infty} \Pr[Y = k] s^k + \Pr[Y = \infty] s^\infty \\ &= p_0 s + p_1 p_0 s^2 + (p_1^2 p_0 + p_2 p_0^2) s^3 + \cdots + \Pr[Y = \infty] \cdot 1_{\{1\}}(s) \\ &= s \cdot [p_0 + p_1 p_0 s + (p_1^2 p_0 + p_2 p_0^2) s^2 + \cdots] + \Pr[Y = \infty] \cdot 1_{\{1\}}(s). \end{aligned}$$

Thus $\Phi(0) = 0$, $\Phi(s)$ is nondecreasing and convex on $[0, 1]$,

$$\Phi(1-) = \Pr[Y < \infty] = u \leq 1, \tag{4.31}$$

and $\Phi(1) = 1$, so Φ *has a jump of size* $1 - u$ *at* $s = 1$:



Proposition 4.2. $\Phi(s)$ is the unique solution to the functional equation

$$\Phi(s) = s \cdot \phi(\Phi(s)), \quad 0 \le s < 1. \tag{4.32}$$

Proof. Since $X_0 = 1$, condition on X_1 and apply Lemma 4.1 to obtain

$$\begin{aligned}\Phi(s) = \mathrm{E}[s^Y] &= \mathrm{E}[s^{1+Y_1+\cdots+Y_{X_1}}] \\ &= s \cdot \phi(\Phi(s)).\end{aligned}$$

Here, $Y_1, \ldots, Y_{X_1}$ denote the total population sizes descending from the X_1 individuals in the first generation. They are iid with each $Y_i \sim Y$.

For uniqueness, set $\psi_s(x) = s\phi(x)$ and rewrite (4.32) as $\Phi = \psi_s(\Phi)$. If $0 < s < 1$, $\psi_s(0) \ge 0$, $\psi_s(1) < 1$, and $\psi_s(x)$ is convex, so the solution is unique (draw figures). If $s = 0$, $\psi_s \equiv 0$ so $\Phi = 0$ is the unique solution. □

Remark 4.2. Since $\phi(s)$ is continuous at $s = 1$, let $s \uparrow 1$ in (4.32) to obtain

$$\Phi(1-) = \phi(\Phi(1-)), \tag{4.33}$$

so $\Phi(1-)$ satisfies $s = \phi(s)$ (as it must, since $\Phi(1-) = u$ by (4.31)). □

Case A: $\Pr[Y < \infty] \equiv u = 1 \iff \mu \equiv \mathrm{E}(\xi) = \phi'(1) \le 1$.

Here Φ is a bona fide pgf (no jump at 1) and can be obtained in principle from (4.32), which also yields $\mathrm{E}(Y)$ directly:

$$\Phi'(s) = s\,\phi'(\Phi(s))\Phi'(s) + \phi(\Phi(s)), \quad 0 \le s \le 1, \tag{4.34}$$

so, letting $s \uparrow 1$,

$$\Phi'(1) = \phi'(1)\Phi'(1) + 1$$

since $\Phi(1) = 1 = \phi(1)$. Therefore

$$\mathrm{E}(Y) \equiv \Phi'(1) = \frac{1}{1-\phi'(1)} \equiv \frac{1}{1-\mu} \begin{cases} < \infty & \text{if } \mu < 1 \textit{ (Case A1)}; \\ = \infty & \text{if } \mu = 1 \textit{ (Case A2)}. \end{cases} \tag{4.35}$$

Case B: $0 < \Pr[Y = \infty] \equiv 1 - u < 1 \iff \mu \equiv \phi'(1) > 1$ and $p_0 > 0$.

Here Φ has a jump at 1 so is not a bona fide pgf, but still can be obtained in principle from (4.32). Here $\mathrm{E}(Y) = \infty$, but we will obtain $\mathrm{E}[Y|Y < \infty]$, *the expected total population size conditioned on eventual extinction.*

To study the conditional distribution $\mathcal{L}(Y|Y < \infty)$ in Case B, let $\tilde{Y} \equiv Y|Y < \infty$ denote the *total population size given extinction* and let

$$\tilde{\Phi}(s) \equiv \Phi_{\tilde{Y}}(s) = \mathrm{E}[s^Y|Y < \infty], \qquad 0 \le s \le 1, \tag{4.36}$$

denote its *conditional pgf.* The conditional range of $\tilde{Y}$ is $\{1, 2, \ldots\}$, so

$$\begin{aligned}
\tilde{\Phi}(s) &= \sum\nolimits_{k=1}^{\infty} \Pr[Y = k|Y < \infty]s^k \\
&= \sum\nolimits_{k=1}^{\infty} \frac{\Pr[Y = k]}{\Pr[Y < \infty]} s^k \\
&= \begin{cases} \frac{1}{u}\Phi(s), & 0 \le s < 1 \\ 1, & s = 1. \end{cases} \qquad (4.37)
\end{aligned}$$

Thus $\tilde{\Phi}(0) = 0$, $\tilde{\Phi}$ is nondecreasing and strictly convex on $[0, 1]$, and

$$\tilde{\Phi}(1-) = \frac{1}{u}\Phi(1-) = 1 = \tilde{\Phi}(1) \tag{4.38}$$

by (4.31), so $\tilde{\Phi}$ *has no jump at* $s = 1$, *that is,* $\tilde{\Phi}$ *is a bona fide pgf and determines the conditional distribution* $\mathcal{L}(\tilde{Y}) \equiv \mathcal{L}(Y|Y < \infty)$. Thus

$$\mathrm{E}(\tilde{Y}) = \tilde{\Phi}'(1) = \tilde{\Phi}'(1-) = \frac{1}{u}\Phi'(1-) \tag{4.39}$$

by (4.37). Also, from (4.32),

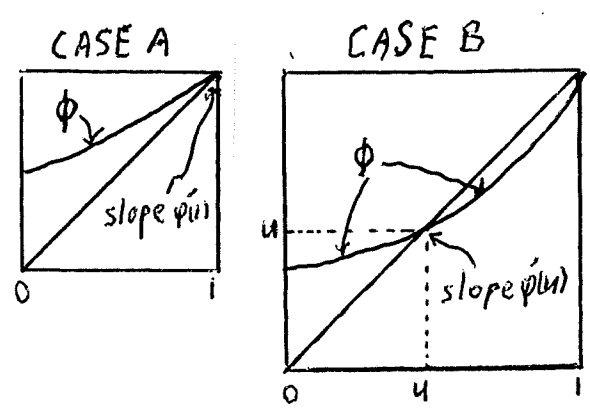
$$\Phi'(s) = s\,\phi'(\Phi(s))\Phi'(s) + \phi(\Phi(s)), \quad 0 \le s < 1$$

(same as (4.34)). Let $s \uparrow 1$ and apply (4.31) and (4.16) to obtain

$$\begin{aligned}
\Phi'(1-) &= \phi'(\Phi(1-))\Phi'(1-) + \phi(\Phi(1-)) \\
&= \phi'(u)\Phi'(1-) + \phi(u) \\
&= \phi'(u)\Phi'(1-) + u,
\end{aligned}$$

so $\quad \Phi'(1-) = \dfrac{u}{1 - \phi'(u)}$, hence

$$\mathrm{E}(\tilde{Y}) = \frac{1}{1 - \phi'(u)} \text{ by (4.39).} \tag{4.40}$$



Remark 4.3. By comparing (4.40) for $\mathrm{E}(\tilde{Y})$ to (4.35) for $\mathrm{E}(Y)$, we see that there is no "Case B2" for $\mathrm{E}(\tilde{Y})$: in Case B $\phi'(u) < 1$ always, whereas $\phi'(1) \le 1$ in Case A (see the preceding figures). □

Example 4.1b *(The linear case, continued).*
Here $p_0 + p_1 = 1$, $\phi(s) = p_0 + p_1 s$, and $\mu = \phi'(1) = p_1 < 1$, so Case A1 holds and (4.32) becomes

$$\Phi(s) = s[p_0 + p_1\Phi(s)].$$

This is easily solved for $\Phi(s)$ to yield

$$\begin{aligned}\Phi(s) = \frac{sp_0}{1 - sp_1} &= sp_0[1 + sp_1 + (sp_1)^2 + (sp_1)^3 + \cdots] \\ &= p_0 s + p_0 p_1 s^2 + p_0 p_1^2 s^3 + p_0 p_1^3 s^4 + \cdots,\end{aligned}$$

hence $\Pr[Y = k] = p_0 p_1^{k-1}$ for $k \ge 1$. Thus $Y \sim$ geometric(p_0), hence $\mathrm{E}(Y) = \frac{1}{p_0}$, which agrees with (4.35). Since $T = Y$ in this linear (no-growth) case, also $T \sim$ geometric(p_0). Thus, as we know, the waiting time to the first success has a geometric distribution. □

Example 4.4a *(The quadratic case).*
Here $p_0 + p_1 + p_2 = 1$ with $p_2 > 0$, so growth is possible, and $\phi(s) = p_0 + p_1 s + p_2 s^2$ is a quadratic function (e.g., Example 4.2 with $r = 2$). Thus $\mu \equiv \phi'(1) = p_1 + 2p_2 = 1 + p_2 - p_0$, so

$$\mu \text{ is } \begin{cases} \le 1 & \text{if } p_0 \ge p_2 \qquad \textit{(Case A)}; \\ > 1 & \text{if } 0 < p_0 < p_2 \quad \textit{(Case B)}; \\ > 1 & \text{if } 0 = p_0 < p_2 \quad \textit{(Case C)}. \end{cases} \tag{4.41}$$

Here (4.16) is equivalent to the quadratic equation

$$p_2 s^2 - (1 - p_1)s + p_0 = 0,$$

which, since $1 - p_1 = p_0 + p_2$, has the solutions

$$\frac{(p_0 + p_2) \pm \sqrt{(p_0 + p_2)^2 - 4p_0p_2}}{2p_2} = \frac{(p_0 + p_2) \pm (p_0 - p_2)}{2p_2} = \left\{1,\ \frac{p_0}{p_2}\right\}.$$

Thus the extinction probability is

$$u = \begin{cases} 1 & \text{if } p_0 \geq p_2 & \textit{(Case A)}; \\ \frac{p_0}{p_2} & \text{if } 0 < p_0 < p_2 & \textit{(Case B)}; \\ 0 & \text{if } 0 = p_0 < p_2 & \textit{(Case C)}. \end{cases} \tag{4.42}$$

Also, in Case B we have

$$\phi'(u) = p_1 + 2p_2 u = 1 - p_0 - p_2 + 2p_2 \tfrac{p_0}{p_2} = 1 + p_0 - p_2,$$

so it follows from (4.35) and (4.40) that

$$\begin{cases} \mathrm{E}(Y) = \frac{1}{1-\phi'(1)} = \frac{1}{p_0 - p_2} & \text{if } p_0 > p_2 & \textit{(Case A1)}; \\ \mathrm{E}(Y) = \frac{1}{1-\phi'(1)} = \infty & \text{if } p_0 = p_2 & \textit{(Case A2)}; \\ \mathrm{E}(\tilde{Y}) = \frac{1}{1-\phi'(u)} = \frac{1}{p_2 - p_0} & \text{if } 0 < p_0 < p_2 & \textit{(Case B)}. \end{cases} \tag{4.43}$$

Next, (4.37) becomes

$$\Phi(s) = s[p_0 + p_1 \Phi(s) + p_2 (\Phi(s))^2], \tag{4.44}$$

a quadratic equation in $\Phi(s)$ that can solved explicitly to obtain $\Phi(s)$ in Case A and thence $\tilde{\Phi}(s)$ in Case B. In principle these determine the distribution $\mathcal{L}(Y)$ in Case A and the conditional distribution $\mathcal{L}(\tilde{Y})$ in Case B. □

Exercise 4.2b *(Geometric offspring distribution, continued).*

(i) Find $\Phi \equiv \Phi_p$ and confirm that $\Phi_p(1-) = u$.

(ii) In Case A find $\mathrm{E}_p(Y)$ and apply the binomial expansion

$$(1-x)^\alpha = \sum\nolimits_{k=0}^{\infty} (-1)^k \binom{\alpha}{k} x^k, \qquad |x| < 1,$$

to $\Phi_p(s)$ to obtain the distribution of Y.

(iii) In Case B find $\tilde{\Phi}_p(s)$ and relate it to a Case A1 pgf, confirm that $\tilde{\Phi}_p(1-) = 1$, find $\mathrm{E}_p(\tilde{Y})$, and use this relation to Case A1 to obtain the conditional distribution $\mathcal{L}_p(\tilde{Y})$. □

Example 4.3b *(Poisson offspring distribution, continued).*
Here $\phi_\lambda(s) = e^{\lambda(s-1)}$, so (4.37) becomes

$$\Phi_\lambda(s) = s\, e^{\lambda(\Phi_\lambda(s)-1)},$$

which cannot be solved for $\Phi_\lambda(s)$ explicitly. However, $\phi'_\lambda(s) = \lambda e^{\lambda(s-1)}$, so

$$\begin{aligned} \phi'_\lambda(1) &= \lambda, \\ \phi'_\lambda(u) &= \lambda e^{\lambda(u-1)} = u\lambda, \end{aligned} \tag{4.45}$$

since u satisfies (4.26). Thus

$$\begin{cases} \mathrm{E}_\lambda(Y) = \frac{1}{1-\phi'_\lambda(1)} = \frac{1}{1-\lambda} & \text{if } \lambda < 1 \quad \textit{(Case A1)}; \\ \mathrm{E}_\lambda(Y) = \frac{1}{1-\phi'_\lambda(1)} = \infty & \text{if } \lambda = 1 \quad \textit{(Case A2)}; \\ \mathrm{E}_\lambda(\tilde{Y}) = \frac{1}{1-\phi'_\lambda(u)} = \frac{1}{1-u\lambda} & \text{if } \lambda > 1 \quad \textit{(Case B)}. \end{cases} \tag{4.46}$$

(From the figure at (4.40), note that $\phi'_\lambda(u) \equiv u\lambda < 1$ in Case B.) □

* * *

Now consider the distribution of the extinction time T: for $n \geq 1$,

$$\Pr[T \leq n] = \Pr[X_n = 0] = \phi_n(0), \tag{4.47}$$

so

$$\Pr[T = n] = \phi_n(0) - \phi_{n-1}(0), \tag{4.48}$$

from which the exact distribution of T can be obtained in principle.

Proposition 4.3. *In Case A,* $\mathrm{E}(T) \begin{cases} < \infty & \textit{if } \mu < 1 \quad \textit{(Case A1)}; \\ = \infty & \textit{if } \mu = 1 \quad \textit{(Case A2)}. \end{cases}$

Proof. *Case A1:* $\mathrm{E}(T) \leq \mathrm{E}(Y) = \frac{1}{1-\mu} < \infty$ by (4.35).

Case A2 (sketch): From (4.47),

$$\mathrm{E}(T) = \sum_{n=0}^{\infty} \Pr[T > n] = \sum_{n=0}^{\infty} [1 - \phi_n(0)], \tag{4.49}$$

where $\phi_0(0) \equiv 0$. Since $\phi(s) = 1$, $\phi'(1) = \mu = 1$, and $\phi(s) > s$ for $0 \le s < 1$ by strict convexity, the Taylor expansion of $\phi(s) - s$ at $s = 1$ is

$$\phi(s) - s = a(1-s)^2 + O((1-s)^3) \tag{4.50}$$

with $a = \frac{1}{2}\phi''(1)$ $(0 \le a \le \infty)$. If $0 < a < \infty$ then for s near 1,

$$1 - \phi(s) \approx (1-s) - a(1-s)^2,$$

so

$$\begin{aligned} \frac{1}{1-\phi(s)} &\approx \frac{1}{1-s}\left[\frac{1}{1-a(1-s)}\right] \\ &\approx \frac{1}{1-s}\left[1 + a(1-s)\right] \\ &= \frac{1}{1-s} + a. \end{aligned} \tag{4.51}$$

Replace s by $\phi_{n-1}(0)$, $\phi_{n-1}(0)$, $\ldots$, $\phi_1(0) \equiv p_0$ to obtain

$$\begin{aligned} \frac{1}{1-\phi_n(0)} &\approx \frac{1}{1-\phi_{n-1}(0)} + a \\ &\approx \frac{1}{1-\phi_{n-2}(0)} + 2a \\ &\vdots \\ &\approx \frac{1}{1-\phi_1(0)} + (n-1)a. \end{aligned} \tag{4.52}$$

Thus by (4.49)

$$\mathrm{E}\,(T) = \sum\nolimits_{n=0}^{\infty}[1 - \phi_n(0)] \approx \sum\nolimits_{n=0}^{\infty} \frac{1}{c + (n-1)a} = \infty. \qquad \square \tag{4.53}$$

Note: If $a = 0$, replace $(1-s)^2$ in (4.50) by $(1-s)^{2r}$ for some $r \ge 2$. Also note that $a < \infty$ iff $\mathrm{E}(\xi^2) < \infty$, since $\phi''(1) = \mathrm{E}[\xi(\xi - 1)]$ and $\mathrm{E}(\xi) = 1$. $\square$

Remark 4.4. Clearly $\mathrm{E}(T) = \infty$ in Cases B and C, since $\Pr[T = \infty] \equiv \Pr[Y = \infty] > 0$. $\square$

Exercise 4.2c* *(Geometric offspring distribution, continued).*

(i) Find an explicit expression for $\phi_n(s)$, the pgf of X_n. When $p < 1/2$, confirm that $\phi_n(0) \uparrow u$.

(ii) When $p \geq 1/2$, so $\mu \leq 1$ and $\Pr_p[T < \infty] = 1$, find the exact distribution of T, i.e., find $\Pr_p[T = n]$ for $n = 1, 2, \ldots$ and find $\mathrm{E}_p(T)$. (Consider the cases $p > 1/2$ and $p = 1/2$ separately.)

(iii) When $p < 1/2$, so $\mu > 1$ and $\Pr_p[T < \infty] < 1$, find the exact conditional distribution $\mathcal{L}_p(T|T < \infty)$, i.e., find $\Pr_p[T = n|T < \infty]$ for $n = 1, 2, \ldots$, and find $\mathrm{E}_p[T|T < \infty]$. □

4.5. Conditional behavior of a Case B GW process given extinction

Consider a Case B GW branching process $\{X_n\}$ with pgf ϕ. The two figures at (4.40), p.96, suggest a relation between the *conditional process*

$$\{\tilde{X}_n\} \equiv \{X_n\} \,|\, Y < \infty,$$

i.e., given extinction, and a Case A GW branching process with pgf $\tilde{\phi}$ similar to that portion of the original pgf ϕ that lies in the square $[0, u]^2$ (see figure for Case B). To convert this to a bona fide pgf $\tilde{\phi}$, re-scale the graph of ϕ by the factor $1/u$ to transform $[0, u]^2$ to the unit square $[0, 1]^2$. That is, for $0 \leq s \leq 1$ define

$$\tilde{\phi}(s) = \frac{\phi(us)}{u} = \sum\nolimits_{k=0}^{\infty} (p_k u^{k-1}) s^k \equiv \sum\nolimits_{k=0}^{\infty} \tilde{p}_k s^k. \tag{4.54}$$

Note that $\tilde{\phi}$ is a Case A1 pgf, e.g., $\tilde{\phi}'(1) = \phi'(u) < 1$.

As a first indication that this relation holds, first note that

$$\mathrm{E}(\tilde{Y}) = \frac{1}{1 - \phi'(u)} = \frac{1}{1 - \tilde{\phi}'(1)} \tag{4.55}$$

by (4.40) p.96, and (4.54), where, as in §4.4, $\tilde{Y}$ is the total population size of the $\{\tilde{X}_n\}$ process, then compare this to (4.35) p.95. In fact it will follow from Theorem 4.1 that *the conditional distribution $\mathcal{L}(\tilde{Y}) \equiv \mathcal{L}(Y|Y < \infty)$ is identical to the distribution of the total population size for the GW process determined by $\tilde{\phi}$.*

To establish this assertion directly, as in §4.4 let $\tilde{\Phi}$ denote the pgf of $\tilde{Y}$. Apply (4.37) p.96, (4.32) p.95, and (4.54) p.101 to obtain

$$\tilde{\Phi}(s) = \frac{\Phi(s)}{u} = \frac{s}{u} \cdot \phi\Big(u\frac{\Phi(s)}{u}\Big) = s \cdot \tilde{\phi}(\tilde{\Phi}(s)), \tag{4.56}$$

so $\tilde{\Phi}$ satisfies the functional equation (4.32) determined by $\tilde{\phi}$. By Proposition 4.2 p.95, the solution to this equation is unique, so $\tilde{\Phi}$ must be the pgf for the total population size for the GW process determined by $\tilde{\phi}$, as asserted.

Theorem 4.1. *For a Case B GW branching process $\{X_n\}$ with pgf ϕ, the distribution of the conditional process $\{\tilde{X}_n\} \equiv \{X_n\} \,|\, Y < \infty$ is the same as the distribution of a Case A1 GW process with generating pgf $\tilde{\phi}$.*

Short Proof. Under the condition $Y < \infty$ the family line of each individual eventually terminates, so the conditional process $\{\tilde{X}_n\}$ behaves like a Case A GW process. [See the Long Proof for a rigorous proof of this fact.] It remains to show that $\tilde{\phi}$ in (4.54) is the pgf for this process. But for this, simply note that for each $k = 0, 1, 2, \ldots$, Bayes formula yields

$$\begin{aligned} \Pr[\tilde{X}_1 = k] &\equiv \Pr[X_1 = k \,|\, Y < \infty] \\ &= \frac{\Pr[Y < \infty \,|\, X_1 = k]\Pr[X_1 = k]}{\Pr[Y < \infty]} \\ &= \frac{u^k p_k}{u} = \tilde{p}_k \end{aligned} \tag{4.57}$$

by (4.54), so the pgf of the conditional process $\{\tilde{X}_n\}$ coincides with $\tilde{\phi}$.

Long Proof. By Proposition 4.1(ii), it suffices to show that for each $n \geq 1$, the *joint pgf* $\phi_{\tilde{X}_1,\ldots,\tilde{X}_n}$ of $(\tilde{X}_1, \ldots, \tilde{X}_n)$ satisfies

$$\phi_{\tilde{X}_1,\ldots,\tilde{X}_n}(s_1, \ldots, s_n) = \tilde{\phi}(s_1\tilde{\phi}(s_2\tilde{\phi}(\ldots s_{n-1}\tilde{\phi}(s_n)\ldots))). \tag{4.58}$$

However, to introduce the idea of the proof, we first establish (4.58) when $s_1 = \cdots = s_{n-1} = 1$, that is, we first show that

$$\phi_{\tilde{X}_n}(1, \ldots, 1, s_n) = \tilde{\phi}(\tilde{\phi}(\tilde{\phi}(\ldots\tilde{\phi}(s_n)\ldots))) \equiv \tilde{\phi}_n(s_n) \tag{4.59}$$

for all $n \geq 1$, which is equivalent to the fact that $\tilde{X}_n$ has the same distribution as the nth stage of the GW process with pgf $\tilde{\phi}$. The proof of this fact proceeds by induction on n.

For $n = 1$, this is (4.57). Now assume that (4.59) holds for $\tilde{X}_n$. Then

$$\begin{aligned}
\mathrm{E}\,[s^{X_{n+1}} \mid Y < \infty] &= \mathrm{E}\,\{\mathrm{E}\,[s^{\xi_1^{(n)}+\cdots+\xi_{X_n}^{(n)}} \mid X_n, Y < \infty] \mid Y < \infty\} \\
&\overset{*}{=} \mathrm{E}\,\{(\tilde{\phi}(s))^{X_n} \mid Y < \infty\} && (4.60) \\
&\overset{**}{=} \tilde{\phi}_n(\tilde{\phi}(s)) && (4.61) \\
&= \tilde{\phi}_{n+1}(s),
\end{aligned}$$

so (4.59) holds for $\tilde{X}_{n+1}$ as required. The equality (*) follows because $Y < \infty$ implies that $\xi_1^{(n)}, \ldots, \xi_{X_n}^{(n)}$ are iid rvs with pgf $\tilde{\phi}$ rather than ϕ, since each individual in generation n must give rise to a family line that becomes extinct. Then (**) follows from the induction hypothesis (4.59) for $\tilde{X}_n$.

We turn to the proof of (4.58), which also proceeds by induction on n. The case $n = 1$ follows from (4.57). Assume that (4.58) holds for n and consider the case $n + 1$. Then

$$\begin{aligned}
&\phi_{\tilde{X}_1,\ldots,\tilde{X}_{n+1}}(s_1, \ldots, s_{n+1}) \\
&= \mathrm{E}\Big[\prod_{i=1}^{n+1} s_i^{X_i} \,\Big|\, Y < \infty\Big] \\
&= \mathrm{E}\Big\{\prod_{i=1}^{n} s_i^{X_i} \cdot \mathrm{E}\Big[s_{n+1}^{\xi_1^{(n)}+\cdots+\xi_{X_n}^{(n)}} \,\Big|\, X_1, \ldots, X_n, Y < \infty\Big] \,\Big|\, Y < \infty\Big\} \\
&\overset{*}{=} \mathrm{E}\Big\{\prod_{i=1}^{n} s_i^{X_i} \cdot (\tilde{\phi}(s_{n+1}))^{X_n} \,\Big|\, Y < \infty\Big\} && (4.62) \\
&= \mathrm{E}\Big\{\prod_{i=1}^{n-1} s_i^{X_i} \cdot (s_n\tilde{\phi}(s_{n+1}))^{X_n} \,\Big|\, Y < \infty\Big\} \\
&\equiv \phi_{\tilde{X}_1,\ldots,\tilde{X}_n}(s_1, \ldots, s_{n-1}, s_n\tilde{\phi}(s_{n+1})) \\
&\overset{**}{=} \tilde{\phi}(s_1\tilde{\phi}(s_2\tilde{\phi}(\ldots s_{n-1}\tilde{\phi}(s_n\tilde{\phi}(s_{n+1}))\ldots))),
\end{aligned}$$

which confirms (4.58) for the case $n + 1$. The equality (*) follows as in (4.60), while (**) follows from the induction hypothesis (4.58). □

Remark 4.5. From (4.54), the 2nd iterate of $\tilde{\phi}$ is

$$\tilde{\phi}_2(s) = \tfrac{1}{u}\phi(u\tfrac{1}{u}\phi(us)) = \tfrac{1}{u}\phi(\phi(us)) = \tfrac{1}{u}\phi_2(us), \tag{4.63}$$

so by induction, the nth iterate of $\tilde{\phi}$ and its first two derivatives are

$$\tilde{\phi}_n(s) = \tfrac{1}{u}\phi_n(us), \tag{4.64}$$

$$\tilde{\phi}'_n(s) = \phi'_n(us), \tag{4.65}$$

$$\tilde{\phi}''_n(s) = u\,\phi''_n(us). \tag{4.66}$$

Using the relation $\phi(u) = u$, it can be shown that (Exercise 4.3(ii))

$$\phi'_n(u) = [\phi'(u)]^n, \tag{4.67}$$

$$\phi''_n(u) = \phi''(u)[\phi'(u)]^{n-1}\left\{\frac{[\phi'(u)]^n - 1}{\phi'(u) - 1}\right\}, \tag{4.68}$$

By (4.59), (4.3), (4.14), (4.65), and (4.67),

$$\mathrm{E}(\tilde{X}_n) = \tilde{\phi}'_n(1) = \phi'_n(u) = [\phi'(u)]^n < [\phi'(1)]^n \equiv \mu^n \equiv \mathrm{E}(X_n), \tag{4.69}$$

by the strict monotonicity of $\phi'(\cdot)$. In fact, setting $\tilde{\mu} = \phi'(u) \equiv \mathrm{E}(\tilde{X}_1)$,

$$\mathrm{E}(\tilde{X}_n) = \tilde{\mu}^n \ll \mu^n \equiv \mathrm{E}(X_n) \qquad \text{for large } n, \tag{4.70}$$

since $\tilde{\mu} < 1 < \mu$ in Case B. Furthermore, by (4.59) and (4.5) for $\tilde{X}_n$,

$$\mathrm{Var}(\tilde{X}_n) = \tilde{\sigma}^2 \cdot \tilde{\mu}^{n-1}\left(\frac{1-\tilde{\mu}^n}{1-\tilde{\mu}}\right), \tag{4.71}$$

where

$$\begin{aligned} \tilde{\sigma}^2 = \mathrm{Var}(\tilde{X}_1) &= \tilde{\phi}''(1) + \tilde{\phi}'(1) - [\tilde{\phi}'(1)]^2 && \text{[verify]} \\ &= u\,\phi''(u) + \phi'(u) - [\phi'(u)]^2 && \text{[by (4.66)].} \\ &\equiv u\,\phi''(u) + \tilde{\mu} - \tilde{\mu}^2. \end{aligned} \tag{4.72}$$

□

Exercise 4.3. (i) Derive (4.67) and (4.68).

(ii) Use (4.66) with $s = 1$ and (4.68) to re-derive (4.71). □

Example 4.4b *(The quadratic case, continued).* Recall that

$$\phi(s) \equiv \phi_{p_0,p_2}(s) = p_0 + p_1 s + p_2 s^2 \tag{4.73}$$

and Case B obtains iff $0 < p_0 < p_2$. In this case the extinction probability $u = p_0/p_2$ and from (4.54) we obtain the interesting relation

$$\begin{aligned}\tilde{\phi}_{p_0,p_2}(s) &= \tfrac{1}{u}\,\phi_{p_0,p_2}(us)\\ &= \tfrac{p_2}{p_0}[p_0 + p_1(\tfrac{p_0}{p_2}s) + p_2(\tfrac{p_0}{p_2}s)^2]\\ &= p_2 + p_1 s + p_0 s^2\\ &= \phi_{p_2,p_0}(s). \qquad (4.74)\end{aligned}$$

By Theorem 4.1, therefore, conditional on extinction the Case B GW process with pgf ϕ_{p_0,p_2} evolves like the unconditional Case A1 GW process with pgf ϕ_{p_2,p_0}. [Interpret? Relate to (4.43).] Furthermore,

$$\tilde{\mu} \equiv \phi'_{p_0,p_2}(u) = p_1 + 2p_2(\tfrac{p_0}{p_2}) = 1 + p_0 - p_2 \ (< 1), \tag{4.75}$$

so it follows from (4.69) that

$$\mathrm{E}_{p_0,p_2}(\tilde{X}_n) = (1 + p_0 - p_2)^n = \mathrm{E}_{p_2,p_0}(X_n),$$

which, by (4.59), agrees with (4.74). Also, by (4.71) and (4.72) [verify],

$$\mathrm{Var}_{p_0,p_2}(\tilde{X}_n) = [p_0 + p_2 - (p_2 - p_0)^2] \cdot \tilde{\mu}^{n-1}\left(\frac{\tilde{\mu}^n - 1}{\tilde{\mu} - 1}\right). \qquad \square$$

Exercise 4.2d *(Geometric offspring distribution, continued).*
(i) As in Exercise 4.2a, let $\xi \sim$ geometric(p), that is, $p_k = pq^k$ for $k = 0, 1, 2, \ldots$ $(0 < p < 1)$, and let $\phi_p(s)$ denote the pgf of ξ. For those values of p such that Case B holds (found in Exercise 4.2a), find $\tilde{\phi}_p(s)$ and show that it remains a geometric pgf, now Case A1. Relate this to Exercise 4.2b(iii).
(ii) Find $\mathrm{E}_p(\tilde{X}_n)$ and $\mathrm{Var}_p(\tilde{X}_n)$. $\square$

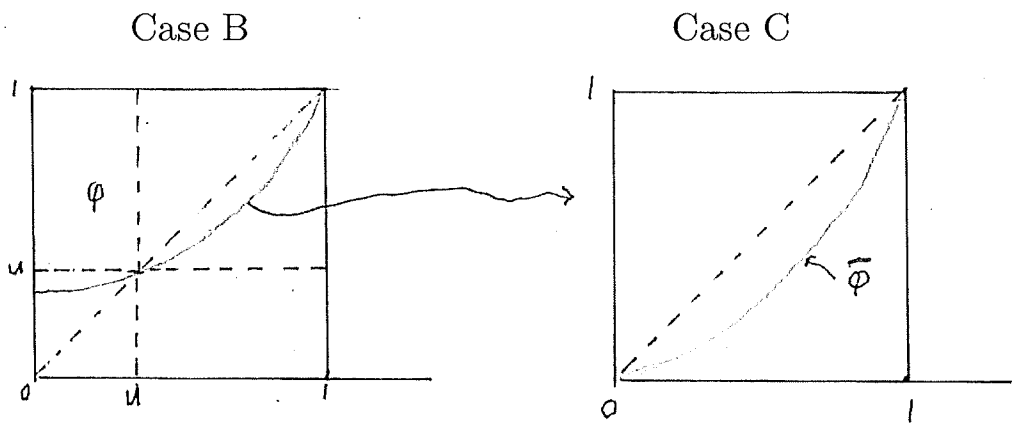
Exercise 4.4a *(Poisson offspring distribution, continued).*
(i) As in Example 4.3a, let $\xi \sim$ Poisson(λ), that is, $p_k = e^{-\lambda}\lambda^k/k!$ for $k = 0, 1, \ldots$, and let $\phi(s) \equiv \phi_\lambda(s) = e^{\lambda(s-1)}$ be the pgf of ξ. It was shown

that Case B holds iff $\lambda > 1$. In this case find $\tilde{\phi}_\lambda(s)$ and show that it remains a Poisson pgf, now Case A1. Relate this to (4.46) in Example 4.3b.

(ii) Find $\mathrm{E}_\lambda(\tilde{X}_n)$ and $\mathrm{Var}_\lambda(\tilde{X}_n)$. □

4.6. Conditional behavior of a Case B GW process given explosion

Again consider a Case B GW branching process $\{X_n\}$ with pgf ϕ. Analogous to the results in §4.5, the two figures below suggest a relation between the conditional process $\{X_n\}$ given explosion, i.e., given $Y = \infty$, and an unconditional Case C process with pgf $\bar{\phi}$ similar to that portion of the original pgf ϕ that lies in the square $[u, 1]^2$, but re-scaled to lie in $[0, 1]^2$.



This re-scaled function $\bar{\phi}(s)$ is given by

$$\bar{\phi}(s) = \frac{\phi((1-u)s + u) - u}{1-u}, \qquad 0 \le s \le 1. \tag{4.76}$$

Note that $\bar{\phi}(0) = 0$ (since $\phi(u) = u$) and $\bar{\phi}(1) = 1$, so if $\bar{\phi}$ is a pgf (see Theorem 4.2), it must be a Case C pgf. Furthermore,

$$\bar{\mu} \equiv \bar{\phi}'(1) = \phi'(1) \equiv \mu > 1. \tag{4.77}$$

However, the relation suggested above is *not valid.* Unlike the case of extinction, where the family line of *every* individual dies out, in the case of explosion only some family lines survive forever, while the rest die out which cannot occur in Case C. Instead we shall show that $\bar{\phi}$ is the pgf of the *partial process* $\{\breve{X}_n\}$ *conditioned on the event* $\{\breve{X}_0 = 1\} \equiv \{Y = \infty\}$, where $\breve{X}_n$ is *the number of individuals at time* n *whose family lines never die out.*

First note that $\breve{X}_0 = 0$ or 1 with probabilities u and $1-u$ respectively. We will show that the conditional process $\{\bar{X}_n\} \equiv \{\breve{X}_n\} \mid \breve{X}_0 = 1$ behaves like a Case C process with pgf $\bar{\phi}$.

First note the following three (unconditional) properties of $\{\breve{X}_n\}$:

(a) $\breve{X}_n \le X_n$ for $n = 0, 1, 2, \ldots$;

(b) $\breve{X}_0 = 0 \Rightarrow \breve{X}_1 = \breve{X}_2 = \cdots = 0$;

(c) $\breve{X}_0 = 1 \Rightarrow 1 \le \breve{X}_1 \le \breve{X}_2 \le \cdots$.

Theorem 4.2. *For a Case B GW branching process $\{X_n\}$ with pgf ϕ, the (conditional) partial process $\{\bar{X}_n\} \equiv \{\breve{X}_n\} \mid \breve{X}_0 = 1$ behaves like a Case C GW process with generating pgf $\bar{\phi}$.*

Proof. Because $\{X_n\}$ is a GW process, it follows directly that the partial process $\{\breve{X}_n\}$ also is a GW process: at each time n we may simply ignore those individuals whose family lines eventually die out, because neither they nor their descendants contribute to subsequent generations $\breve{X}_{n+1}, \breve{X}_{n+2}, \ldots$ [illustrate with a diagram]. It remains to show that the generating pgf for the conditional process $\{\bar{X}_n\} \equiv \{\breve{X}_n\} \mid \breve{X}_0 = 1$ is given by $\bar{\phi}$ in (4.76).

It suffices to show that

$$\mathrm{E}(s^{\bar{X}_1}) \equiv \mathrm{E}[s^{\breve{X}_1} \mid \breve{X}_0 = 1] = \frac{\phi((1-u)s+u) - u}{1-u} \equiv \bar{\phi}(s). \tag{4.78}$$

First, $\bar{p}_0 \equiv \Pr[\bar{X}_1 = 0] = 0$ by (c), while for $k \ge 1$, $\bar{p}_k \equiv \Pr[\bar{X}_1 = k]$ satisfies

$$\begin{aligned}
\bar{p}_k &= \Pr[\bar{X}_1 = k] \\
&= \Pr[\breve{X}_1 = k]/\Pr[\breve{X}_0 = 1] \\
&= \frac{1}{1-u} \sum_{l \ge k} \Pr[\breve{X}_1 = k \,|\, X_1 = l] \Pr[X_1 = l] \quad \text{[by (a)]} \\
&= \frac{1}{1-u} \sum_{l \ge k} \Pr[\, k \text{ of } l \text{ live forever, } l-k \text{ go extinct} \,|\, X_1 = l]\, p_l \\
&= \frac{1}{1-u} \sum_{l \ge k} \binom{l}{k} (1-u)^k u^{l-k} p_l, \qquad (4.79)
\end{aligned}$$

where $p_l = \Pr[X_1 = l]$. Thus

$$\begin{aligned}
\mathrm{E}[s^{\bar{X}_1}] &\equiv \sum_{k\geq 0} \bar{p}_k s^k \\
&= \frac{1}{1-u} \sum_{k\geq 1} \Big[\sum_{l\geq k} p_l \binom{l}{k} (1-u)^k u^{l-k} \Big] s^k \\
&= \frac{1}{1-u} \sum_{l\geq 1} p_l \sum_{k=1}^{l} \binom{l}{k} [(1-u)s]^k u^{l-k} \\
&= \frac{1}{1-u} \sum_{l\geq 1} p_l \Big[\sum_{k=0}^{l} \binom{l}{k} [(1-u)s]^k u^{l-k} - u^l \Big] \\
&= \frac{1}{1-u} \sum_{l\geq 1} p_l \Big[((1-u)s+u)^l - u^l \Big] \\
&= \frac{1}{1-u} \sum_{l\geq 0} p_l \Big[((1-u)s+u)^l - u^l \Big] \\
&= \frac{\phi((1-u)s+u) - \phi(u)}{1-u}
\end{aligned}$$

which confirms (4.78), since $\phi(u) = u$. □

Remark 4.6. The distribution $\mathcal{L}(\bar{X}_n) \equiv \mathcal{L}(\breve{X}_n \mid \breve{X}_0 = 1)$ is determined by the nth iterate $\bar{\phi}_n$ of $\bar{\phi}$. The 2nd iterate is

$$\begin{aligned}
\bar{\phi}_2(s) \equiv \bar{\phi}(\bar{\phi}(s)) &= \frac{\phi\left\{(1-u)\left[\frac{\phi((1-u)s+u)-u}{1-u}\right] + u\right\} - u}{1-u} \\
&= \frac{\phi(\phi((1-u)s+u)) - u}{1-u} \\
&= \frac{\phi_2((1-u)s+u) - u}{1-u}.
\end{aligned}$$

By induction, the nth iterate of $\bar{\phi}$ is

$$\bar{\phi}_n(s) = \frac{\phi_n((1-u)s+u) - u}{1-u}, \tag{4.80}$$

from which the (conditional) mean of $\bar{X}_n$ is found to be

$$\mathrm{E}(\bar{X}_n) = \bar{\phi}_n'(1) = \phi_n'(1) = \mathrm{E}(X_n) \equiv \mu^n. \tag{4.81}$$

Furthermore, as in (4.71)-(4.72), the (conditional) variance of $\bar{X}_n$ is

$$\begin{aligned}
\mathrm{Var}(\bar{X}_n) &= \bar{\sigma}^2 \cdot \bar{\mu}^{n-1}\Big(\frac{\bar{\mu}^n - 1}{\bar{\mu} - 1}\Big) \\
&= \bar{\sigma}^2 \cdot \bar{\mu}^{n-1}(1 + \bar{\mu} + \cdots + \bar{\mu}^{n-1}) && (4.82)\\
&= (\sigma^2 - u\phi''(1)) \cdot \mu^{n-1}(1 + \mu + \cdots + \mu^{n-1}) \\
&< \sigma^2 \cdot \mu^{n-1}(1 + \mu + \cdots + \mu^{n-1}) && (4.83)\\
&= \mathrm{Var}(X_n),
\end{aligned}$$

where $\bar{\mu} \equiv \mathrm{E}(\bar{X}_1) = \bar{\phi}'(1) = \mu$ by (4.77) and where

$$\begin{aligned}
\bar{\sigma}^2 \equiv \mathrm{Var}(\bar{X}_1) &= \bar{\phi}''(1) + \bar{\phi}'(1) - [\bar{\phi}'(1)]^2 && \text{[verify]} \\
&= (1-u)\phi''(1) + \mu - \mu^2 && \text{[by (4.76)]} \\
&= \sigma^2 - u\phi''(1). && \square
\end{aligned}$$

Remark 4.7. From (4.81) the unconditional expectation of $\breve{X}_n$ is

$$\begin{aligned}
\mathrm{E}(\breve{X}_n) &= \mathrm{E}[\breve{X}_n \mid \breve{X}_0 = 1]\mathrm{Pr}[\breve{X}_0 = 1] + \underbrace{\mathrm{E}[\breve{X}_n \mid \breve{X}_0 = 0]}_{=0} \mathrm{Pr}[\breve{X}_0 = 0] \\
&= \mathrm{E}(\bar{X}_n)(1-u) = \mathrm{E}(X_n)(1-u) = \mu^n(1-u). && (4.84)
\end{aligned}$$

Thus unconditionally, $\mathrm{E}(\breve{X}_n) < \mathrm{E}(X_n)$ as expected, since $\breve{X}_n \le X_n$. $\square$

Remark 4.8. Because the (conditional) process $\{\bar{X}_n\}$ is a Case C GW process, it must hold that $\Pr[\bar{X}_n \to \infty] = 1$. $\square$

Example 4.4c *(The quadratic case, continued).* Recall that

$$\phi_{p_0,p_2}(s) = p_0 + p_1 s + p_2 s^2$$

and Case B obtains iff $0 < p_0 < p_2$. In this case the extinction probability $u = p_0/p_2$ and from (4.76) we find that

$$\begin{aligned}
\bar{\phi}_{p_0,p_2}(s) &= \tfrac{1}{1-u}\,\phi_{p_0,p_2}((1-u)s+u) \\
&= \tfrac{1}{1-\frac{p_0}{p_2}}[p_0 + p_1((1-\tfrac{p_0}{p_2})s + \tfrac{p_0}{p_2}) + p_2((1-\tfrac{p_0}{p_2})s + \tfrac{p_0}{p_2})^2] \\
&= (\text{check algebra})\cdots \\
&= 0 + (1+p_0-p_2)s + (p_2-p_0)s^2 \qquad (4.85) \\
&= s\cdot[(1+p_0-p_2) + (p_2-p_0)s] \\
&\equiv \eta(s)\cdot\psi(s), \qquad (4.86)
\end{aligned}$$

where we used the fact that $p_1 = 1-p_0-p_2$. Here (4.85) shows that $\bar{\phi}_{p_0,p_2} = \phi_{0,p_2-p_0}$, a quadratic-case Case C pgf. Each individual in the partial process $\{\bar{X}_n\}$ is replaced by either 1 individual with probability $1-(p_2-p_0)$ or by 2 individuals with probability p_2-p_0. By Theorem 4.2, conditional on explosion the partial process $\{\bar{X}_n\}$ evolves like the supercritical Case C GW process with pgf ϕ_{0,p_2-p_0}.

In fact, (4.86) shows that $\bar{\phi}_{p_0,p_2}$ is the product of the two pgfs η and ψ, where $\eta(s) \equiv s$ is the pgf of a random variable degenerate at 1 and $\psi(s)$ is the pgf of a Bernoulli(p_2-p_0) rv U. (Recall the linear case in Example 4.1a.) Thus, if we let $\bar{\xi}$ denote the offspring rv of the process $\{\bar{X}_n\}$, then $\bar{\xi} \stackrel{d}{=} 1+U$, because the pgf of a sum of independent rvs is the product of their pgfs [verify]. Note that ψ is a Case A1 pgf.

Lastly, note that

$$\bar{\phi}'_{p_0,p_2}(1) = (1+p_0-p_2) + 2(p_2-p_0) = 1+p_1-p_0 = \phi'_{p_0,p_2}(1),$$

which agrees with (4.81) for $n=1$. □

Exercise 4.2e *(Geometric offspring distribution, continued).*
As in Exercise 4.2a, let $\xi \sim$ geometric(p), that is, $p_k = pq^k$ for $k = 0,1,2,\ldots$ $(0<p<1)$, and let ϕ_p denote the pgf of ξ. For those values of p such that Case B holds (found in Exercise 4.2a), show that $\bar{\phi}_p$ is a Case C pgf. How is it related to the geometric pgf ϕ_u? Find $\bar{p}_k$ for $k = 0,1,2,\ldots$. □

Exercise 4.4b *(Poisson offspring distribution, continued).*
As in Example 4.3a, let $\xi \sim \text{Poisson}(\lambda)$, that is, $p_k = e^{-\lambda}\lambda^k/k!$ for $k = 0, 1, \ldots$, so $\phi_\lambda(s) = e^{\lambda(s-1)}$ is the pgf of ξ. It was shown that Case B holds iff $\lambda > 1$. Find $\bar{\phi}_\lambda(s)$ and show that it is the conditional pgf of a Poisson rv K given $K \geq 1$, which is a Case C pgf. Find $\bar{p}_k$ for $k = 0, 1, 2, \ldots$. □

4.7. Stochastic comparisons among conditional Case B processes

We now consider possible stochastic orderings among $\{X_n\}$ and the following four conditional processes based on the original Case B GW branching process $\{X_n\}$:

$$\begin{aligned}
\{\tilde{X}_n\} &\equiv \{X_n\} \,|\, Y < \infty; && \text{(a GW process)}\\
\{\bar{X}_n\} &\equiv \{\breve{X}_n\} | \breve{X}_0 = 1; && \text{(a GW process)}\\
\{\dot{X}_n\} &\equiv \{X_n\} \,|\, Y = \infty; && \text{(not a GW process?)}\\
\{\check{X}_n\} &\equiv \{X_n\} \,|\, \text{no individual dies} && \text{(a GW process).}
\end{aligned}$$

Recall that $\{\breve{X}_0 = 1\} = \{Y = \infty\}$, and that some individuals still may die even though $Y = \infty$, so $\{\dot{X}_n\}$ may not be a GW process (see (7) below).

Because $p_0 > 0$ in our Case B process, the event {no individual dies} is in fact null, so cannot actually be considered as a conditioning event. Instead, the process $\{\check{X}_n\}$ is defined to be the modified GW branching process with offspring distribution $\check{\mathbf{p}} \equiv (\check{p}_0, \check{p}_1, \check{p}_2, \ldots)$, where $\check{p}_0 = 0$ and

$$\check{p}_k = \frac{p_k}{1 - p_0} \qquad \text{for } k \geq 1. \tag{4.87}$$

That is, the modified offspring rv $\check{\xi}$ has the conditional distribution of the original ξ given that $\xi \geq 1$, i.e., death is not allowed. Thus $\{\check{X}_n\}$ is a Case C GW process with pgf

$$\check{\phi}(s) = \frac{\phi(s) - p_0}{1 - p_0}. \tag{4.88}$$

Note that $\{\check{X}_n\}$ behaves differently than $\{\dot{X}_n\}$ because, as already noted, some individuals in the original process may die even though the population

explodes, i.e., even though $Y = \infty$. Also, $\{\check{X}_n\}$ behaves differently than the partial process $\{\bar{X}_n\}$: for example, by (4.88) and (4.81),

$$\begin{aligned} \mathrm{E}(\check{X}_n) = \check{\phi}_n'(1) = [\check{\phi}'(1)]^n &= \left[\tfrac{\phi'(1)}{1-p_0}\right]^n \\ &\gg [\phi'(1)]^n = [\bar{\phi}'(1)]^n = \bar{\phi}_n'(1) = \mathrm{E}(\bar{X}_n) \end{aligned} \tag{4.89}$$

since $p_0 > 0$ in Case B.

Although $\{\dot{X}_n\}$ is not a GW process, we can obtain the pgf $\dot{\phi}_{[n]}$ of $\dot{X}_n$ and thence its moments by the following mixture representation for X_n:

$$X_n = \begin{cases} \tilde{X}_n & \text{with probability } u, \\ \dot{X}_n & \text{with probability } 1-u. \end{cases} \tag{4.90}$$

Therefore

$$\begin{aligned} \phi_n(s) = \mathrm{E}(s^{X_n}) &= \mathrm{E}[s^{X_n} \mid Y < \infty]\,\Pr[Y < \infty] + \mathrm{E}[s^{X_n} \mid Y = \infty]\,\Pr[Y = \infty] \\ &= \mathrm{E}(s^{\tilde{X}_n})\,u + \mathrm{E}(s^{\dot{X}_n})\,(1-u) \\ &= \tilde{\phi}_n(s)\,u + \dot{\phi}_{[n]}(s)\,(1-u), \end{aligned} \tag{4.91}$$

so

$$\dot{\phi}_{[n]}(s) = \frac{\phi_n(s) - \tilde{\phi}_n(s)\,u}{1-u} = \frac{\phi_n(s) - \phi_n(us)}{1-u} \tag{4.92}$$

by (4.64). Similarly, by (4.69), (4.70), and (4.81),

$$\mathrm{E}(\dot{X}_n) = \frac{\mathrm{E}(X_n) - \mathrm{E}(\tilde{X}_n)\,u}{1-u} = \frac{\mu^n - \tilde{\mu}^n\,u}{1-u} > \mu^n = \mathrm{E}(X_n) = \mathrm{E}(\bar{X}_n). \tag{4.93}$$

(Note that in Case B, $\mu > 1 > \tilde{\mu}$.) Also, by (4.89) and (4.93),

$$\mathrm{E}(\check{X}_n) = \left(\frac{\mu}{1-p_0}\right)^n \gg \frac{\mu^n}{1-u} > \frac{\mu^n - \tilde{\mu}^n\,u}{1-u} = \mathrm{E}(\dot{X}_n), \tag{4.94}$$

where the strong inequality holds when $n \gg \frac{\log(1-u)}{\log(1-p_0)}$ (which ratio is > 1, because $0 < p_0 < u < 1$ in Case B).

Thus far we have established the following strict inequalities:

$$\mathrm{E}(\tilde{X}_n) \overset{(4.69)}{<} \mathrm{E}(X_n) \overset{(4.81)}{=} \mathrm{E}(\breve{X}_n) \begin{cases} \overset{(4.93)}{<} \mathrm{E}(\dot{X}_n) \\ \overset{(4.89)}{<} \mathrm{E}(\breve{X}_n), \end{cases} \tag{4.95}$$

and

$$\mathrm{E}(\dot{X}_n) \overset{(4.94)}{<} \mathrm{E}(\breve{X}_n) \quad \text{for sufficiently large } n. \tag{4.96}$$

This suggests that the following six strict stochastic orderings may hold among the original GW process $\{X_n\}$ and the four conditional processes based upon it:

$$\tilde{X}_n \begin{cases} \overset{(1)}{<}_{\mathrm{st}} \; X_n \begin{cases} \overset{(2)}{<}_{\mathrm{st}} \; \dot{X}_n \\ \overset{(3)}{<}_{\mathrm{st}} \; \breve{X}_n \end{cases} \\ \overset{(4)}{<}_{\mathrm{st}} \; \breve{X}_n \begin{cases} \overset{(5)}{<}_{\mathrm{st}} \; \dot{X}_n \\ \overset{(6)}{<}_{\mathrm{st}} \; \breve{X}_n, \end{cases} \end{cases} \tag{4.97}$$

and possibly a seventh ordering :

$$\dot{X}_n \overset{(7)}{\leq}_{\mathrm{st}} \breve{X}_n. \tag{4.98}$$

We now examine these seven possible stochastic orderings.

The orderings (1), (2), and (3) are intuitively obvious by the definitions of $\tilde{X}_n$ and $\dot{X}_n$ [recall]. They can be proved as follows:

(1) is valid:

The processes $\{\tilde{X}_n\}$ and $\{X_n\}$ are GW branching processes with offspring distributions $\tilde{\xi} \sim (\tilde{p}_0, \tilde{p}_1, \ldots)$ and $\xi \sim (p_0, p_1, \ldots)$ respectively. By the fundamental reproductive property (4.2) of a GW process and induction, to establish the stochastic ordering (1) it suffices to show that $\tilde{X}_1 <_{\mathrm{st}} X_1$, i.e., $\tilde{\xi} <_{\mathrm{st}} \xi$. This will follow from monotone likelihood ratio (MLR) theory [explain!] if it can be shown that the pair $(\tilde{\xi}, \xi)$ has a strictly increasing likelihood ratio, that is, the ratio $\frac{p_k}{\tilde{p}_k}$ is strictly increasing in k for those k s.t. $p_k > 0$. But this ratio equals $1/u^{k-1}$ by (4.57), hence is strictly increasing since $0 < u < 1$. $\square$

(2) is valid:

We might (incorrectly!) argue as in (1) as follows: Both processes $\{X_n\}$ and $\{\dot{X}_n\}$ are GW branching processes with offspring distributions $\xi \sim (p_0, p_1, \ldots)$ and $\dot{\xi} \sim (\dot{p}_0, \dot{p}_1, \ldots)$ respectively. By the fundamental reproductive property (4.2), to establish the ordering (2) it suffices to show that $\xi <_{\text{st}} \dot{\xi}$. But $\dot{p}_0 \equiv \Pr[\dot{X}_1 = 0] = 0$, while for $k \geq 1$, (4.92) with $n = 1$ implies that

$$\dot{p}_k = p_k \left(\frac{1 - u^k}{1 - u} \right) \qquad \text{[verify!]} \tag{4.99}$$

Thus, here too the LR $\frac{\dot{p}_k}{p_k}$ is strictly increasing in k for those k s.t. $p_k > 0$, hence as with (1), the ordering (2) follows from MLR theory.

However this argument is incorrect, or at least incomplete, because it is not obvious[8] that $\{\dot{X}_n\} \equiv \{X_n\} \,|\, Y = \infty$ is a GW process: even though $Y = \infty$, the family lines of some individuals may go extinct so these individuals would not evolve according to $\dot{\xi}$. Therefore we cannot appeal to the fundamental reproductive property (4.2). Instead, however, (2) can be established simply by means of the mixture representation (4.90): for any integer $k \geq 1$,

$$\begin{aligned} \Pr[X_n \geq k] &= \Pr[\tilde{X}_n \geq k]\, u + \Pr[\dot{X}_n \geq k]\,(1 - u) \\ &< \Pr[X_n \geq k]\, u + \Pr[\dot{X}_n \geq k]\,(1 - u) \end{aligned} \tag{4.100}$$

by (1), so

$$\Pr[X_n \geq k] < \Pr[\dot{X}_n \geq k], \tag{4.101}$$

hence (2) holds. □

(3) is valid:

The processes $\{X_n\}$ and $\{\check{X}_n\}$ are GW branching processes with offspring distributions $\xi \sim (p_0, p_1, \ldots)$ and $\check{\xi} \sim (\check{p}_0, \check{p}_1, \ldots)$ respectively. Thus by property (4.2), to establish (3) it suffices to show that $\xi <_{\text{st}} \check{\xi}$. But this follows directly from (4.87) since $p_0 > 0$. □

[8] In fact $\{\dot{X}_n\}$ is not a GW process – see (7) below.

We now examine the possible stochastic inequalities (4), (5), and (6), each of which involves $\breve{X}_n$.

(4) is valid for $u \le \frac{1}{2}$ but not necessarily for $u > \frac{1}{2}$:
By (4.60) with $k = 0$,

$$\Pr[\tilde{X}_1 \ge 1] = 1 - \frac{p_0}{u} < 1 = \Pr[\breve{X}_1 \ge 1] \tag{4.102}$$

for all values of u. Also, it follows from (4.79) and (4.60) that for $r \ge 2$,

$$\begin{aligned}
\Pr[\breve{X}_1 \ge r] &= \sum_{k \ge r} \breve{p}_k \\
&= \frac{1}{1-u} \sum_{k \ge r} \sum_{l \ge k} p_l \binom{l}{k} (1-u)^k u^{l-k} \\
&= \frac{1}{1-u} \sum_{l \ge r} p_l\, u^l \sum_{k=r}^{l} \binom{l}{k} \left(\frac{1-u}{u}\right)^k, \\
&= \frac{1}{1-u} \left(\frac{1-u}{u}\right)^r \sum_{l \ge r} p_l\, u^l \underbrace{\sum_{m=0}^{l-r} \binom{l}{m+r} \left(\frac{1-u}{u}\right)^m}_{\ge\, l-r+1} \\
&\ge \frac{1}{u} \sum_{l \ge r} p_l\, u^l \qquad (4.103) \\
&= \sum_{l \ge r} \tilde{p}_l \\
&= \Pr[\tilde{X}_1 \ge r],
\end{aligned}$$

where the inequality (4.103) holds provided that $u \le \frac{1}{2}$. Thus, in this case $\tilde{X}_1 <_{\text{st}} \breve{X}_1$ and therefore (4) holds (i.e., $\tilde{X}_n <_{\text{st}} \breve{X}_n$), since both $\{\tilde{X}_n\}$ and $\{\breve{X}_n\}$ are GW processes hence satisfy (4.2).

In the geometric case (Exercise 4.2f), in fact (4) holds for all $n \ge 1$ when $u \le \frac{\sqrt{5}-1}{2} \approx 0.618$ and for sufficiently large n (depending on u) when $u > \frac{\sqrt{5}-1}{2}$. In the quadratic case (Example 4.4d(4) below), however, $\tilde{X}_n \not\le_{\text{st}} \breve{X}_n$ for all $n \ge 1$ when $u > \frac{1}{2}$, so (4) fails in this case. This may be somewhat surprising in view of the facts that in general:

• the conditioning events for $\tilde{X}_n$ and $\breve{X}_n$ are $\{Y < \infty\}$ and $\{Y = \infty\}$ respectively;

• $\tilde{X}_n \to 0$ (because $\{\tilde{X}_n\}$ is a Case A GW process) while $\breve{X}_n \to \infty$ (because $\{\breve{X}_n\}$ is a Case C GW process);

• $E(\tilde{X}_n) \ll \mathrm{E}(\breve{X}_n)$ for large n (by (4.70) and (4.81)). □

(5) is valid:

Clearly $\breve{X}_n \leq_{\mathrm{st}} \dot{X}_n$ because $\breve{X}_n \leq \dot{X}_n$ on their common conditioning event $\{\breve{X}_0 = 1\} = \{Y = \infty\}$. But

$$\begin{aligned} \Pr[\breve{X}_n < \dot{X}_n] &= \mathrm{E}\{\Pr[\breve{X}_n < \dot{X}_n \mid \dot{X}_n]\} \\ &= \mathrm{E}\{1 - \Pr[\breve{X}_n = \dot{X}_n \mid \dot{X}_n]\} \\ &= \mathrm{E}\{1 - (1-u)^{\dot{X}_n}\} \\ &> 0 \end{aligned} \tag{4.104}$$

since $u > 0$ and $\dot{X}_n \geq 1$, hence $\breve{X}_n <_{\mathrm{st}} \dot{X}_n$. □

(6) is not necessarily valid:

Example 4.4d(6) and Exercises 4.2f and 4.4c show that (6) holds in the quadratic, geometric, and Poisson cases previously considered. However, the cubic Example 4.5 below shows that $\breve{X}_n <_{\mathrm{st}} \breve{X}_n$ does not hold in general, despite the fact that in general, $\mathrm{E}(\breve{X}_n) \ll \mathrm{E}(\breve{X}_n)$ for large n by(4.89). □

(7) is not necessarily valid:

Exercise 4.2f shows that (7) holds for sufficiently large n in the geometric case. In the quadratic Example 4.4d(7) below, however, $\dot{X}_n \not\leq_{\mathrm{st}} \breve{X}_n$ for all n, so (7) does not hold in general. This may be somewhat surprising in view of the facts that $\mathrm{E}(\dot{X}_n) \ll \mathrm{E}(\breve{X}_n)$ for large n by (4.94) and that no deaths are allowed for the process $\{\breve{X}_n\}$.

In fact for $n = 1$, $\breve{X}_1 <_{\mathrm{st}} \dot{X}_1$ *always,*[9] seen as follows. Clearly $\dot{p}_0 = \breve{p}_0 = 0$, while for $k \geq 1$, (4.87) and (4.99) combine to yield

$$\frac{\dot{p}_k}{\breve{p}_k} = \left(\frac{1-u^k}{1-u}\right)(1-p_0),$$

[9] Unless $p_k > 0$ for only one $k \geq 1$ – see Example 4.4d(7).

which is strictly increasing in k. Thus $(\check{X}_1, \dot{X}_1)$ satisfies the strict MLR property so $\check{X}_1 <_{\text{st}} \dot{X}_1$. *This confirms the fact that* $\{\dot{X}_n\}$ *does not evolve according to a GW branching process,* for otherwise (4.2) would imply that $\check{X}_n <_{\text{st}} \dot{X}_n$ for all n, contradicting the inequality $E(\dot{X}_n) < \mathrm{E}(\check{X}_n)$. □

Here is a summary of these results:

$$\tilde{X}_n \begin{cases} <_{\text{st}} \quad X_n \quad <_{\text{st}} \begin{Bmatrix} \dot{X}_n \\ \check{X}_n \end{Bmatrix} \\ \overset{(u \le \frac{1}{2})}{<_{\text{st}}} \quad \breve{X}_n \quad \begin{Bmatrix} <_{\text{st}} \quad \dot{X}_n \\ \overset{(-)}{<_{\text{st}}} \quad \check{X}_n \end{Bmatrix} \end{cases}, \tag{4.105}$$

$$\check{X}_1 < \dot{X}_1, \qquad \dot{X}_n \overset{(-)}{\le}_{\text{st}} \check{X}_n \text{ for } n \ge 2, \tag{4.106}$$

where "(-)" indicate that no general result holds

Remark 4.9. Because $\check{X}_1 <_{\text{st}} \dot{X}_1$, it follows from (4.94) with $n = 1$ that

$$\frac{\mu}{1 - p_0} = \mathrm{E}(\check{X}_1) < \mathrm{E}(\dot{X}_1) = \frac{\mu - \tilde{\mu} u}{1 - u},$$

from which a lower bound for the extinction probability u is obtained:

$$u > \frac{\mu p_0}{\mu - \tilde{\mu}(1 - p_0)} \tag{4.107}$$

□

We now examine the stochastic orderings (1)-(7) specifically for the quadratic example previously considered.

Example 4.4d *(The quadratic case, continued).*
From Examples 4.4a-4.4c, (4.87), and (4.99), we find that Case B obtains iff $0 < p_0 < p_2$, that $u = p_0/p_2$, $\mu = 1 + p_2 - p_0 > 1$, and $\tilde{\mu} = 1 + p_0 - p_2 < 1$, and that

$$\phi_{p_0,p_2}(s) = p_0 + (1 - p_0 - p_2)s + p_2 s^2, \tag{4.108}$$

$$\tilde{\phi}_{p_0,p_2}(s) = p_2 + (1 - p_0 - p_2)s + p_0 s^2 \quad \equiv \phi_{p_2,p_0}(s), \tag{4.109}$$

$$\breve{\phi}_{p_0,p_2}(s) = (1 + p_0 - p_2)s + (p_2 - p_0)s^2 \equiv \phi_{0,p_2-p_0}(s), \tag{4.110}$$

$$\dot{\phi}_{p_0,p_2}(s) = (1 - p_0 - p_2)s + (p_0 + p_2)s^2 \equiv \phi_{0,p_0+p_2}(s), \tag{4.111}$$

$$\check{\phi}_{p_0,p_2}(s) = \frac{1 - p_0 - p_2}{1 - p_0} s + \frac{p_2}{1 - p_0} s^2 \quad \equiv \phi_{0,\frac{p_2}{1-p_0}}(s). \tag{4.112}$$

From (4.109), (4.108), (4.112), and (4.111), the distributions of $\tilde{X}_1$, X_1, $\check{X}_1$, and $\dot{X}_1$ on $(0,1,2)$ are (p_2, p_1, p_0), (p_0, p_1, p_2), $(0, \frac{1-p_0-p_2}{1-p_0}, \frac{p_2}{1-p_0})$, and $(0, 1-p_0-p_2,\ p_0+p_2)$, respectively. Because $p_0 < p_2 < \frac{p_2}{1-p_0} \leq p_0 + p_2$, therefore

$$\tilde{X}_1 <_{\text{st}} X_1 <_{\text{st}} \check{X}_1 \leq_{\text{st}} \dot{X}_1, \tag{4.113}$$

the last inequality being strict unless $p_0+p_2 = 1$ (in which case $\check{X}_1 = \dot{X}_1 = 2$ with probability 1). As in Example 4.4c, it follows from (4.110) - (4.112) that

$$\breve{\xi} \equiv \breve{X}_1 \stackrel{d}{=} 1 + \text{Bernoulli}\,(p_2 - p_0), \tag{4.114}$$

$$\dot{\xi} \equiv \dot{X}_1 \stackrel{d}{=} 1 + \text{Bernoulli}\,(p_0 + p_2), \tag{4.115}$$

$$\check{\xi} \equiv \check{X}_1 \stackrel{d}{=} 1 + \text{Bernoulli}\,(\tfrac{p_2}{1-p_0}), \tag{4.116}$$

hence, since $p_2 - p_0 < \frac{p_2}{1-p_0}$, additionally

$$\breve{X}_1 <_{\text{st}} \check{X}_1. \tag{4.117}$$

Now consider the orderings (1) - (7) for $n \geq 2$ in this quadratic example:

(1) $\tilde{X}_n <_{\text{st}} X_n$:
Because both $\{\tilde{X}_n\}$ and $\{\tilde{X}_n\}$ are GW processes, the result (1) follows from (4.113) and (4.2) by induction on n.

(2) $X_n <_{\text{st}} \dot{X}_n$:
$\{\dot{X}_n\}$ is not a GW process and there are no simple expressions for the pgfs of X_n and $\dot{X}_n$, but (2) follows from the general result (4.101).

(3) $X_n <_{\text{st}} \check{X}_n$:
Because both $\{X_n\}$ and $\{\check{X}_n\}$ are GW processes, (3) follows from (4.113) and (4.2) by induction on n.

(4) $\tilde{X}_n <_{\text{st}} \breve{X}_n$ *when* $u \leq \frac{1}{2}$ *but not when* $u > \frac{1}{2}$:
From (4.102),

$$\Pr[\tilde{X}_1 \geq 1] < \Pr[\breve{X}_1 \geq 1]$$

for all u, while from (4.109) and (4.110),

$$\Pr[\tilde{X}_1 \geq 2] = \Pr[\tilde{X}_1 = 2] = p_0 \leq p_2 - p_0 = \Pr[\breve{X}_1 = 2] = \Pr[\breve{X}_1 \geq 2]$$

when $u \equiv \frac{p_0}{p_2} \leq \frac{1}{2}$. Thus $\tilde{X}_1 <_{\text{st}} \breve{X}_1$ when $u \leq \frac{1}{2}$, so (4) holds in this case by induction on n because both $\{\tilde{X}_n\}$ and $\{\breve{X}_n\}$ are GW processes.

When $u > \frac{1}{2}$, because $\tilde{X}_n \leq 2^n$ and $\breve{X}_n \leq 2^n$ we see that

$$\begin{aligned} \Pr[\tilde{X}_n \geq 2^n] &= \Pr[\tilde{X}_n = 2^n] = p_0^n \\ &> (p_2 - p_0)^n = \Pr[\breve{X}_n = 2^n] = \Pr[\breve{X}_n \geq 2^n] \end{aligned} \tag{4.118}$$

for all $n \geq 1$. Therefore (4) fails when $u > \frac{1}{2}$, i.e., $\tilde{X}_n \not\leq_{\text{st}} \breve{X}_n$.

(5) $\breve{X}_n <_{\text{st}} \dot{X}_n$:

$\{\dot{X}_n\}$ is not a GW process and there are no simple expressions for the pgfs of X_n and $\dot{X}_n$ for $n \geq 2$, so we must appeal to the general result (4.104).

(6) $\breve{X}_n <_{\text{st}} \check{X}_n$:

Because both $\{\breve{X}_n\}$ and $\{\check{X}_n\}$ are GW processes, (6) follows from (4.117) and (4.2) by induction on n.

(7) $\dot{X}_n \not\leq_{\text{st}} \check{X}_n$:

Here too $\dot{X}_n \leq 2^n$ and $\check{X}_n \leq 2^n$, so

$$\Pr[\check{X}_n \geq 2^n] = \Pr[\check{X}_n = 2^n] = \frac{\Pr[X_n = 2^n]}{1 - p_0} = \frac{p_2^n}{1 - p_0}.$$

Similarly from the mixture representation (4.90),

$$\begin{aligned} \Pr[\dot{X}_n \geq 2^n] = \Pr[\dot{X}_n = 2^n] &= \frac{\Pr[X_n = 2^n] - \Pr[\tilde{X}_n = 2^n]\, u}{1 - u} \\ &= \frac{p_2^n - p_0^n \left(\frac{p_0}{p_2}\right)}{1 - \left(\frac{p_0}{p_2}\right)}. \end{aligned}$$

Therefore

$$\begin{aligned} \Pr[\dot{X}_n \geq 2^n] - \Pr[\check{X}_n \geq 2^n] &= \frac{p_2^n - p_0^n \left(\frac{p_0}{p_2}\right)}{1 - \left(\frac{p_0}{p_2}\right)} - \frac{p_2^n}{1 - p_0} \\ &= \frac{p_0[p_2^n(1 - p_2) - p_0^n(1 - p_0)]}{(p_2 - p_0)(1 - p_0)} \\ &= \frac{p_0[p_2^{n-1} p_2(1 - p_2) - p_0^{n-1} p_0(1 - p_0)]}{(p_2 - p_0)(1 - p_0)} \\ &> 0 \end{aligned} \tag{4.119}$$

because $n \geq 1$, $p_2 > p_0$, and $p_2(1-p_2) \geq p_0(1-p_0)$. Thus $\dot{X}_n \not\leq_{\text{st}} \check{X}_n$ as asserted, even though, by (4.94), $\mathrm{E}(\dot{X}_n) \ll \mathrm{E}(\check{X}_n)$ for $n > \frac{\log(1-\frac{p_0}{p_2})}{\log(1-p_0)}$.

Summarizing: In the quadratic case,

$$\tilde{X}_n \begin{cases} <_{\text{st}} \; X_n <_{\text{st}} \begin{Bmatrix} \dot{X}_n \\ \check{X}_n \end{Bmatrix} \\ \overset{(u \leq \frac{1}{2})}{<_{\text{st}}} \; \breve{X}_n <_{\text{st}} \begin{Bmatrix} \dot{X}_n \\ \check{X}_n \end{Bmatrix}, \end{cases} \tag{4.120}$$

$$\check{X}_1 <_{\text{st}} \dot{X}_1, \qquad \dot{X}_n \not\leq_{\text{st}} \check{X}_n \text{ for } n \geq 2. \tag{4.121}$$

□

Exercise 4.5 *(The quadratic case, continued).*
In the quadratic case, express $\mathrm{E}(\dot{X}_n)$ and $\mathrm{E}(\check{X}_n)$ in terms of p_0 and p_2, and show directly that $\mathrm{E}(\check{X}_n)/\mathrm{E}(\dot{X}_n) \to \infty$ as $n \to \infty$. □

Example 4.5 *(A cubic case).*
Consider the GW process with cubic pgf

$$\phi(s) = \delta + (1-3\delta)s + 2\delta s^3, \tag{4.122}$$

where $0 < \delta < \frac{1}{3}$. Then $p_0 = \delta > 0$ and $\mu = \phi''(1) = 1 + 3\delta > 1$ so Case B obtains. The extinction probability u satisfies the cubic equation $\phi(s) = s$, equivalently

$$2s^3 - 3s + 1 = 0, \tag{4.123}$$

which has the unique solution $u \approx 0.36602$ in $(0,1)$. From (4.79),

$$\begin{aligned} \breve{p}_1 &= \frac{1}{1-u} \sum_{l \geq 1} p_l \binom{l}{1} (1-u) u^{l-1} \\ &= (1-3\delta) + 2\delta \cdot 3u^2, \end{aligned}$$

while from (4.87), $\check{p}_1 = \frac{1-3\delta}{1-\delta}$. Therefore

$$\begin{aligned} \breve{p}_1 < \check{p}_1 &\iff (1-3\delta) + 2\delta \cdot 3u^2 < \frac{1-3\delta}{1-\delta} \\ &\iff 6u^2 < \frac{1-3\delta}{1-\delta} \qquad \text{[verify!]} \\ &\iff 0.8038 < \frac{1-3\delta}{1-\delta} \\ &\iff \delta < 0.0893. \end{aligned}$$

Because $\breve{p}_0 = \check{p}_0 = 0$, this implies that for all $n \geq 1$,

$$\Pr[\breve{X}_n \leq 1] = \Pr[\breve{X}_n = 1] = (\breve{p}_1)^n < (\check{p}_1)^n = \Pr[\check{X}_n = 1] = \Pr[\check{X}_n \leq 1]$$

whenever $\delta < 0.0893$ (approximately), hence $\breve{X}_n \not\leq_{\text{st}} \check{X}_n$, so (6) does not hold in this case. □

Exercise 4.2f *(Geometric offspring distribution, continued).*
As in Exercise 4.2a, let $\xi \sim \text{geometric}(p)$, that is, $p_k = pq^k$ for $k = 0, 1, 2, \ldots$ $(0 < p < 1)$, and let ϕ_p denote the pgf of ξ. For those values of p such that Case B holds (found in Exercise 4.2a), use the pgfs ϕ_p, $\tilde{\phi}_p$, $\breve{\phi}_p$, $\dot{\phi}_p$, and $\check{\phi}_p$ as in Example 4.4d to determine and/or refine the possible stochastic orderings (1) - (7) among X_n, $\tilde{X}_n$, $\breve{X}_n$, $\dot{X}_n$, and $\check{X}_n$. □

Exercise 4.2g *(Geometric offspring distribution, continued).*
In the geometric case, express $\mathrm{E}(\dot{X}_n)$ and $\mathrm{E}(\check{X}_n)$ in terms of μ and show directly that $\mathrm{E}(\check{X}_n)/\mathrm{E}(\dot{X}_n) \to \infty$ as $n \to \infty$. □

Exercise 4.4c *(Poisson offspring distribution, continued).*
As in Examples 4.3a,b, let $\xi \equiv \xi_\lambda \sim \text{Poisson}(\lambda)$, that is, $p_k \equiv p_{\lambda,k} = e^{-\lambda}\lambda^k/k!$ for $k = 0, 1, \ldots$, so $\phi_\lambda(s) \equiv \phi_\lambda(s) = e^{\lambda(s-1)}$ is the pgf of ξ. It was shown that Case B holds iff $\lambda > 1$. In this case, as in Example 4.4d use the pgfs ϕ_λ, $\tilde{\phi}_\lambda$, $\breve{\phi}_\lambda$, $\dot{\phi}_\lambda$, and $\check{\phi}_\lambda$ to determine the possible stochastic orderings (1) - (7) among X_n, $\tilde{X}_n$, $\breve{X}_n$, $\dot{X}_n$, and $\check{X}_n$. □

Exercise 4.4d *(Poisson offspring distribution, continued).*
In Case B $(\lambda > 1)$, derive these upper and lower bounds for the extinction probability u:

$$\frac{1}{e^\lambda - 1} < u < e^{1-\lambda}. \tag{4.124}$$

□

5. Poisson Processes (PP) in One and Several Dimensions

One-dimensional Poisson processes $\{N_t \mid 0 \le t < \infty\}$ were introduced in Section 1. We now derive further properties aimed at statistical applications, i.e. testing "randomness" of a set of points in d-dimensional space $\mathbf{R}^d$, $d \ge 1$, and at methods for generating PPs by simulation. We'll see that there is a close relationship between the random distribution of points from a homogeneous PP and the uniform distribution in time or space.

A (nonhomogeneous) PP on $\mathbf{R}^1_+$ with intensity function $\lambda(t) \ge 0$ has *independent Poisson increments:*

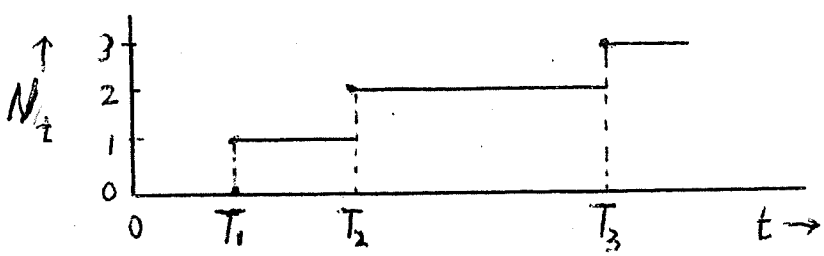
• for any $0 \equiv t_0 < t_1 < t_2 < \cdots$ the increments $N_{t_1} - N_{t_0}$, $N_{t_2} - N_{t_1}$, ... are mutually independent with

$$N_{t_i} - N_{t_{i-1}} \sim \text{Poisson}\left(\int_{t_{i-1}}^{t_i} \lambda(t)dt\right). \tag{5.1}$$

If $\lambda(t) \equiv \lambda > 0$ is constant, the PP is homogenous and (5.1) reduces to

$$N_{t_i} - N_{t_{i-1}} \sim \text{Poisson}(\lambda\,(t_i - t_{i-1})). \tag{5.2}$$

Its sample paths $\{N_t\}$ start at $N_0 = 0$ and are nonnegative, nondecreasing, integer-valued step functions, consisting of a series of jumps of size 1 at random times T_1, T_2, $T_3 \ldots$, where $0 < T_1 < T_2 < T_3 < \cdots$.



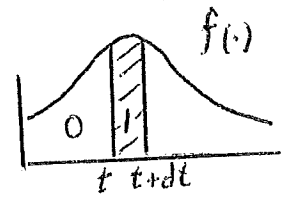
5.1. The homogeneous PP on $\mathbf{R}^1_+$: distribution of waiting times

In the homogeneous case we saw that T_1, $T_2 - T_1$, $T_3 - T_2$, ...are iid exponential(λ) rvs, hence $T_k \sim$ gamma(k, λ) (recall Proposition 1.9 and Exercise 1.7). Here is an alternate proof.

Begin with T_1; denote its pdf by $f(t)$. This was our first proof:

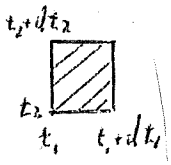
$$f(t) = \frac{d\Pr[T_1 \le t]}{dt} = \frac{d\{1 - \Pr[T_1 > t]\}}{dt} = \frac{-d\Pr[N_t = 0]}{dt} = \frac{-d(e^{-\lambda t})}{dt} = \lambda e^{-\lambda t}.$$

An alternative proof proceeds as follows: for a small increment dt,

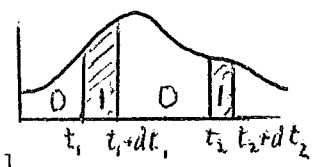


$$\begin{aligned} f(t)dt &\approx \Pr[t < T_1 \le t + dt] \\ &\approx \Pr[0 \text{ jumps in } (0,t],\ 1 \text{ jump in } (t,\ t+dt]\,] \\ &= \Pr[N_t = 0,\ N_{t+dt} - N_t = 1] \\ &= e^{-\lambda t} \cdot e^{-\lambda dt} \lambda dt && \text{[by (5.2)]} \\ &\approx \lambda e^{-\lambda t} dt. \end{aligned}$$

Now cancel "dt" to obtain $f(t) = \lambda e^{-\lambda t}$.



Now extend the second method, as follows: For $0 < t_1 < \cdots < t_k$, let $f(t_1, \ldots, t_k)$ denote the joint pdf of $T_1, \ldots, T_k$. Then



$$\begin{aligned} & f(t_1, \ldots, t_k)\, dt_1 \cdots dt_k \\ \approx\ & \Pr[t_1 < T_1 \le t_1 + dt_1,\ t_2 < T_2 \le t_2 + dt_2, \ldots, t_k < T_k \le t_k + dt_k] \\ \approx\ & \Pr[0 \text{ jumps in } (0, t_1],\ 1 \text{ jump in } (t_1,\ t_1 + dt_1], \\ & \quad 0 \text{ jumps in } (t_1 + dt_1,\ t_2],\ 1 \text{ jump in } (t_2,\ t_2 + dt_2], \ldots, \\ & \quad 0 \text{ jumps in } (t_{k-1} + dt_{k-1},\ t_k],\ 1 \text{ jump in } (t_k,\ t_k + dt_k]] \\ =\ & e^{-\lambda t_1} \cdot e^{-\lambda dt_1} \lambda dt_1 \cdot e^{-\lambda(t_2 - t_1 - dt_1)} \cdot e^{-\lambda dt_2} \lambda dt_2 \cdots e^{-\lambda(t_k - t_{k-1} - dt_{k-1})} \cdot e^{-\lambda dt_k} \lambda dt_k \\ \approx\ & \lambda^k e^{-\lambda t_k} dt_1 \cdots dt_k. \end{aligned}$$

Cancel "$dt_1 \cdots dt_k$" to obtain

$$f(t_1, \ldots, t_k) = \begin{cases} \lambda^k e^{-\lambda t_k} & \text{if } 0 < t_1 < \cdots < t_k; \\ 0 & \text{otherwise.} \end{cases} \tag{5.3}$$

Continuing, define $W_1 = T_1$, $W_2 = T_2 - T_1, \ldots, W_k = T_k = T_{k-1}$. The range of $(W_1, \ldots, W_k)$ is $\mathbf{R}_+^k \equiv \{(w_1, \ldots, w_k) \mid 0 \le w_i < \infty\}$. The inverse transformation is triangular:

$$\begin{aligned} T_1 &= W_1 \\ T_2 &= W_1 + W_2 \\ &\ \ \vdots \\ T_k &= W_1 + W_2 + \cdots + W_k \end{aligned} \tag{5.4}$$

with Jacobian $= 1$. Thus by (5.3) and (5.4), the pdf of $(W_1, \ldots, W_k)$ is

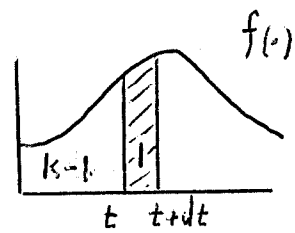
$$f(w_1, \ldots, w_k) = \lambda^k e^{-\lambda(w_1 + \cdots + w_k)} \quad \text{on } \mathbf{R}_+^k. \tag{5.5}$$

Thus, as asserted, $W_1, \ldots, W_k$ are iid exponential(λ) rvs. By (5.4), this implies that $T_k \sim \text{gamma}(k, \lambda)$. □

We present two direct demonstrations that $T_k \sim \text{gamma}(k, \lambda)$:

(i) Let $f(t)$ denote the pdf of T_k. Then for $t > 0$,

$$\begin{aligned} f(t)dt &\approx \Pr[t < T_k \le t + dt] \\ &\approx \Pr[k-1 \text{ jumps in } (0, t],\ 1 \text{ jump in } (t,\ t + dt] \\ &= e^{-\lambda t} \frac{(\lambda t)^{k-1}}{(k-1)!} \cdot e^{-\lambda dt} \lambda dt \\ &\approx \frac{\lambda^k t^{k-1}}{(k-1)!} e^{-\lambda t} dt, \end{aligned}$$



so

$$f(t) = \frac{\lambda^k}{\Gamma(k)} t^{k-1} e^{-\lambda t}, \qquad (0 < t < \infty). \tag{5.6}$$

(ii) As above for T_1,

$$\begin{aligned} f(t) &= \frac{d \Pr[T_k \le t]}{dt} = \frac{-d \Pr[T_k > t]}{dt} = \frac{-d \Pr[N_t \le k-1]}{dt} \\ &= \frac{-d \sum_{i=0}^{k-1} e^{-\lambda t} \frac{(\lambda t)^i}{i!}}{dt} \\ &= \sum_{i=0}^{k-1} e^{-\lambda t} \frac{\lambda^{i+1} t^i}{i!} - \sum_{i=1}^{k-1} e^{-\lambda t} \frac{\lambda^i t^{i-1}}{(i-1)!} \\ &= e^{-\lambda t} \left\{ \sum_{i=0}^{k-1} \frac{\lambda^{i+1} t^i}{i!} - \sum_{j=0}^{k-2} \frac{\lambda^{j+1} t^j}{j!} \right\} \\ &= e^{-\lambda t} \frac{\lambda^k t^{k-1}}{(k-1)!}, \end{aligned}$$

from which (5.6) again follows (see PK p.242). □

5.2. Relation between the jump times of a homogeneous PP on $\mathbf{R}^1_+$ and the order statistics from the uniform distribution

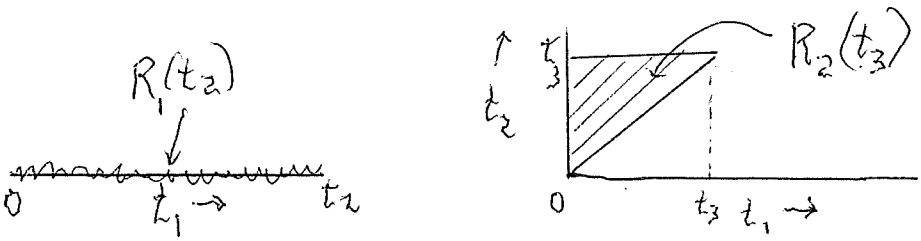
From (5.3) we see that the pdf $f(t_1, \ldots, t_k)$ of $(T_1, \ldots, T_k)$ depends only on t_k, not on $t_1, \ldots, t_{k-1}$. Thus the conditional pdf of $(T_1, \ldots, T_{k-1}) \mid T_k$,

$$f(t_1, \ldots, t_{k-1} \mid t_k) \equiv \frac{f(t_1, \ldots, t_{k-1}, t_k)}{f(t_k)}, \tag{5.7}$$

does not depend on $(t_1, \ldots, t_{k-1})$, hence *the conditional distribution of* $(T_1, \ldots, T_{k-1}) \mid T_k$ *is uniform over the conditional range*

$$\mathbf{R}_{k-1}(t_k) \equiv \{(t_1, \ldots, t_{k-1}) \mid 0 < t_1 < \cdots < t_{k-1} < t_k\}. \tag{5.8}$$

$k = 2$: $k = 3$:



Set $V_1 = \frac{T_1}{T_k}$, $V_2 = \frac{T_2}{T_k}$, $\ldots$, $V_{k-1} = \frac{T_{k-1}}{T_k}$, so $0 < V_1 < \cdots < V_{k-1} < 1$. For $k = 2$, $T_1 \mid T_2$ is uniformly distributed on $\mathbf{R}_1(t_2) \equiv (0, t_2)$, so

$$V_1 \mid T_2 \sim \text{uniform}(0, 1) \equiv \text{uniform}(\mathbf{R}_1(1)), \quad \text{so } V_1 \perp\!\!\!\perp T_2. \tag{5.9}$$

For $k = 3$, $(T_1, T_2) \mid T_3$ is uniform on $\mathbf{R}_2(t_3) \equiv \{0 < t_1 < t_2 < t_3\}$, hence

$$(V_1, V_2) \mid T_3 \sim \text{uniform}(\mathbf{R}_2(1)), \quad \text{so } (V_1, V_2) \perp\!\!\!\perp T_3. \tag{5.10}$$

For general k,

$$(V_1, \ldots, V_{k-1}) \mid T_k \sim \text{uniform}(\mathbf{R}_{k-1}(1)), \quad \text{so } (V_1, \ldots, V_{k-1}) \perp\!\!\!\perp T_k. \tag{5.11}$$

Therefore, unconditionally as well,

$$(V_1, \ldots, V_{k-1}) \sim \text{uniform}(\mathbf{R}_{k-1}(1)). \tag{5.12}$$

Note that this unconditional distribution of $(V_1, \ldots, V_{k-1})$ does not depend on the parameter λ.

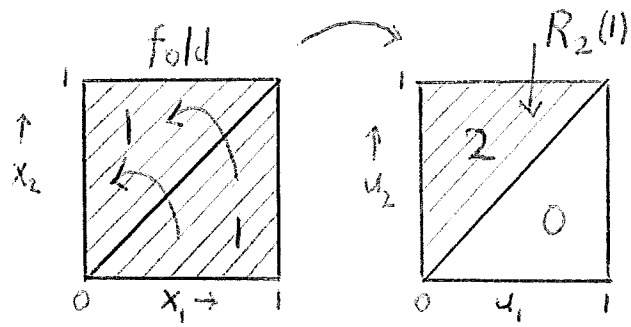
Proposition 5.1. *The distribution of the order statistics $U_1 < \cdots < U_{k-1}$ based on an iid sample $X_1, \ldots, X_{k-1}$ from the uniform distribution on $(0,1)$ is the uniform distribution on $\mathbf{R}_{k-1}(1)$, hence coincides with the distribution of $(V_1, \ldots, V_{k-1})$ in (5.12).*

Thus we can test the hypothesis that the observed jump points come from a homogeneous PP by testing $V_1, \ldots, V_{k-1}$ for uniformity on $(0,1)$.[10]

Proof. The assertion is trivial for $k = 2$. For $k = 3$, (X_1, X_2) is distributed uniformly over the unit square. The distribution of the order statistics (U_1, U_2) is obtained by "folding" the square about its main diagonal onto $\mathbf{R}_2(1)$ [see figure]. Since this folding preserves uniformity, the distribution of (U_1, U_2) coincides with (5.10).

$U_1 = \min(X_1, X_2)$:

$U_2 = \max(X_1, X_2)$



Here is a precise argument for $k = 3$: for $0 < u_1 < u_2 < 1$,

$$
\begin{aligned}
(5.13) \qquad f(u_1, u_2) &= \frac{\partial^2}{\partial u_1 \partial u_2} \Pr[U_1 \le u_1,\, U_2 \le u_2] \\
&= \frac{\partial^2}{\partial u_1 \partial u_2} \{\Pr[U_2 \le u_2] - \Pr[U_1 > u_1,\, U_2 \le u_2]\} \\
&= -\frac{\partial^2}{\partial u_1 \partial u_2} \Pr[u_1 < X_1 \le u_2,\, u_1 < X_2 \le u_2] \\
&= \frac{-\partial^2}{\partial u_1 \partial u_2} (u_2 - u_1)^2 \\
&= 2 \cdot 1_{\mathbf{R}_2(1)}(u_1, u_2),
\end{aligned}
$$

hence $(U_1,\, U_2)$ is uniformly distributed on $\mathbf{R}_2(1)$ as asserted. □

Exercise 5.1. Prove Proposition 5.1 for $k \ge 4$. □

[10] Use Pearson's chi-square statistic or the Kolmogorov-Smirnov statistic, for example.

Since $T_1 = V_1 T_k, \ldots, T_{k-1} = V_{k-1} T_k$, we obtain the following corollary:

Corollary 5.1. *The conditional distribution of* $(T_1, \ldots, T_{k-1}) \mid T_k$ *coincides with the distribution of the order statistics based on an iid sample of size* $k-1$ *from the uniform distribution on* $(0, T_k)$*, namely, the uniform distribution on* $\mathbf{R}_{k-1}(T_k)$*.* □

In Proposition 5.1, the uniform distribution arises from a homogeneous PP by conditioning on the *random time* T_k. The uniform distribution also arises when by conditioning on N_t for a *fixed time* t.

Proposition 5.2. *The conditional distribution of* $(T_1, \ldots, T_n) \mid N_t = n$ *is the uniform distribution on* $\mathbf{R}_n(t)$*, hence coincides with the distribution of the order statistics based on an iid sample of size* n *from the uniform distribution on* $(0, t)$*.*

Thus we can test the hypothesis that the observed jump points come from a homogeneous PP by testing $T_1/t, \ldots, T_n/t$ for uniformity on $(0, 1)$.

Proof. First consider $n = 1$. Then for $0 < t_1 < t$,

$$\begin{aligned}
(5.14)\quad \Pr[T_1 \le t_1 \mid N_t = 1] &= \Pr[T_1 \le t_1,\, N_t = 1] / \Pr[N_t = 1] \\
&= \Pr[N_{t_1} = 1,\, N_t - N_{t_1} = 0] / \Pr[N_t = 1] \\
&= e^{-\lambda t_1} \frac{(\lambda t_1)^1}{1!} \cdot e^{-\lambda(t-t_1)} \frac{(\lambda(t-t_1))^0}{0!} / e^{-\lambda t} \frac{(\lambda t)^1}{1!} \\
&= t_1/t,
\end{aligned}$$

so

$$(5.15) \qquad f(t_1|t) \equiv \frac{d(t_1/t)}{dt_1} = \frac{1}{t} 1_{(0,t)}(t_1),$$

the pdf of the uniform distribution on $(0, t) \equiv \mathbf{R}_1(t)$ as asserted.

For $n = 2$ we can proceed as in (5.13): for $0 < t_1 < t_2 < t$,

$$\begin{aligned}
(5.16)\quad & f(t_1,\, t_2 \mid N_t = 2) \\
&= \frac{\partial^2}{\partial t_1 \partial t_2} \Pr[T_1 \le t_1,\, T_2 \le t_2 \mid N_t = 2] \\
&= \frac{\partial^2}{\partial t_1 \partial t_2} \{\Pr[T_1 \le t_1 \mid N_t = 2] - \Pr[T_1 \le t_1,\, T_2 > t_2 \mid N_t = 2]\}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{-\partial^2}{\partial t_1 \partial t_2} \Pr[T_1 \le t_1,\, T_2 > t_2 \mid N_t = 2] \\
&= \frac{-\partial^2}{\partial t_1 \partial t_2} \left\{ \frac{\Pr[1 \text{ jump in } (0, t_1],\, 0 \text{ jumps in } (t_1, t_2],\, 1 \text{ jump in } (t_2, t]]}{\Pr[N_t = 2]} \right\} \\
&= \frac{-\partial^2}{\partial t_1 \partial t_2} \left\{ \frac{e^{-\lambda t_1} \lambda t_1 \cdot e^{-\lambda(t_2 - t_1)} \cdot e^{-\lambda(t - t_2)} \lambda (t - t_2)}{e^{-\lambda t} (\lambda t)^2 / 2!} \right\} \\
&= \frac{-\partial^2}{\partial t_1 \partial t_2} \left\{ \frac{2 t_1 (t - t_2)}{t^2} \right\} \\
&= \frac{2!}{t^2},
\end{aligned}
$$

the pdf of the uniform distribution on $\mathbf{R}_2 \equiv \{(t_1, t_2) \mid 0 < t_1 < t_2 < t\}$.

The preceding argument can be extended to the case $n \ge 3$. Alternatively, we can proceed as follows: for $0 < t_1 < \cdots < t_n < t$,

$$
\begin{aligned}
(5.17) \quad & \Pr[T_1 \le t_1,\, T_2 \le t_2, \ldots, T_n \le t_n \mid N_t = n] \\
&= \Pr[T_1 \le t_1,\, T_1 < T_2 \le t_2, \ldots, T_{n-1} < T_n \le t_n,\, T_{n+1} > t] / \Pr[N_t = n] \\
&\overset{*}{=} \int_0^{t_1} \int_{s_1}^{t_2} \cdots \int_{s_{n-1}}^{t_n} \int_t^{\infty} \lambda^{n+1} e^{-\lambda s_{n+1}} ds_{n+1} ds_n \cdots ds_2 ds_1 / [e^{-\lambda t} (\lambda t)^n / n!] \\
&= \frac{n!}{t^n} \int_0^{t_1} \int_{s_1}^{t_2} \cdots \int_{s_{n-1}}^{t_n} ds_n \cdots ds_2 ds_1,
\end{aligned}
$$

where * follows from (5.3), p.124, with $k = n+1$. Therefore the conditional pdf of $(T_1, \ldots, T_n) \mid N_t = n$ is obtained by differentiating (5.17) successively w.r.to $t_1, t_2, \ldots t_n$ to obtain

$$
(5.18) \qquad f(t_1, \ldots, t_n \mid N_t = n) = \frac{n!}{t^n},
$$

the pdf of the uniform distribution on $\mathbf{R}_n(t)$, as asserted. □

Remark 5.1. It follows from Proposition 5.2 and the stationarity of a homogeneous PP that for any interval $[a, b]$, given that exactly n jumps have occurred in this interval, the locations of the jumps are distributed as the order statistics from an iid sample distributed uniformly on $[a, b]$. □

Exercise 5.2. Clearly $N_{T_k} = k$ for fixed k (set $T_0 = 0$). For fixed $t > 0$, find the distribution of T_{N_t}, the time of the last jump before time t, and evaluate $\mathrm{E}[T_{N_t}]$. Show that $\mathrm{E}[T_{N_t}] \uparrow t$ at rate $O(\frac{1}{\lambda})$ as $\lambda \to \infty$.

Hint: Since $T_0 = 0$, the distribution of T_{N_t} is a mixture of a point mass at 0 and a continuous pdf on $(0, t]$. □

Corollary 5.1 and Proposition 5.2 can be viewed as *backward memory-free* properties of a homogeneous PP: given either the time T_k of the kth jump or the number N_t of jumps by time t, these results show that *the past behavior of the process is uniformly distributed over its entire history.* Past events are equally likely to have occurred in the recent, moderate, or distant past. This should be contrasted with the *forward* memory-free property of the exponential waiting times $T_1, T_2, \ldots$ (Proposition 1.9).

The following result is another manifestation of the backward memory-free nature of the homogeneous PP. It extends PK Theorem 3.5.6, p.244.

Proposition 5.3. *Suppose that $N_t = n$, i.e., exactly n events occurred in the interval $(0, t]$. Then for any partition $0 \equiv t_0 < t_1 < \cdots < t_{k-1} < t_k \equiv t$ of $[0, t]$ into k subintervals, the n events are allocated to these subintervals at random according to a multinomial distribution. Specifically,*

$$\begin{aligned}&(N_{t_1}, N_{t_2} - N_{t_1}, \ldots, N_t - N_{t_{k-1}}) \mid \{N_t = n\}\\ &\quad \sim \mathrm{M}_k(n;\, p_1, p_2, \ldots, p_k),\end{aligned} \tag{5.19}$$

where the ith probability $p_i \equiv (t_i - t_{i-1})/t$ is proportional to the length of the ith interval.

Proof. For any set of nonnegative integers $n_1, n_2, \ldots, n_k$ with $\sum n_i = n$,

$$\begin{aligned}&\Pr[N_{t_1} = n_1,\, N_{t_2} - N_{t_1} = n_2, \ldots,\, N_t - N_{t_{k-1}} = n_k \mid N_t = n]\\ =&\frac{e^{-\lambda t_1}(\lambda t_1)^{n_1}}{n_1!} \cdot \frac{e^{-\lambda(t_2 - t_1)}(\lambda(t_2 - t_1))^{n_2}}{n_2!} \cdots \frac{e^{-\lambda(t - t_{k-1})}(\lambda(t - t_{k-1}))^{n_{k-1}}}{n_{k-1}!}\\ &\div \frac{e^{-\lambda t}(\lambda t)^n}{n!}\\ =&\frac{n!}{n_1! \cdots n_k!}\left(\frac{t_1}{t}\right)^{n_1}\left(\frac{t_2 - t_1}{t}\right)^{n_2} \cdots \left(\frac{t - t_{k-1}}{t}\right)^{n_k},\end{aligned}$$

which confirms (5.19). □

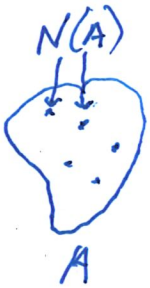
Remark 5.2. The conditional distributions in Corollary 5.1 (conditioning on T_k) and in Propositions 5.2 and 5.3 (conditioning on $N_t = n$) do not depend on the intensity parameter λ. Suppose that we wish to make an inference about the value of λ based on observing the PP either up to the random time T_k of the kth jump or up to the fixed time t. Then *once we know the value of T_k or N_t, neither the prior jump times $T_1, \ldots, T_{k-1}$ nor the prior values N_s, $0 < s < t$, provide any further information relevant to the value of λ.* We say, therefore, that T_k or N_t is a *sufficient statistic* for the parameter λ. □

Remark 5.3. One way to simulate a homogeneous PP with rate λ is to simulate iid exponential rvs W_1, W_2, ..., then set $T_1 = W_1$, $T_2 = W_1 + W_2$, ..., in order to simulate the jump times of the process (recall (5.4)).

A second method to simulate the jump times, based on Corollary 5.1, is to simulate $T_k \sim \text{gamma}(k, \lambda)$ then independently simulate $T_1 < \ldots < T_{k-1}$ as the order statistics from an iid sample of size $k-1$ from the uniform distribution on $[0, T_k]$.

A third method, which simulates the PP(λ) on a fixed time interval $[0, t]$ rather than a random time interval T_k, is based on Proposition 5.2: simulate $N_t \sim \text{Poisson}(\lambda t)$, then simulate $T_1 < \ldots < T_{N_t}$ as the order statistics from an iid sample of size N_t from the uniform distribution on $[0, t]$. Here the time interval $[0, t]$ is fixed while the number of jumps n in this interval is random; the reverse was true in the first two methods. □

5.3. Poisson processes in $\mathbf{R}^d$

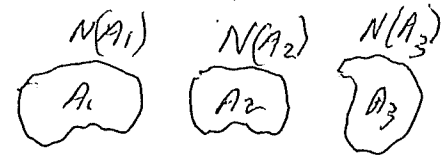


A PP in $\mathbf{R}^d$ represents the occurrence of discrete events, or jumps, at random points in $\mathbf{R}^d$. For any region $A \subseteq \mathbf{R}^d$ we denote the (random) number of events that occur in A by $N(A)$ and extend the defining property (5.1) of a 1-dimensional PP to a d-dimensional PP as follows:

Let $\lambda(t)$ be a nonnegative function on $\mathbf{R}^d$, where $t = (t_1, \ldots, t_d)$. A (nonhomogeneous) PP on $\mathbf{R}^d$ with intensity function $\lambda(\cdot)$ has *independent Poisson "increments"* in the following sense:

- for any collection of mutually disjoint regions A_1, $A_2, \ldots$ in $\mathbf{R}^d$ such that $\int_{A_i} \lambda(t)dt < \infty$, the random counts $N(A_1)$, $N(A_2)$, ... are mutually inde-

pendent with



$$N(A_i) \sim \text{Poisson}\left(\int_{A_i} \lambda(t)dt\right). \tag{5.20}$$

If $\lambda(\cdot) \equiv \lambda > 0$ is constant, the PP is homogenous and (5.20) reduces to

$$N(A_i) \sim \text{Poisson}\left(\lambda \cdot |A_i|\right), \tag{5.21}$$

where $|A|$ denotes the d-dimensional volume of A. Note that it is no longer possible to think in terms of "sample paths" of the PP (but see Proposition 5.6); instead we should visualize a collection of random points in $\mathbf{R}^d$.

Proposition 5.3 extends to $\mathbf{R}^d$ as follows:

Proposition 5.4. *Suppose that $N(A) = n$, i.e., exactly n events have occurred in the region $A \subseteq \mathbf{R}^d$. Then for any partition $A = A_1 \dot\cup \cdots \dot\cup A_k$ of A into k disjoint subregions, the n events are allocated to these subregions at random according to a multinomial distribution. Specifically,*

$$(N(A_1), \ldots, N(A_k)) \mid \{N(A) = n\} \sim \text{Multinomial}_k(n;\, p_1, \ldots, p_k), \tag{5.22}$$

where $p_i = \int_{A_i} \lambda(t)dt / \int_A \lambda(t)dt \equiv \omega_i/\omega$. [Note: $\sum \omega_i = \omega$ so $\sum p_i = 1$.]

Proof. For any set of nonnegative integers $n_1, \ldots, n_k$ with $\sum n_i = n$,

$$\begin{aligned}
&\Pr[N(A_1) = n_1, \ldots, N(A_k) = n_k \mid N(A) = n] \\
&= \frac{e^{-\omega_1}\omega_1^{n_1}}{n_1!} \cdots \frac{e^{-\omega_k}\omega_k^{n_k}}{n_k!} \Big/ \frac{e^{-\omega}\omega^n}{n!} \\
&= \frac{n!}{n_1! \cdots n_k!} \left(\frac{\omega_1}{\omega}\right)^{n_1} \cdots \left(\frac{\omega_k}{\omega}\right)^{n_k},
\end{aligned}$$

which confirms (5.22). □

Proposition 5.4 can be used to extend Proposition 5.2 to $\mathbf{R}^d$ as follows:

Proposition 5.5. *Let $A \subseteq \mathbf{R}^d$. Given that $N(A) = n$, the conditional distribution of the n random points in A is that of an iid sample of size n from the distribution with pdf on A given by*

$$f_\lambda(\cdot) = \frac{\lambda(\cdot)}{\int_A \lambda(t)dt} \tag{5.23}$$

Thus we can test the hypothesis that the observed points are generated by the PP on A with intensity function $\lambda(\cdot)$ by testing the goodness-of-fit of the n observed points to the pdf $f_\lambda(\cdot)$. □

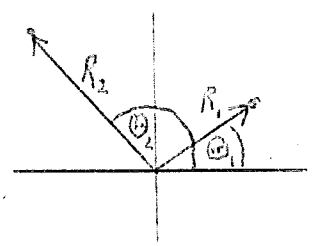
Exercise 5.3. Prove Proposition 5.5. □

Remark 5.4. One can simulate the PP with intensity function $\lambda(\cdot)$ on A by simulating $N(A) \sim \text{Poisson}(\int_A \lambda(t)dt)$, then, given $N(A) = n$, independently simulate n random points from the pdf $f_\lambda(\cdot)$ on A. □

Propositions 1.9 and 5.1 and Corollary 5.1 concern the waiting times $T_1, T_2, \ldots$ of a 1-dimensional homogeneous PP hence have no direct counterparts in $\mathbf{R}^d$. However we can obtain analogs of these waiting times for a d-dimensional homogeneous PP(λ) by using polar coordinates.

For simplicity we consider $d = 2$. Fix an arbitrary point t_0 in $\mathbf{R}^2$, for simplicity consider $t_0 = (0,0)$. Let (R_i, Θ_i) be the polar coordinates of the random point (X_i, Y_i) of the PP that is the ith closest to 0, so $0 < R_1 < R_2 < \cdots$. For $i = 0, 1, 2, \ldots$, define

$$\begin{aligned} T_i &= R_i^2, \\ W_i &= T_i - T_{i-1} = R_i^2 - R_{i-1}^2, \end{aligned} \tag{5.24}$$



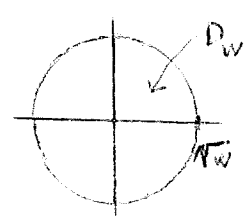
where $R_0 = 0$. Thus $0 < T_1 < T_2 < \cdots$ and each $W_i > 0$.

Proposition 5.6. *The random variables $W_1, \Theta_1, W_2, \Theta_2, \ldots$ are mutually independent; each $W_i \sim$ exponential($\lambda\pi$) and each $\Theta_i \sim$ uniform$(0, 2\pi)$.*

Proof. The result is similar to Proposition 1.9 and can be proved as in Exercise 1.7 or by the alternate method in §5.1, which we adopt here.

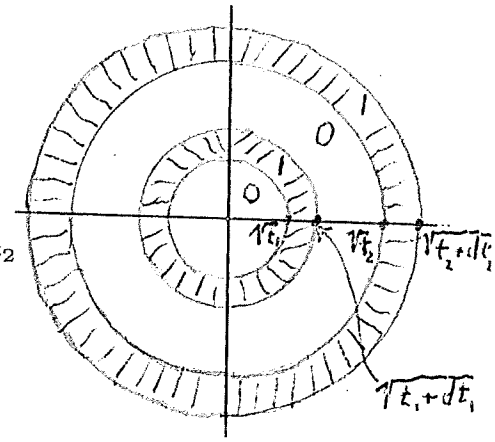
To begin gently, let $f(w)$ denote the pdf of $W_1 \equiv R_1^2$. Then for $w > 0$,

$$\begin{aligned} \Pr[W_1 > w] &= \Pr[0 \text{ points in the disk } D_w \text{ of radius } \sqrt{w}] \\ &= \Pr[N(D_w) = 0] \\ &= \Pr[\text{Poisson}(\lambda\pi w) = 0] \\ &= e^{-\lambda\pi w}, \end{aligned} \tag{5.25}$$



so $f(w) = \lambda\pi e^{-\lambda\pi w}$, hence $W_1 \sim$ exponential($\lambda\pi$) as asserted.

Next, consider the pdf $f(t_1, t_2)$ of (T_1, T_2): for $0 < t_1 < t_2$,



$$
\begin{aligned}
&f(t_1,\, t_2)\, dt_1 dt_2 \\
&\approx \Pr[\, t_1 < T_1 < t_1 + dt_1,\ t_2 < T_2 < t_2 + dt_2\,] \\
&\approx \Pr[0 \text{ points in } D_{t_1},\ 1 \text{ point in } D_{t_1+dt_1} \setminus D_{t_1}, \\
&\quad 0 \text{ points in } D_{t_2} \setminus D_{t_1+dt_1},\ 1 \text{ point in } D_{t_2+dt_2} \setminus D_{t_2}] \\
&\approx e^{-\lambda\pi t_1} \cdot \lambda\pi dt_1 e^{-\lambda\pi dt_1} \cdot e^{-\lambda\pi[t_2-(t_1+dt_1)]} \cdot \lambda\pi dt_2 e^{-\lambda\pi dt_2} \\
&\approx \lambda^2\pi^2 e^{-\lambda\pi t_2}\, dt_1 dt_2,
\end{aligned}
$$

so

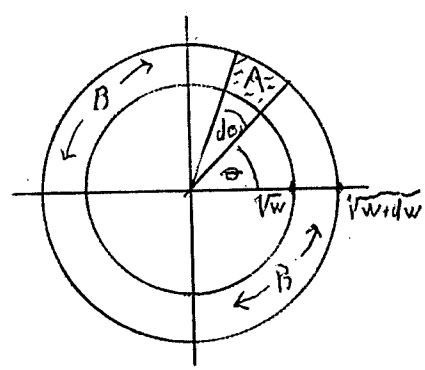
$$
f(t_1,\, t_2) = \lambda^2\pi^2 e^{-\lambda\pi t_2}, \quad 0 < t_1 < t_2 < \infty. \tag{5.26}
$$

Make the transformation $W_1 = T_1$, $W_2 = T_2 - T_1$ (Jacobian $= 1$) to obtain the pdf of $(W_1,\, W_2)$:

$$
f(w_1,\, w_2) = \lambda\pi e^{-\lambda\pi w_1} \cdot \lambda\pi e^{-\lambda\pi w_2}, \quad 0 < w_1, w_2 < \infty. \tag{5.27}
$$

Thus W_1 and W_2 are iid exponential$(\lambda\pi)$ rvs.

Now consider the joint pdf $f(w, \theta)$ of (W_1, Θ_1): for $w > 0,\ 0 \le \theta < 2\pi$,



$$
\begin{aligned}
&f(w, \theta)\, dw d\theta \\
&\approx \Pr[\, w < W_1 < w + dw,\ \theta < \Theta_1 < \theta + d\theta\,] \\
&\approx \Pr[0 \text{ points in } D_w,\ 1 \text{ point in } A,\ 0 \text{ points in } B\,] \\
&\approx e^{-\lambda\pi w} \cdot \lambda |A| e^{-\lambda|A|} \cdot e^{-\lambda|B|} \\
&\approx e^{-\lambda\pi w} \cdot \lambda\, (\pi dw \cdot \tfrac{d\theta}{2\pi}) e^{-\lambda(\pi dw \cdot \frac{d\theta}{2\pi})} \cdot e^{-\lambda(\pi dw)} \\
&\approx \lambda\pi e^{-\lambda\pi w} \cdot \tfrac{1}{2\pi}\, dw\, d\theta,
\end{aligned}
$$

hence

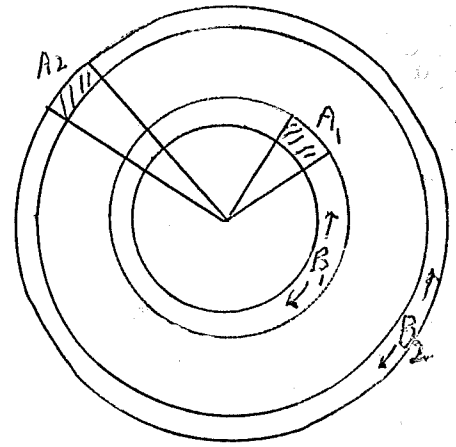
$$
f(w, \theta) = \lambda\pi e^{-\lambda\pi w} \cdot \tfrac{1}{2\pi} \equiv f(w) \cdot f(\theta). \tag{5.28}
$$

Thus $W_1 \sim$ exponential$(\lambda\pi)$, $\Theta_1 \sim$ uniform$(0, 2\pi)$, and $W_1 \perp\!\!\!\perp \Theta_1$.

Next, consider the pdf $f(t_1, \theta_1, t_2, \theta_2)$ of $(T_1, \Theta_1, T_2, \Theta_2)$. A similar argument yields the following: for $0 < t_1 < t_2 < \infty$ and $0 \le \theta_1, \theta_2 < 2\pi$,

$$
f(t_1, \theta_1, t_2, \theta_2)\, dt_1 d\theta_1 dt_2 d\theta_2
$$

$$\approx e^{-\lambda\pi t_1} \cdot \lambda|A_1|e^{-\lambda|A_1|} \cdot e^{-\lambda|B_1|}$$
$$\cdot\, e^{-\lambda\pi(t_2-t_1)} \cdot \lambda|A_2|e^{-\lambda|A_2|} \cdot e^{-\lambda|B_2|}$$
$$\approx e^{-\lambda\pi t_1} \cdot \lambda\,(\pi dt_1 \tfrac{d\theta_1}{2\pi})e^{-\lambda(\pi dt_1 \frac{d\theta_1}{2\pi})} \cdot e^{-\lambda(\pi dt_1)}$$
$$e^{-\lambda\pi(t_2-t_1)} \cdot \lambda\,(\pi dt_2 \tfrac{d\theta_2}{2\pi})e^{-\lambda(\pi dt_2 \frac{d\theta_2}{2\pi})} \cdot e^{-\lambda(\pi dt_2)}$$
$$\approx \lambda^2\pi^2 e^{-\lambda\pi t_2} \cdot (\tfrac{1}{2\pi})^2\, dt_1 d\theta_1 dt_2 d\theta_2,$$



hence (see (5.26))

$$\begin{aligned} f(t_1, \theta_1, t_2, \theta_2) &= \lambda^2\pi^2 e^{-\lambda\pi t_2} \cdot (\tfrac{1}{2\pi})^2 \\ &\equiv f(t_1,\, t_2) \cdot f(\theta_1) \cdot f(\theta_2), \end{aligned} \tag{5.29}$$

so $(T_1,\, T_2) \perp\!\!\!\perp \Theta_1 \perp\!\!\!\perp \Theta_2$. By (5.27), therefore,

$$f(w_1,\, \theta_1,\, w_2,\, \theta_2) = f(w_1) \cdot f(w_2) \cdot f(\theta_1) \cdot f(\theta_2), \tag{5.30}$$

with $W_i \sim$ exponential$(\lambda\pi)$ and $\Theta_i \sim$ uniform$(0, 2\pi)$, $i = 1, 2$.

The general result for $W_1,\, \Theta_1, \ldots, W_k,\, \Theta_k$ is proved similarly. □

Remark 5.5. Thus one can simulate the PP(λ) on $\mathbf{R}^2$ by simulating $W_1, \Theta_1, W_2, \Theta_2, \ldots$ as in Proposition 5.6, setting $R_i = \sqrt{W_1 + \cdots + W_i}$, and taking

$$(X_i, Y_i) = (R_i \cos\Theta_i,\; R_i \sin\Theta_i), \quad i = 1, 2, \ldots, \tag{5.25}$$

to be the random points of the PP.

To simulate a homogeneous PP(λ) on the first quadrant $\mathbf{R}^2_+$, proceed as above but take $\Theta_i \sim$ uniform$(0, \pi/2)$ [verify]. □

Exercise 5.5. Extend Proposition 5.6 and Remark 5.5 to homogeneous PPs in $\mathbf{R}^d$, $d \geq 3$. □

5.4. Compound/Marked Poisson processes